Pharmacogenomics

DRUGS AND THE PHARMACEUTICAL SCIENCES

DRUGS AND THE PHARMACEUTICAL SCIENCES

A Series of Textbooks and Monographs

1. Pharmacokinetics, *Milo Gibaldi and Donald Perrier*
2. Good Manufacturing Practices for Pharmaceuticals: A Plan for Total Quality Control, *Sidney H. Willig, Murray M. Tuckerman, and William S. Hitchings IV*
3. Microencapsulation, *edited by J. R. Nixon*
4. Drug Metabolism: Chemical and Biochemical Aspects, *Bernard Testa and Peter Jenner*
5. New Drugs: Discovery and Development, *edited by Alan A. Rubin*
6. Sustained and Controlled Release Drug Delivery Systems, *edited by Joseph R. Robinson*
7. Modern Pharmaceutics, *edited by Gilbert S. Banker and Christopher T. Rhodes*
8. Prescription Drugs in Short Supply: Case Histories, *Michael A. Schwartz*
9. Activated Charcoal: Antidotal and Other Medical Uses, *David O. Cooney*
10. Concepts in Drug Metabolism (in two parts), *edited by Peter Jenner and Bernard Testa*
11. Pharmaceutical Analysis: Modern Methods (in two parts), *edited by James W. Munson*
12. Techniques of Solubilization of Drugs, *edited by Samuel H. Yalkowsky*
13. Orphan Drugs, *edited by Fred E. Karch*
14. Novel Drug Delivery Systems: Fundamentals, Developmental Concepts, Biomedical Assessments, *Yie W. Chien*
15. Pharmacokinetics: Second Edition, Revised and Expanded, *Milo Gibaldi and Donald Perrier*
16. Good Manufacturing Practices for Pharmaceuticals: A Plan for Total Quality Control, Second Edition, Revised and Expanded, *Sidney H. Willig, Murray M. Tuckerman, and William S. Hitchings IV*
17. Formulation of Veterinary Dosage Forms, *edited by Jack Blodinger*
18. Dermatological Formulations: Percutaneous Absorption, *Brian W. Barry*
19. The Clinical Research Process in the Pharmaceutical Industry, *edited by Gary M. Matoren*
20. Microencapsulation and Related Drug Processes, *Patrick B. Deasy*
21. Drugs and Nutrients: The Interactive Effects, *edited by Daphne A. Roe and T. Colin Campbell*
22. Biotechnology of Industrial Antibiotics, *Erick J. Vandamme*
23. Pharmaceutical Process Validation, *edited by Bernard T. Loftus and Robert A. Nash*

24. Anticancer and Interferon Agents: Synthesis and Properties, *edited by Raphael M. Ottenbrite and George B. Butler*
25. Pharmaceutical Statistics: Practical and Clinical Applications, *Sanford Bolton*
26. Drug Dynamics for Analytical, Clinical, and Biological Chemists, *Benjamin J. Gudzinowicz, Burrows T. Younkin, Jr., and Michael J. Gudzinowicz*
27. Modern Analysis of Antibiotics, *edited by Adjoran Aszalos*
28. Solubility and Related Properties, *Kenneth C. James*
29. Controlled Drug Delivery: Fundamentals and Applications, Second Edition, Revised and Expanded, *edited by Joseph R. Robinson and Vincent H. Lee*
30. New Drug Approval Process: Clinical and Regulatory Management, *edited by Richard A. Guarino*
31. Transdermal Controlled Systemic Medications, *edited by Yie W. Chien*
32. Drug Delivery Devices: Fundamentals and Applications, *edited by Praveen Tyle*
33. Pharmacokinetics: Regulatory • Industrial • Academic Perspectives, *edited by Peter G. Welling and Francis L. S. Tse*
34. Clinical Drug Trials and Tribulations, *edited by Allen E. Cato*
35. Transdermal Drug Delivery: Developmental Issues and Research Initiatives, *edited by Jonathan Hadgraft and Richard H. Guy*
36. Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms, *edited by James W. McGinity*
37. Pharmaceutical Pelletization Technology, *edited by Isaac Ghebre-Sellassie*
38. Good Laboratory Practice Regulations, *edited by Allen F. Hirsch*
39. Nasal Systemic Drug Delivery, *Yie W. Chien, Kenneth S. E. Su, and Shyi-Feu Chang*
40. Modern Pharmaceutics: Second Edition, Revised and Expanded, *edited by Gilbert S. Banker and Christopher T. Rhodes*
41. Specialized Drug Delivery Systems: Manufacturing and Production Technology, *edited by Praveen Tyle*
42. Topical Drug Delivery Formulations, *edited by David W. Osborne and Anton H. Amann*
43. Drug Stability: Principles and Practices, *Jens T. Carstensen*
44. Pharmaceutical Statistics: Practical and Clinical Applications, Second Edition, Revised and Expanded, *Sanford Bolton*
45. Biodegradable Polymers as Drug Delivery Systems, *edited by Mark Chasin and Robert Langer*
46. Preclinical Drug Disposition: A Laboratory Handbook, *Francis L. S. Tse and James J. Jaffe*

47. HPLC in the Pharmaceutical Industry, *edited by Godwin W. Fong and Stanley K. Lam*
48. Pharmaceutical Bioequivalence, *edited by Peter G. Welling, Francis L. S. Tse, and Shrikant V. Dinghe*
49. Pharmaceutical Dissolution Testing, *Umesh V. Banakar*
50. Novel Drug Delivery Systems: Second Edition, Revised and Expanded, *Yie W. Chien*
51. Managing the Clinical Drug Development Process, *David M. Cocchetto and Ronald V. Nardi*
52. Good Manufacturing Practices for Pharmaceuticals: A Plan for Total Quality Control, Third Edition, *edited by Sidney H. Willig and James R. Stoker*
53. Prodrugs: Topical and Ocular Drug Delivery, *edited by Kenneth B. Sloan*
54. Pharmaceutical Inhalation Aerosol Technology, *edited by Anthony J. Hickey*
55. Radiopharmaceuticals: Chemistry and Pharmacology, *edited by Adrian D. Nunn*
56. New Drug Approval Process: Second Edition, Revised and Expanded, *edited by Richard A. Guarino*
57. Pharmaceutical Process Validation: Second Edition, Revised and Expanded, *edited by Ira R. Berry and Robert A. Nash*
58. Ophthalmic Drug Delivery Systems, *edited by Ashim K. Mitra*
59. Pharmaceutical Skin Penetration Enhancement, *edited by Kenneth A. Walters and Jonathan Hadgraft*
60. Colonic Drug Absorption and Metabolism, *edited by Peter R. Bieck*
61. Pharmaceutical Particulate Carriers: Therapeutic Applications, *edited by Alain Rolland*
62. Drug Permeation Enhancement: Theory and Applications, *edited by Dean S. Hsieh*
63. Glycopeptide Antibiotics, *edited by Ramakrishnan Nagarajan*
64. Achieving Sterility in Medical and Pharmaceutical Products, *Nigel A. Halls*
65. Multiparticulate Oral Drug Delivery, *edited by Isaac Ghebre-Sellassie*
66. Colloidal Drug Delivery Systems, *edited by Jörg Kreuter*
67. Pharmacokinetics: Regulatory • Industrial • Academic Perspectives, Second Edition, *edited by Peter G. Welling and Francis L. S. Tse*
68. Drug Stability: Principles and Practices, Second Edition, Revised and Expanded, *Jens T. Carstensen*
69. Good Laboratory Practice Regulations: Second Edition, Revised and Expanded, *edited by Sandy Weinberg*
70. Physical Characterization of Pharmaceutical Solids, *edited by Harry G. Brittain*

71. Pharmaceutical Powder Compaction Technology, *edited by Göran Alderborn and Christer Nyström*
72. Modern Pharmaceutics: Third Edition, Revised and Expanded, *edited by Gilbert S. Banker and Christopher T. Rhodes*
73. Microencapsulation: Methods and Industrial Applications, *edited by Simon Benita*
74. Oral Mucosal Drug Delivery, *edited by Michael J. Rathbone*
75. Clinical Research in Pharmaceutical Development, *edited by Barry Bleidt and Michael Montagne*
76. The Drug Development Process: Increasing Efficiency and Cost Effectiveness, *edited by Peter G. Welling, Louis Lasagna, and Umesh V. Banakar*
77. Microparticulate Systems for the Delivery of Proteins and Vaccines, *edited by Smadar Cohen and Howard Bernstein*
78. Good Manufacturing Practices for Pharmaceuticals: A Plan for Total Quality Control, Fourth Edition, Revised and Expanded, *Sidney H. Willig and James R. Stoker*
79. Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms: Second Edition, Revised and Expanded, *edited by James W. McGinity*
80. Pharmaceutical Statistics: Practical and Clinical Applications, Third Edition, *Sanford Bolton*
81. Handbook of Pharmaceutical Granulation Technology, *edited by Dilip M. Parikh*
82. Biotechnology of Antibiotics: Second Edition, Revised and Expanded, *edited by William R. Strohl*
83. Mechanisms of Transdermal Drug Delivery, *edited by Russell O. Potts and Richard H. Guy*
84. Pharmaceutical Enzymes, *edited by Albert Lauwers and Simon Scharpé*
85. Development of Biopharmaceutical Parenteral Dosage Forms, *edited by John A. Bontempo*
86. Pharmaceutical Project Management, *edited by Tony Kennedy*
87. Drug Products for Clinical Trials: An International Guide to Formulation • Production • Quality Control, *edited by Donald C. Monkhouse and Christopher T. Rhodes*
88. Development and Formulation of Veterinary Dosage Forms: Second Edition, Revised and Expanded, *edited by Gregory E. Hardee and J. Desmond Baggot*
89. Receptor-Based Drug Design, *edited by Paul Leff*
90. Automation and Validation of Information in Pharmaceutical Processing, *edited by Joseph F. deSpautz*
91. Dermal Absorption and Toxicity Assessment, *edited by Michael S. Roberts and Kenneth A. Walters*

92. Pharmaceutical Experimental Design, *Gareth A. Lewis, Didier Mathieu, and Roger Phan-Tan-Luu*
93. Preparing for FDA Pre-Approval Inspections, *edited by Martin D. Hynes III*
94. Pharmaceutical Excipients: Characterization by IR, Raman, and NMR Spectroscopy, *David E. Bugay and W. Paul Findlay*
95. Polymorphism in Pharmaceutical Solids, *edited by Harry G. Brittain*
96. Freeze-Drying/Lyophilization of Pharmaceutical and Biological Products, *edited by Louis Rey and Joan C. May*
97. Percutaneous Absorption: Drugs–Cosmetics–Mechanisms–Methodology, Third Edition, Revised and Expanded, *edited by Robert L. Bronaugh and Howard I. Maibach*
98. Bioadhesive Drug Delivery Systems: Fundamentals, Novel Approaches, and Development, *edited by Edith Mathiowitz, Donald E. Chickering III, and Claus-Michael Lehr*
99. Protein Formulation and Delivery, *edited by Eugene J. McNally*
100. New Drug Approval Process: Third Edition, The Global Challenge, *edited by Richard A. Guarino*
101. Peptide and Protein Drug Analysis, *edited by Ronald E. Reid*
102. Transport Processes in Pharmaceutical Systems, *edited by Gordon L. Amidon, Ping I. Lee, and Elizabeth M. Topp*
103. Excipient Toxicity and Safety, *edited by Myra L. Weiner and Lois A. Kotkoskie*
104. The Clinical Audit in Pharmaceutical Development, *edited by Michael R. Hamrell*
105. Pharmaceutical Emulsions and Suspensions, *edited by Francoise Nielloud and Gilberte Marti-Mestres*
106. Oral Drug Absorption: Prediction and Assessment, *edited by Jennifer B. Dressman and Hans Lennernäs*
107. Drug Stability: Principles and Practices, Third Edition, Revised and Expanded, *edited by Jens T. Carstensen and C. T. Rhodes*
108. Containment in the Pharmaceutical Industry, *edited by James P. Wood*
109. Good Manufacturing Practices for Pharmaceuticals: A Plan for Total Quality Control from Manufacturer to Consumer, Fifth Edition, Revised and Expanded, *Sidney H. Willig*
110. Advanced Pharmaceutical Solids, *Jens T. Carstensen*
111. Endotoxins: Pyrogens, LAL Testing, and Depyrogenation, Second Edition, Revised and Expanded, *Kevin L. Williams*
112. Pharmaceutical Process Engineering, *Anthony J. Hickey and David Ganderton*
113. Pharmacogenomics, *edited by Werner Kalow, Urs A. Meyer, and Rachel F. Tyndale*
114. Handbook of Drug Screening, *edited by Ramakrishna Seethala and Prabhavathi B. Fernandes*

115. Drug Targeting Technology: Physical • Chemical • Biological Methods, *edited by Hans Schreier*

116. Drug–Drug Interactions, *edited by A. David Rodrigues*

117. Handbook of Pharmaceutical Analysis, *edited by Lena Ohannesian and Anthony J. Streeter*

118. Pharmaceutical Process Scale-Up, *edited by Michael Levin*

119. Dermatological and Transdermal Formulations, *edited by Kenneth A. Walters*

120. Clinical Drug Trials and Tribulations: Second Edition, Revised and Expanded, *edited by Allen Cato, Lynda Sutton, and Allen Cato III*

121. Modern Pharmaceutics: Fourth Edition, Revised and Expanded, *edited by Gilbert S. Banker and Christopher T. Rhodes*

122. Surfactants and Polymers in Drug Delivery, *Martin Malmsten*

123. Transdermal Drug Delivery: Second Edition, Revised and Expanded, *edited by Richard H. Guy and Jonathan Hadgraft*

124. Good Laboratory Practice Regulations: Second Edition, Revised and Expanded, *edited by Sandy Weinberg*

125. Parenteral Quality Control: Sterility, Pyrogen, Particulate, and Package Integrity Testing: Third Edition, Revised and Expanded, *Michael J. Akers, Daniel S. Larrimore, and Dana Morton Guazzo*

126. Modified-Release Drug Delivery Technology, *edited by Michael J. Rathbone, Jonathan Hadgraft, and Michael S. Roberts*

127. Simulation for Designing Clinical Trials: A Pharmacokinetic-Pharmacodynamic Modeling Perspective, *edited by Hui C. Kimko and Stephen B. Duffull*

128. Affinity Capillary Electrophoresis in Pharmaceutics and Biopharmaceutics, *edited by Reinhard H. H. Neubert and Hans-Hermann Rüttinger*

129. Pharmaceutical Process Validation: An International Third Edition, Revised and Expanded, *edited by Robert A. Nash and Alfred H. Wachter*

130. Ophthalmic Drug Delivery Systems: Second Edition, Revised and Expanded, *edited by Ashim K. Mitra*

131. Pharmaceutical Gene Delivery Systems, *edited by Alain Rolland and Sean M. Sullivan*

132. Biomarkers in Clinical Drug Development, *edited by John C. Bloom and Robert A. Dean*

133. Pharmaceutical Extrusion Technology, *edited by Isaac Ghebre-Sellassie and Charles Martin*

134. Pharmaceutical Inhalation Aerosol Technology: Second Edition, Revised and Expanded, *edited by Anthony J. Hickey*

135. Pharmaceutical Statistics: Practical and Clinical Applications, Fourth Edition, *Sanford Bolton and Charles Bon*

136. Compliance Handbook for Pharmaceuticals, Medical Devices, and Biologics, *edited by Carmen Medina*

137. Freeze-Drying/Lyophilization of Pharmaceutical and Biological Products: Second Edition, Revised and Expanded, *edited by Louis Rey and Joan C. May*

138. Supercritical Fluid Technology for Drug Product Development, *edited by Peter York, Uday B. Kompella, and Boris Y. Shekunov*

139. New Drug Approval Process: Fourth Edition, Accelerating Global Registrations, *edited by Richard A. Guarino*

140. Microbial Contamination Control in Parenteral Manufacturing, *edited by Kevin L. Williams*

141. New Drug Development: Regulatory Paradigms for Clinical Pharmacology and Biopharmaceutics, *edited by Chandrahas G. Sahajwalla*

142. Microbial Contamination Control in the Pharmaceutical Industry, *edited by Luis Jimenez*

143. Generic Drug Product Development: Solid Oral Dosage Forms, *edited by Leon Shargel and Izzy Kanfer*

144. Introduction to the Pharmaceutical Regulatory Process, *edited by Ira R. Berry*

145. Drug Delivery to the Oral Cavity: Molecules to Market, *edited by Tapash K. Ghosh and William R. Pfister*

146. Good Design Practices for GMP Pharmaceutical Facilities, *edited by Andrew Signore and Terry Jacobs*

147. Drug Products for Clinical Trials, Second Edition, *edited by Donald Monkhouse, Charles Carney, and Jim Clark*

148. Polymeric Drug Delivery Systems, *edited by Glen S. Kwon*

149. Injectable Dispersed Systems: Formulation, Processing, and Performance, *edited by Diane J. Burgess*

150. Laboratory Auditing for Quality and Regulatory Compliance, *Donald Singer, Raluca-Ioana Stefan, and Jacobus van Staden*

151. Active Pharmaceutical Ingredients: Development, Manufacturing, and Regulation, *edited by Stanley Nusim*

152. Preclinical Drug Development, *edited by Mark C. Rogge and David R. Taft*

153. Pharmaceutical Stress Testing: Predicting Drug Degradation, *edited by Steven W. Baertschi*

154. Handbook of Pharmaceutical Granulation Technology: Second Edition, *edited by Dilip M. Parikh*

155. Percutaneous Absorption: Drugs–Cosmetics–Mechanisms–Methodology, Fourth Edition, *edited by Robert L. Bronaugh and Howard I. Maibach*

156. Pharmacogenomics: Second Edition, *edited by Werner Kalow, Urs A. Meyer and Rachel F. Tyndale*

Pharmacogenomics

Second Edition

Werner Kalow
University of Toronto
Toronto, Ontario, Canada

Urs A. Meyer
University of Basel
Basel, Switzerland

Rachel F. Tyndale
Centre for Addiction and Mental Health
University of Toronto
Toronto, Ontario, Canada

Taylor & Francis
Taylor & Francis Group
New York London

Published in 2005 by
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 1-57444-878-1 (Hardcover)
International Standard Book Number-13: 978-1-57444-878-8 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress

T&F informa

Taylor & Francis Group
is the Academic Division of T&F Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

Preface

The editors of *Pharmacogenomics* were pleased with the success of the book's first edition, indicated by substantial sales and by its translation into Japanese. We hope that the second edition creates equal satisfaction; thus, many descriptions of fundamental issues will be retained. However, since pharmacogenomics is a rapidly expanding science, seven new chapters have been added: Variability in Induction of Human Drug Metabolizing Enzymes, Pharmacogenetics and Cardiac Ion Channels, Regulatory Perspectives on Pharmacogenomics, Metabonomics, Haplotype Structure and Pharmacogenomics, Pharmacoepigenetics: From Basic Epigenetics to Therapeutic Applications, Genome Variation Influencing Gene Copy Number and Disease. In addition, the first edition chapters have been revised by new and previous authors. However, due to space constraints we have been unable to include several related topics such as genetic variation in transcriptional control and RNA editing. In short, the second edition is meant to be enriched but also to retain all the essentials of the first edition.

The editors can only hope that readers of the book, or of its parts, will be satisfied by our new efforts.

Werner Kalow
Urs A. Meyer
Rachel F. Tyndale

Contents

Contributors

Martin Beaulieu Sequenom Inc., San Diego, California, U.S.A.

Leif Bertilsson Division of Clinical Pharmacology, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

Paul C. Boutros Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada

Angela R. Brooks-Wilson Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

Paul Buckland Department of Psychological Medicine, University of Wales College of Medicine, Cardiff, U.K.

Lyle D. Burgoon Department of Pharmacology and Toxicology, Center for Integrative Toxicology, National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan, USA

Charles Cantor Sequenom Inc., San Diego, California, U.S.A.

Guy del Mistro Sequenom Inc., San Diego, California, U.S.A.

Mathias Ehrich Sequenom Inc., San Diego, California, U.S.A.

Xavier Estivill Program of Genes and Diseases, Center for Genomic Regulation, Barcelona, Spain

William E. Evans St. Jude Children's Research Hospital, Colleges of Pharmacy and Medicine, University of Tennessee, Memphis, Tennessee, U.S.A.

Lars Feuk Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada

James Flanagan The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada

David A. Flockhart Division of Clinical Pharmacology, Indiana University School of Medicine, Indianapolis, Indiana, U.S.A.

Denis M. Grant Department of Pharmacology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

Julius V. Halaschek-Wiener Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

D. M. Helmeste University of California, Irvine, California, U.S.A.

Christiane Honisch Sequenom Inc., San Diego, California, U.S.A.

Werner Kalow Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada

Richard B. Kim Division of Clinical Pharmacology, Vanderbilt University, Nashville, Tennessee, U.S.A.

Lois D. Lehman-McKeeman Discovery Toxicology, Bristol Myers Squibb Co., Princeton, New Jersey, U.S.A.

Bernard Lerer Biological Psychiatry Laboratory, Department of Psychiatry, Hadassah-Hebrew University Medical Center, Ein Karem, Jerusalem, Israel

Lawrence J. Lesko Food and Drug Administration, Rockville, Maryland, U.S.A.

Ulrika Liljedahl Department of Medical Sciences, Uppsala University, Uppsala, Sweden

Marco A. Marra Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

Catia Marzolini Division of Clinical Pharmacology, Vanderbilt University, Nashville, Tennessee, U.S.A.

Oliver G. McDonald Department of Molecular Physiology and Biological Physics, University of Virginia Health Sciences Center, Charlottesville, Virginia, U.S.A.

Urs A. Meyer Division of Pharmacology/Neurobiology, Biozentrum, University of Basel, Basel, Switzerland

Glenn A. Miller Genzyme Corporation, Westborough, Massachusetts, U.S.A.

Arno G. Motulsky University of Washington, Seattle, Washington, U.S.A.

Paul Oeth Sequenom Inc., San Diego, California, U.S.A.

Allan B. Okey Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada

Vural Ozdemir Biomarker and Clinical Pharmacology Unit, VA Long Beach Healthcare System, Southern California Institute for Research and Education and Department of Psychiatry and Human Behavior, College of Medicine, University of California, Irvine, California, U.S.A.

Georgios Panagiotidis Division of Clinical Pharmacology, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

Arturas Petronis The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada

Peter N. Ray Division of Molecular Genetics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

Dan M. Roden Oates Institute for Experimental Therapeutics, Vanderbilt University School of Medicine, Nashville, Tennessee, U.S.A.

Stephen W. Scherer Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada

Katherine A. Siminovitch Department of Medicine, University of Toronto, Toronto, Ontario, Canada

Tracy L. Stockley Division of Molecular Genetics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

Ann-Christine Syvänen Department of Medical Sciences, Uppsala University, Uppsala, Sweden

S. W. Tang University of Hong Kong and University of California, Irvine, California, U.S.A.

Glenys Thomson Department of Integrative Biology, University of California, Berkeley, California, U.S.A.

Rommel G. Tirona Division of Clinical Pharmacology, Vanderbilt University, Nashville, Tennessee, U.S.A.

Ana Maria Valdes Department of Integrative Biology, University of California, Berkeley, California, U.S.A.

Dirk van den Boom Sequenom Inc., San Diego, California, U.S.A.

Mu Wang Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana, U.S.A.

Wendell W. Weber Department of Pharmacology, University of Michigan, Ann Arbor, Michigan, U.S.A.

Richard F. Wintle Ellipsis Biotherapeutics Corporation, Toronto, Ontario, Canada

Frank A. Witzmann Department of Cellular and Integrative Physiology, Indiana University School of Medicine, Indianapolis, Indiana, U.S.A.

Janet Woodcock Food and Drug Administration, Rockville, Maryland, U.S.A.

Timothy R. Zacharewski Department of Biochemistry and Molecular Biology, Center for Integrative Toxicology, National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan, U.S.A.

1

Historical Aspects of Pharmacogenetics

Werner Kalow

Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada

1. INTRODUCTION

The occurrence of genetic influences upon one or other drug response was predicted by Sir Archibald Garrot in his 1931 book *Inborn Factors in Diseases* (1), and by J.B.S. Haldane in 1949 in an article entitled "Disease and Evolution" (2). Pharmacogenetics, as we know it today, arose as a new scientific entity in the late 1950s as a marriage of the older sciences of Pharmacology and Genetics.

Pharmacogenetics deals with heredity and the effect of drugs. It is a branch of science devoted to efforts of explaining variability of one or other drug response, and to search for the genetic basis of such variations or differences. It started by looking at differences between individual subjects, but as it developed, it also became concerned with genetic differences between populations. Many pharmacogeneticists happen to be mostly concerned with the human species but the science applies in principle to all living subjects on earth, primitive or complex, capable of responding to a drug or to a toxic chemical.

Of many genetic responses to environmental impacts, Pharmacogenetics is only one (3). Human variation in pharmacogenetics is similar to human variation in response to foods (4). For instance, modern salt intake causes members of populations who come from salt-poor areas to develop cardio-vascular disease (5). Populations adjusted to frequent periods of starvation tend to show a high incidence of type 2 diabetes (6). There are different genetic mechanisms to fight infections. There is a gene conveying

resistance to tuberculosis, acting before any immune response sets in (7). The mechanics of AIDS differ between Caucasians and Africans (8). Thus, pharmacogenetics is not a unique affair, but let us still look at its development.

2. THE INITIAL PHASE OF PHARMACOGENETICS

2.1. My Unexpected Dive into Pharmacogenetics: A Personal Story

In Berlin in 1948, there were still incidences of malnutrition. Because of this, there were patients who suffered fatal poisoning from the generally safe, local anaestetic drug procaine. This became my impetus to study the esterase that hydrolysed procaine (9). When invited to Philadelphia, I continued these studies with the superior equipment there available to me. I found that the procaine-splitting esterase was butyrylcholinesterase, then called pseudo- or plasma-cholinesterase, and I explored a method using UV-spectrophotometry which elegantly and precisely indicated the esterase activity (10).

I then transferred to Toronto where pseudo-cholinesterase had been discovered, and where it was still being investigated. I proposed the use of my new UV-method to replace the tedious gasometric method then in place. My proposal seemed acceptable to the responsible biochemist, provided I could demonstrate the efficiency of my method by comparing and testing plasma from patients with known high and low cholinesterase activity. Thus I came to test the esterase of patients known to show abnormal effects of succinylcholine, whose esterase had been designated by a government laboratory as having low activity. I was surprised to see that the cholinesterase activity was not low but that it displayed abnormal kinetics with grossly reduced affinity for its substrate, and thus appeared to be low. This could be explained only by an abnormal enzyme structure, and that could only be genetic; I could prove that point by family investigations (11). I was excited by this observation of interplay between genetics and the abnormal effect of a drug. I tried to find out whether there were other established examples.

My literature search was successful. In 1937 Waldenstrom's story on porphyria was published (12). A major find was a report on hemolysis caused by the antimalarial drug primaquine in some American soldiers in the 1940s (13), later shown to be due to glucose-6-phosphate dehydrogenase (G6PD) deficiency. I became excited by a 1957 paper of Arno Motulsky (14), a paper entitled "Drug Reactions, Enzymes, and Biochemical Genetics," a paper sponsored by the Council on Drugs of the American Medical Association. I found a report describing genetic differences of the metabolism of isoniazid (15), a then revolutionary antituberculosis drug. These and several other reports encouraged me to write a book on this new topic of Pharmacogenetics (16), citing all examples which I was able to find. Pharmacogenetics had become my scientific life blood.

2.2. The Growth of Pharmacogenetics

Many pertinent observations came in the following years. For instance, Vesell and Page (17) used twin studies to show genetic control of the metabolism of several drugs. von Wartburg and Schurch (18) described a variant of alcohol dehydrogenase. However, a report on deficient metabolism of debrisoquine became a milestone in pharmacogenetics (19); Dr. Robert Smith in London (20) had personally experienced the deficiency as a life-threatening drop of blood pressure after taking the drug experimentally. Before that, Eichelbaum (21) had reported in a thesis a metabolic deficiency of sparteine metabolism; both defects are now known to be due to deficiency of the P450 cytochrome CYP2D6. This deficiency affects the metabolism of more than 50 drugs; whether the deficiency is clinically important for any given drug depends on a number of drug-associated criteria and safety factors (22). More than 70 different variants of CYP2D6 are known, many of these are without any trace of activity (23). On the other hand, gene duplication or multiplication in some subjects cause extremely high CYP2D6 activity (24). The medical impact of these variations is indicated by a recent medline search which recorded 2000 papers dealing with CYP2D6 variation. The CYP2D6 is the most variable of human CYPs, but its variability is not unique: a list of 11 different drug-metabolizing CYPs (excluding CYP2D6) named 168 alleles, many carrying several mutations, causing 97 protein changes (25).

It is not surprising, that most initial discoveries in pharmacogenetics pertained to drug-metabolizing enzymes; measurements of drugs and drug metabolites required chemical analytical methods of more or less traditional nature. Investigation of variation of drug receptors usually requires knowledge of the receptor's DNA sequence, so that deviations can be discovered by testing the receptor genes in white blood cells of different people (26). Thus, most studies of the pharmacogenetics of receptors, or similarly of ion channels and of transporters, have recent origins. Much of these topics will be covered in subsequent chapters.

It is only now that we can appreciate the magnitude of genetic variation. Let us look at a few numbers: the human genome contains about 3 billion base pairs, and single-base variations (called SNPs, for single nucleotide polymorphisms) are on the average as frequent as 1/1000 bases (27). This means that many human proteins show genetic variation. There are known, e.g., more than 100 cancer-promoting oncogenes and about three dozen cancer-suppressing genes, and their functions may be controlled by genes determining DNA repair, cell division, metabolism, immune responses, embryonic development, and cell migration (28). Thus, the number of opportunities, and the magnitude of human genetic variation explain that there is what we call pharmacogenetics.

One should not forget that many environmental factors are able to influence drug responses. Some of these, including drugs, produce this effect by altering gene expressions. One pharmacological effect has been known for many years, when it was realized that certain drugs, besides foods, and hormones, may induce formation of drug-metabolizing enzymes. Thus, some drugs are able to change gene function! Disregarding tradition, one might consider these drug-caused alterations of gene expression as an aspect of pharmacogenetics. At the present time, studies of gene expression strongly affect the science of genetics. Parts of this topic will be covered in special chapters.

3. PHARMACOGENETICS AND POPULATIONS

Pharmacogenetics is largely still considered a story of person-to-person differences in drug metabolism and response. A broader view becomes effective when we look at simple organisms. To appreciate the fact that pharmacogenetic variation can be a protective commodity for a population, let us consider insect-resistance to insecticides (3), or bacterial resistance to antibiotics (29). Pharmacogenetic resistance of an individual insect to the killing effects of an insecticide causes this individual to survive an exposure, so that the offspring of that insect can multiply and in the long run create the resistant strain. Bacterial resistance to antibiotics represents the same mechanism. We cannot see this dramatic effect of pharmacogenetics in people because environmental hazards are usually not so directly killing, and the human-generation time is too long. The initial emphasis upon differences between individuals is one-sided; pharmacogenetic diversities also characterize different human populations (30,31).

From reading the literature, I knew early on of some pharmacogenetic differences between human populations (16). However, let me tell the story how the importance of such differences was driven home to me. Toronto in Canada became more and more often the home of immigrants from China. In the early 1970s, there were a few Chinese in the class of 140 medical students at the University of Toronto. My colleagues and I were studying at the time the metabolism of the then frequently used drug amobarbital, a barbiturate, and we observed a family with impaired amobarbital metabolism (32). In order to find out whether this was a frequent or rare deficiency, we asked the medical students to volunteer for an amobarbital study. After we had our laboratory results, I noticed that seven data did not fit to the rest. I suspected an error of measurement and wanted to repeat the study of these subjects. I only knew the student numbers and asked a colleague for their names; after a while, he came back—visibly shaken—that all the student numbers came from student with Chinese names. Further investigations confirmed that one of the metabolic alterations of amobarbital was on the average distinctly faster in Chinese than in Caucasian students (33).

The deficiency of debrisoquine metabolism (19) was also tested in our laboratory, and we found a different metabolic ratio between Chinese and Caucasian students (34). These observations, together with the old G6PD and NAT2 data and some additional comparisons, firmly planted in my mind the idea that differences in drug metabolism are not only a matter of individuals but frequently occur also between the human populations. I published a review article that probably was the first exclusively concerned with interethnic differences of drug metabolism (35). Knowledge of such differences has become very important for the pharmaceutical industry.

4. MONOGENIC AND MULTIGENIC VARIATIONS OF DRUG RESPONSES

The occurrence of the response to a drug that differs between persons can have many different causes, for instance, variability of drug metabolism as indicated above. Other potential differences may be caused by variation of drug targets or receptors, or of the transporters of drugs that operate at sites of absorption, of the blood–brain barrier, or of cellular membranes in general.

Dealing with variation of specific genes is a relatively simple affair, and so far has characterized most aspects of pharmacogenetics. However, we cannot neglect the fact that most differences between people are due to differences between many genes, in addition to environmental influences. Pharmacology became a science only after the ever-present differences between subjects were recognized, and the concept formalized by Trevan's introduction of the term "ED50" (36), indicating the dose sufficient to produce a given effect in 50% of a tested population. Whether somebody belongs to the 50% needing a lower or a higher dose may depend on many factors, including drug absorption, volume of drug distribution, drug transport, blood flow, target reaction, metabolic destruction, and elimination via kidney, bile, or gut. All may contribute to a difference between two people. Every one of these factors may depend on one or more products. Multigenic variation is important.

Let us consider a single reaction, the metabolic alteration of a drug: the metabolism may fail because of a genetic change of the structure of the responsible enzyme, as discussed above. However, it may also fail because not enough enzyme was formed, perhaps due to a failure of transcription or translation. Was there the absence of an inducer or regulator, perhaps a hormone, not formed or too fast degraded? Perhaps a genetic abnormality of the promoting region prevented the normal response to the inducer. Perhaps the drug could not reach the enzyme because of too much albumin binding, or because of a transporter deficiency. Thus, most genetic differences between people are complex and many genes contribute.

Because of the complexity, the existence of inter-individual differences is taken for granted and the causes are usually ignored. However, multifactorial differences between population deserve some attention. As an example, let us consider a comparison between Swedish and Chinese populations and their capacity to metabolize codeine (37). The drug undergoes three primary metabolic reactions: Glucuronidation, *N*-demethylation to form nor-codeine, and *O*-demethylation to form morphine. All these reactions differ between the two populations. The slow morphine formation in Chinese reflects the known ethnic variation of CYP2D6, but no single enzyme change is known to account for the metabolic differences in glucuronidation and in nor-codeine formation. Particularly the glucuronidation curves show a normal distribution in both populations, suggesting multifactorial variation of potential medical significance. This raises the question how one should deal mathematically with multifactorial differences between population.

One answer is illustrated in Figure 1. Assure that the clearance rate of a drug is measured in two populations; the average clearance differs somewhat, in this illustration by one standard deviation. For practical purposes, an average difference of that size can usually be neglected; due to the normal scatter of the data, the values for most members of the population are overlapping.

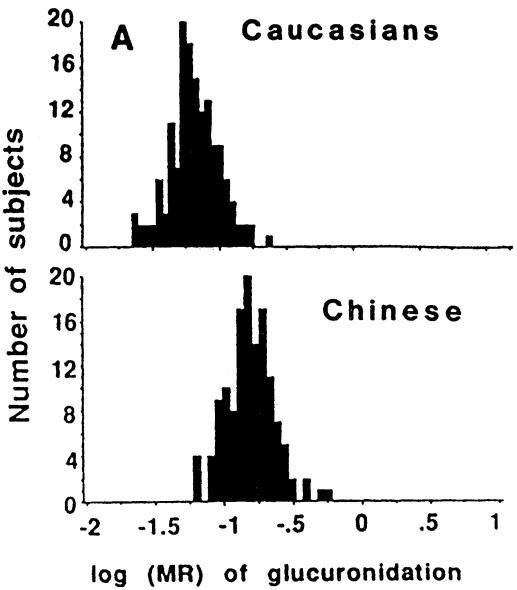


Figure 1 Frequencies of codeine glucuronidation in Caucasian and Chinese populations. The data are shown in terms of the logarithms of the metabolic ratios in 149 Caucasians and 133 Chinese. *Source*: From Ref. 35.

Important is what I like to call the "edge effect" (38,39). Slow clearance delays drug removal, or causes higher blood levels of the drug on repeated administration, and thereby tends to cause drug toxicity. Thus in population B, a much larger percentage of subjects would get adverse drug effects than in population A (Fig. 2). Population averages would not convey this information, only consideration of the edges of the distribution curves would do so.

A problem with multifactorial variation is the question to what extent is it determined by heritable factors, and to what extent by environmental determinants; both are probably contributing to the variation. The answer may guide investigations of the problem: scientific inquiries may be directed to primarily look at genes or to search for environmental influences. Traditionally, twin studies were used to determine the heritability of any human variant. Since drug effects come and go, it is possible in pharmacology to avoid twins, to collect a group of people, and to give each subject a drug two or more times; this will allow a statistical comparison of inter- and intra-subject variation. The comparison can be used to calculate the genetic component contributing to any pharmacological variation (40–42).

5. ECOGENETICS AND PHARMACOGENOMICS

Observations of inter-individual differences in metabolism of drugs, and therefore in different drug responses, led to the development of pharmacogenetics. However, it was not before long that investigators without a particular interest in drugs noted similar differences in response to environmental toxicants. Thus, the term "Ecogenetics" was coined by Brewer in 1971. Brewer asked whether geneticists with their exploding science were sufficiently concerned with humans "facing an environmental crisis of such proportions that our very existence is threatened." The term was taken up and used when concerned with genetic differences in the tolerance of food items (4), such as e.g., of lactose in milk products. Calabrese (44) was much concerned with occupational diseases, and wrote a book about ecogenetics. The World Health Organization arranged in 1989 a meeting on ecogenetics, which led to a subsequent book (45). Thus, ecogenetics is firmly established as a term and a special branch of science. The principal concepts embodied by the terms pharmacogenetics and ecogenetics are indistinguishable.*

The word "Pharmacogenomics" (46,47) reflects, in the first place, the change of the human technical ability to investigate the genome, pinpointing

* There is an underlying linguistic problem: the word "pharmakon" in ancient Greek refers to both drugs and poisonous substances. Therefore, the term "pharmacology" means for many medical scientists a topic dealing with both therapeutic and toxic agents. For others, pharmacology invokes thoughts of drugs as therapeutic medicines, distributed by pharmacists. Thus, some geneticists use the term ecogenetics, while the term seemed redundant to many pharmacologists.

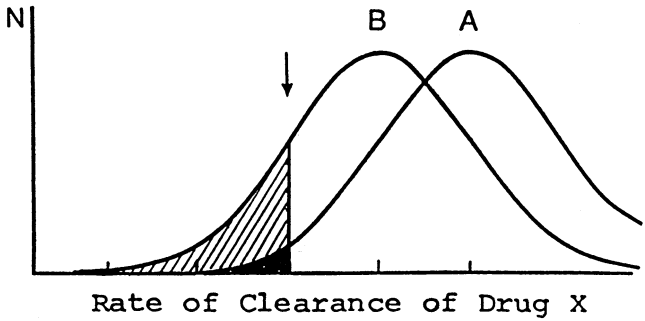


Figure 2 The edge effect of pharmacological distributions. The hypothetical curves A and B represent normal frequency distributions for the elimination capacity of drug X in two populations; the abscissa registers the rate of drug clearance, the ordinate indicates the number of subjects who show a particular rate of clearance. The arrow indicates the critical rate below which the drug causes toxicity. Curves A and B have identical standard deviations but their means are one standard deviation apart. That is, the difference between the means is small compared with the range of variation within each population. The arrow is located two and one standard deviations below the means of A and B, respectively. These data imply that approximately 2% of population A (*black area*) and 16% of population B (*shaded plus black area*) would suffer toxicity from the standard dose of drug X. The eight fold difference would grow substantially with the shift of the arrow towards lower values on the abscissa (from Ref. 37). When only the population means or averages are considered, the difference may register as "statistically not significant," thereby giving a clinically misleading message. *Source*: From Ref. 37.

multiple variations, interactions, and functions of genes (48,49), rather than merely single genes. This is an important change: it is now realized that only a few percent of the human genome are the standard protein-producing genes; functions of the total genome are being explored by comparing the genomes of different species (50). From a medical point of view, it is mainly three aspects that will make pharmacogenetics and pharmacogenomics different subjects:

1. Use of genomic methods will improve and broaden genotyping. Phenotyping methods will have fewer diagnostic uses, but they will remain important as a means to assess the medical significance of a genetic variation.
2. While Pharmacogenetics was mostly concerned with structural variation of genes by mutation, pharmacogenomics will in addition deal with drugs acting on genes, thereby changing gene expression.
3. The use of genomic methods should bring us new drug targets. Such could be variant human genes or their proteins that promote the occurrence of common diseases, that is, diseases like blood pressure elevation, asthma, or schizophrenia. By the same token,

genomic looks at infectious germs might reveal spots in them that are crucial for their survival and which might be attacked by new antibiotics.

6. CONCLUSIONS

In summary, pharmacogenomics will in the long run lead to a better understanding of the interaction between drugs and gene products. The promise of pharmacogenomics is that the choice of drug to combat a disease will be determined more and more by which gene or genes contribute to the disease in a given subject; in other words, we can expect to see the development of individualized, gene-dependend drug therapy.

Pharmacogenetics was historically most concerned with drug safety which will remain a concern, but the main effect of pharmacogenomics promises to be an improvement of drug efficacy.

REFERENCES

1. Garrod AE. Inborn factors in disease: an essay. New York: Oxford University, Press, 1931.
2. Haldane JBS. Disease and evolution. Ric Sci 1949; 19:68–75.
3. Kalow W. Pharmacogenetics in biological perspective. Pharmacol Rev 1997; 49:369–379.
4. Motulsky AG. Pharmacogenetics and ecogenetics in 1991. Pharmacogenetics 1991; 1:2–3.
5. Grim CE, Robinson M. Blood pressure variation in blacks: genetic factors. Semin Nephrol 1996; 16:83–93.
6. Neel JV. The "thrifty genotype" in 1998. Nutr Rev 1998; 57:S2–S9.
7. Skamene E. The Bcg gene story. Immunobiology 1994; 191:451–460.
8. Dean M, Jacobson LP, McFarlane G, Margolick JB, Jenkins FJ, Howard OM, Dong HF, Goedert JJ, Buchbinder S, Gomperts E, Vlahov D, Oppenheim JJ, O'Brien SJ, Carington M. Reduced risk of AIDS lymphoma in individuals heterozygous for the CCR5–delta32 mutation. Cancer Res 1999; 1:3561–3564.
9. Herken H, Kalow W. Photometometrische Bestimmung der Enzymatischen Novocain-Hydrolyse. Klin Wochenschr 1951; 29:90–91.
10. Kalow W. Hydrolysis of local anesthetics by human serum cholinsterase. J Pharmacol Exp Ther 1952; 104:122–134.
11. Kalow W. Familial incidence of low pseudocholinesterase level. Lancet 1956; 2:576–577.
12. Waldenstrom J. Studien uber Porphyrie. Acta Med Scand 1937; 82:1–254.
13. Dern RJ, Beutler E, Alving AS. The hemolytic effect of primaquine. J Lab Clin Med 1954; 44:171–176.
14. Motulsky AG. Drug reactions, enzymes, and biochemical genetics. JAMA 1957; 165:835–837.
15. Bonicke R, Lisboa BP. Uber die Erbeddingtheit der intraindividuellen Konstanz der Isoniazidausscheidung beim Menschen. Naturwissenschaften 1957; 44:314.

16. Kalow W. Pharmacogenetics. Heredity and the Response to Drugs. Philadelphia: W.B. Saunders, 1962.
17. Vesell ES, Page JG. Genetic control of dicumerol levels in man. J Clin Invest 1968; 47:2657–2663.
18. von Wartburg JP, Schurch PM. Atypical human liver alcohol dehydrogenase. Ann NY Acad Sci 1968; 151:936–946.
19. Mahgoub A, Dring LG, Idle JR, Lancaster R, Smith RL. Polymorphic hydroxylation of debrisoquine in man. Lancet 1977; 2:584–586.
20. Smith RL. Introduction: human genetic variations in oxidative drug metabolism. Xenobiotica 1986; 16:361–365.
21. Eichelbaum M. Ein neuentdeckter Defekt im Arzneimittel–stoffwechsel des Menschen: Die fehlende N–Oxidation des Spartein. Bonn: Medizinische Fakultat Rheinischen Friedrich Wilhelms–Universitat, 1975.
22. Meyer UA, Zanger UM. Molecular mechanisms of genetic polymorphisms of drug metabolism. Annu Rev Pharmacol Toxicol 1997; 37:269–296.
23. Marez D, Legrand M, Sabbagh N, Lo Guidice JM, Spire C, Lafitte JJ, Meyer UA, Broly F. Polymorphiism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles, their frequencies and evolution. Pharmacogenetics 1997; 7:193–202.
24. Dahl ML, Johansson I, Bertilsson L, Ingelman-Sundberg M, Sjoqvist F. Ultrarapid hydroxylation of debrisoquine in a Swedish population. Analysis of the molecular genetic basis. J Pharmacol Exp Ther 1995; 274:516–520.
25. Cascorbi I. Pharmacogenetics of cytochrome p4502D6: genetic background and clinical implication. Eur J Clin Invest 2003; 13(suppl 2):17–22.
26. Propping P, Nothen MM. Genetic variation of CNS receptors—a new perspective for pharmacogenetics. Pharmacogenetics 1995; 5:318–325.
27. Zhao Z, Fu YX, Hewett-Emmet D, Boerwinkle E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene 2003; 312:207–213.
28. Garber K. More SNPs on the way: genome research. Science 1998; 281:1788.
29. Bennett PM. The spread of drug resistance. In: Baumberg S, Young JPW, Wellington EMH, Saunders JR, eds. Population Genetics of Bacteria. Cambridge: Cambridge University Press, 1995:317–344.
30. Kalow W. Pharmacogenetics and evolution. Pharmacogenetics 2000; 10:1–3.
31. Kalow W, Tang BK, Kadar D, Endrenyi L, Chan F-Y. A method to study drug metabolism in populations: racial differences in amobarbital metabolism. Clin Pharmacol Ther 1979; 26:766–776.
32. Kalow W, Otton SV, Kadar D, Endrenyi L, Inaba T. Ethnic difference in drug metabolism: debrisoquine 4-hydroxylation in Caucasians and Orientals. Can J Physiol Pharmacol 1980; 58:1142–1144.
33. Kalow W. Ethnic differences in drug metabolism. Clin Pharmacokinet 1982; 7:373–400.
34. Trevan JW. The error of determination of toxicity. R Soc Lond Proc 1 Ser B 1927; 101:483–514.
35. Yue Q, Sawe K. Interindividual and interethnic differences in codeine metabolism. In: Kalow W, ed. Pharmacogenetics of Drug Metabolism. New York: Pergamon Press, 1992:721–727.

36. Kalow W. Pharmacoanthropology and the genetics of drug metabolism. In: Kalow W, ed. Pharmacogenetics of Drug Metabolism. New York: Pergamon Press, 1992:865–877.
37. Kalow W, Bertilsson L. Interethnic factors affecting drug response. Adv Drug Res 1994; 25:1–59.
38. Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. Pharmacogenetics 1998; 8:283–289.
39. Kalow W, Ozdemir V, Tang BK, Tothfalusi L, Endrenyi L. The science of pharmacological variability: an essay. Clin Pharmacol Ther 1999; 66: 445–447.
40. Ozdemir V, Kalow W, Tang BK, Paterson AD, Walker SE, Endrenyi L, Kashuba ADM. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration (RDA) method. Pharmacogenetics 2000; 10:373–388.
41. Brewer GJ. Annotation: human ecology, an expanding role for the human geneticist. Am J Hum Genet 1971; 23:92–94.
42. Calabrese EJ. Ecogenetics: Genetic Variation in Susceptibility to Environmental Agents. New York: John Wiley & Sons, 1984.
43. Grandjean P, Kello D, Rohrborn G, Tarkowski S. Ecogenetics. Genetic Predisposition to the Toxic Effects of Chemicals. New York: World Health Organization/Chapman & Hall, 1991.
44. Grant DM. Pharmacogenomics and the changing face of clinical pharmacology. Can J Clin Pharmacol 1999; 6:131–132.
45. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999; 15:487–491.
46. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. New goals for the U.S. human genome project: 1998–2003. Science 1998; 23: 682–689.
47. Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin D, Sali A, Studier FW, Swaminathan S. Structural genomics: beyond the human genome project. Nat Genet 1999; 23:151–157.
48. O'Brien SJ, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, Stanyon R, Copeland NG, Jenkins NA, Womack JE, Marshall Graves JA. The promise of comparative genomics in mammals. Science 1999; 15: 458–481.
49. Kennedy GC, Matsuzaki H, Dong S, 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 MS, Boice-Jacino MT, Fodor SP, Jones KW. Large-scale genotyping of complex DNA. Nat Biotechnol 2003; 21:1233–1237.
50. Zhou XJ, Gibson G. Cross-species comparison of genome-wide expression patterns. Genome Biol 2004; 5:232–234.

2

Pharmacogenomics and the Promise of Personalized Medicine

Vural Ozdemir

Biomarker and Clinical Pharmacology Unit, VA Long Beach Healthcare System, Southern California Institute for Research and Education and Department of Psychiatry and Human Behavior, College of Medicine, University of California, Irvine, California, U.S.A.

Bernard Lerer

Biological Psychiatry Laboratory, Department of Psychiatry, Hadassah-Hebrew University Medical Center, Ein Karem, Jerusalem, Israel

1. INTRODUCTION

The concept of individualized drug therapy has been a central focus in clinical pharmacology and medicine for many decades (1). In practice, there are two key elements concerning customized therapies: choice of the dosing regimen and/or the drug itself. Both of these factors contribute to personalized medicine. Despite its apparent simplicity, individualization of treatment is a daunting challenge due to molecular heterogeneity in human diseases and marked patient-to-patient and between-population differences in drug effects as well as therapeutic dose requirements.

Therapeutics as a science relies on the degree of predictability of drug effects and the mechanisms governing their predictability (2). A clear understanding of the factors that influence dose–response and time-course of drug action allows rational choice of drugs and their dosages (3). There are also degrees of personalized medicine. The resolution of customization may vary,

on the one hand, from drugs that are targeted for a very small group of individuals, to those that are intended for use in most members of the population without regard to individual characteristics of the patients. The latter group is also known as blockbuster drugs and represents majority of the medications that are currently in clinical use. Thus, the notion of personalized medicine reflects a fundamental conceptual departure from the traditional lore of pharmacotherapy that asserts the use of pharmaceuticals uniformly in broad patient populations, rather than smaller subpopulations wherein drugs may exhibit enhanced efficacy and optimal safety (Fig. 1).

Interest in personalized medicine and rational choice of therapies can be traced back to the first formal comparative trial in the 18th century. James Lind, a naval surgeon, demonstrated in 1747 that citrus juice, and not the other leading remedies recommended by physicians of the day, cured scurvy (2). At the turn of the 20th century the British physician, Archibald Garrod wrote presciently on the topic of chemical individuality (4). More recently in the second half of the 20th century, the impetus for customized pharmacotherapy was fueled by adverse drug reactions (ADRs). The early reports of aplastic anemia in patients exposed to chloramphenicol, followed by the thalidomide disaster in 1961 led to recognition of at risk populations, or conditions (e.g., renal failure) and drug–drug interactions that may

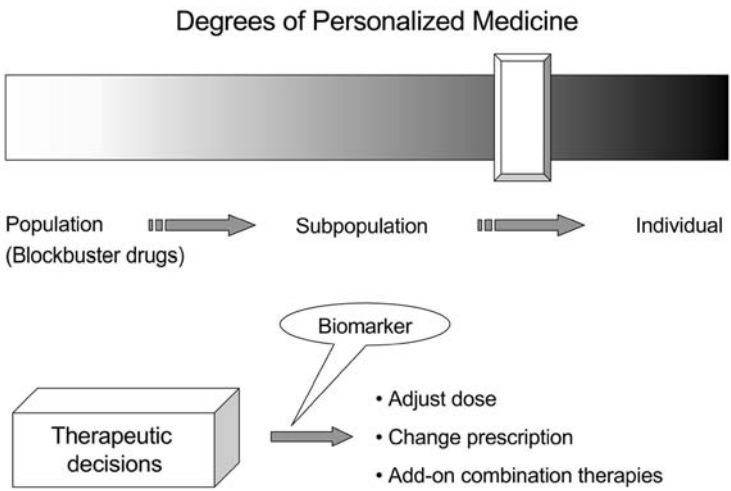


Figure 1 The degree of customization of drug therapy ranging from nil (blockbuster drugs—targeted for the entire population) to those highly personalized for an individual patient. A realistic degree of customization, at the level of subpopulations, is depicted by the vertical column on this spectrum. The three decisions pertinent to customization of drug therapy are shown below. Biomarkers are predictive tests that guide the latter therapeutic decisions (see also Sec. 3.2).

predispose to drug toxicity. These developments and the ensuing scientific scrutiny culminated in the publication, for the first time, of formal principles pertaining to individualization of drug treatment based on disease, genetic, or environmental chemical influences (2,5).

Why is there a strong emphasis on personalized medicine *now*? Interindividual variability in drug effects and the lack of reliable predictors of this variability are increasingly recognized as important barriers to science-based therapeutics (6–8). Moreover, the public health consequences of uncertainty in drug efficacy and safety have been increasingly documented by recent pharmacovigilance studies (9–13). A number of policy initiatives, particularly those on ADRs, further added to the momentum for personalized therapies in health care (10,14). Notably, present genetics-based efforts to individualize drug therapy have a more *mechanistic* focus in contrast to the previous empirical initiatives on personalized medicine using less precise demographic and descriptive clinical characteristics of the patients or the attendant disease (6,15,16). Hence, a common thread that runs through current pharmacogenomic strategies aimed at personalized medicine is the intent to relate the descriptive results of pharmaceutical interventions or patient characteristics to molecular and mechanistic underpinnings of drug response that, by extension, may allow more precise and rational predictions on clinical outcomes. The advances made by completion of the Human Genome Project (HGP) 4 years ago and the high throughput genomic technologies that spun off from this effort now make it entirely feasible to apply this vast knowledgebase in the clinical practice and diagnosis of human diseases as well as drug efficacy and safety in each individual patient (17). The HGP presents unique research opportunities to identify novel drug targets for common complex diseases. This offers the promise in the near future for improvements not only in drug safety but also therapeutic efficacy by selective prescriptions in subpopulations identified by genetic testing of drug targets (18). The individualization of drug therapy is thus evolving from traditional dose titration methods to more radical approaches concerning prescription decisions and the choice of drugs based on individuals' genetic make-up (19). These changes, understandably, are attracting much attention from all stakeholders including patients, physicians, academic investigators, pharmaceutical industry, insurers and experts in regulatory science. Collectively, these recent developments have placed the study of human genetics and personalized medicine firmly on the social policy agenda, raising a vast amount of public interest as well as close scrutiny of its promises and actual impact on patient care (20–24).

Pharmacogenomics is a term introduced in late 1990s and is broadly defined as the study of variability in drug safety and efficacy using information from the entire genome of a patient. Variations in both gene sequence and expression are of interest to pharmacogenomic inquiries. By contrast, the term pharmacogenetics has been established since 1950s and refers to investigations on specific candidate genes in relation to individual differences

in drug effects. Candidate genes in pharmacogenetic studies are selected based on a priori observations of disease susceptibility, drug absorption, metabolism, transport, and excretion as well as drug targets, as opposed to the genome-wide hypothesis-free approach in pharmacogenomics. Despite these differences, there is also interdependency between the two disciplines (16). Once the genes or genetic markers relevant to mechanism of drug action or safety are identified through the genome-wide pharmacogenomics search, each individual gene requires further clinical validation by focused and hypothesis-driven pharmacogenetic approaches before they can be routinely applied at point of care in the clinic. We herein chose to use the term pharmacogenomics but many of the ensuing discussion and concepts will also be applicable to pharmacogenetics.

The aim of the present chapter is to introduce the reader to (1) broad pharmacological, genetic and societal drivers as well as the promise of personalized medicine, (2) specific pharmacogenomic strategies to individualize drug therapy early in drug discovery, clinical development and during routine therapy at point of care in the clinic, and (3) conceptual and practical barriers to pharmacogenomic-guided personalized medicine and the broader ethical issues associated with customized therapies.

2. DRIVERS AND THE PROMISE OF PERSONALIZED MEDICINE

2.1. Rationale for Customized Drug Therapy: Variability and the Present State of Drug Safety and Efficacy in the Clinic

Most current medications and the recommended dosing regimens come with a significant risk of drug toxicity or treatment-failure in the clinic (9,11). Drug-related morbidity and mortality were estimated to cost up to $137 billion annually in the United States (25). A recent meta-analysis of prospective studies in the United States suggests that serious and fatal ADRs occur in 6.7% and 0.32% of hospitalized patients, respectively (11). This translates into more than two million serious ADRs and an annual death rate of 106,000 patients in the United States alone. These estimates rank ADRs as the fourth leading cause of death, ahead of accidents, diabetes and pneumonia (11). Subsequent extended analysis of pharmacoepidemiology data in 32 non-U.S. studies from industrialized countries support the contention that fatal ADRs are a significant public health problem in many countries around the world (12). It is noteworthy that the serious ADRs noted above were observed during treatment with usual doses of drugs that had already been introduced for clinical use and despite the exclusion of cases due to intentional or accidental overdose, human errors in drug administration or non-compliance, drug abuse, and therapeutic failures (11,12). Similarly, an independent study of 2227 ADRs in hospitalized patients showed that about 50% had no readily discernible or preventable cause (13). Nearly

16% (193) of the 1232 pharmaceutical products listed in the Physicians' Desk Reference in the U.S. were deemed to carry a significant ADR risk to warrant a "black box" warning on the drug label (9,26). The societal and global burden of ADRs can be compounded further by consideration of drug morbidity and mortality in ambulatory settings and nursing homes as well as non-industrialized countries. Collectively, these epidemiological observations suggest that more fundamental and previously unaccounted reasons, possibly genetic in nature, may underlie a significant portion of such unpreventable and apparently idiosyncratic ADRs in the clinic. Consistent with this hypothesis, a detailed analysis of 27 drugs frequently cited in ADR studies found that 59% are metabolized by one or more enzyme with a variant allele associated with deficient metabolism (10). By contrast, only 7–22% of drugs selected at random were influenced by a genetically polymorphic metabolic pathway (10).

Although drug safety has traditionally received much research and media attention, drug efficacy is an equally important and yet, often overlooked dimension of pharmacotherapy. In comparison to ADRs, the public health consequences and economic costs of therapeutic-failure associated with drugs have not been well investigated. A recent review of published data on the efficacy of major drugs used in several important diseases is instructive in this regard (9,26). Spear et al. concluded that the response rates vary markedly across various therapeutic areas with 80% of patients responding to Cox-2 inhibitors while the response rate was as low as 25% in cancer chemotherapy and 30% in Alzheimer's disease (Table 1). Overall, it appears that only about 50% of patients respond to drugs in major therapeutic classes (or conversely, 50% of patients, on average, do not respond to pharmacotherapy) (9).

While drugs are often life-saving in some patients, the existing armamentarium of drugs, taken together, are only moderately effective in the general population and carry significant liabilities in the form of serious and fatal ADRs. The traditional notion of blockbuster drugs overlooks this interindividual variability in drug efficacy and safety despite its adverse public health consequences and economic burden on society at large. In fact, one may argue, in light of the pharmacovigilance data noted above, that there are really very few robust examples of blockbuster drugs in the clinic. The term "blockbuster drug" has more economic underpinnings and relates to the commercial promise of a medication. Unfortunately, this term is often misused in a scientific context to refer to a broad array of drugs that are assumed to work in most members of the population with optimal safety whereas, in essence, it indicates the pharmaceutical products that are developed and marketed with the general population in mind. The predicaments associated with safety and efficacy of the existing medications clearly call for more focused science-based therapeutics and rational approaches to selection of drugs and their dosing regimens to customize pharmacotherapy based on, for example, individual patient genetic make-up. Understanding the mechanisms and cause–effect

Table 1 Response Rates of Patients to Major Drug Classes in Selected Therapeutic Areas

Therapeutic area	Efficacy rate (%)
Alzheimer's	30
Analgesics (Cox-2)	80
Asthma	60
Cardiac arrhythmias	60
Depression (SSRI)	62
Diabetes	57
HIV	47
Incontinence	40
Migraine (acute)	52
Migraine (prophylaxis)	50
Oncology	25
Osteoporosis	48
Rheumatoid arthritis	50
Schizophrenia	60

Source: From Ref. 9.

relationships for apparently unpreventable and idiosyncratic ADRs and treatment-failures is crucial for early identification and prevention of such drug related problems in the 21st century healthcare.

2.2. Rationale for Genetics: Can It Explain and Predict Variability in Drug Efficacy and Safety?

A prerequisite implicit assumption for any pharmacogenomic study is that the targeted pharmacological trait (or phenotype) is subject to appreciable genetic control (27). To this end, it is important to recognize that drug effects are usually elicited against the background of disease phenomena. Many of the human diseases display genetic components with varying degrees. Further, some of the biological pathways underlying diseases may serve as targets for drug interventions or alternatively, hold the potential to modify, or counteract the direct pharmacological effects of drugs via homeostatic mechanisms. Thus, genetic regulation of diseases or physiological pathways may indirectly influence variability in drug effects.

More direct evidence for the role of heredity in pharmacology can be observed in pharmacokinetic processes. The origins of pharmacogenomics date to the 1950s when monogenic variations in drug metabolism were the primary focus of research interest. Early on, a number of seminal twin studies in healthy volunteers under uniform basal environmental conditions demonstrated a markedly higher reproducibility of pharmacokinetic indices in monozygotic twins (nearly 100% identity in the genome) compared to

dizygotic twins who share, on average, only 50% of their genome (28–30). These observations provided the first unequivocal evidence that heredity plays a prominent role in drug metabolism, despite the multitude of other environmental and clinical factors that may potentially influence variability in drug exposure. Subsequently, the debrisoquine/sparteine (CYP2D6) polymorphism was first identified by Smith in London, England and Eichelbaum in Bonn, Germany (31,32). Since then, numerous genetic polymorphisms have been firmly documented and functionally characterized in various drug metabolizing enzymes and drug transporters (33,34). Most experts in the field of pharmacogenomics now agree that hereditary factors play an important role in drug disposition.

Although the focus of studies on the clinical relevance of genetic variations in pharmacokinetic pathways has been mostly on drug safety, there is both theoretical basis and empirical evidence to suggest that drug efficacy can also be influenced. In this regard, an early anecdotal obser vation made by Smith during the discovery of debrisoquine polymorphism can be instructive. In a pharmacokinetic study of debrisoquine in 1970s, a subtherapeutic dose resulted in unexpected and profound decrease in blood pressure of a subject, who, in fact, was one of the study investigators (31,35,36). Implicit in this historical account is that the side effect (hypotension) experienced by Smith himself is essentially an *extension* of the primary intended pharmacological effect of the antihypertensive drug debrisoquine. Such ADRs typified by an augmentation of the primary ''therapeutic'' effect of a drug are classified as Type A drug reactions (37). Type A reactions are common, dose- or concentration-dependent and also include ADRs due to overdose as well as drug–drug interactions. Thus, drug efficacy and toxicity are usually observed on a successive concentration gradient. This means that any genetic or environmental factor that can influence drug concentrations may also explain individual variations in both efficacy and safety. Dramatic differences in cure rates of helicobacter pylori infection and peptic ulcer with omeprazole (a CYP2C19 substrate) and amoxacillin among patients with different CYP2C19 genotypes further attest to the relevance of pharmacogenomic variability in drug metabolism with respect to drug efficacy (38). In Chapter 9, Meyer and Flockhart provide specific examples of genetic contributions in pharmacokinetics and their clinical relevance for drug safety and efficacy.

Genetic variations in receptors, ion channels and other types of drug targets constitute a more recent but growing body of evidence in favor of heredity and its role in drug efficacy and safety (39,40). Already, there are accumulating data, for example, documenting the clinical significance of variations in genes encoding arachidonate 5-lipoxygenase (ALOX5) and ß$_2$-adrenoreceptor for response to ALOX5 inhibitors and β_2-agonists, respectively, in asthma; serotonin receptors in response to atypical antipsychotic clozapine; and dopamine D3 receptor gene (*DRD3*) for predisposition to

the movement disorder, tardive dyskinesia, induced by typical antipsychotic drugs (41–44). We note that the commonly occurring and unpreventable ADRs discussed in Sec. 2.1 may be attributable in part to genetic variations in drug targets and/or presently unknown genetic differences in drug disposition, for instance, in phase II drug metabolizing enzymes and drug transporters (45–48).

Genes rarely act in isolation and hence, genetic contributions to pharmacological variability are subject to influences by gene–environment interactions as well as epistatic interactions among various genetic loci (49–51). Indeed, the most likely scenario is that the underlying basis of genetically determined variability in treatment response and common ADRs is polygenic. It is therefore often difficult to estimate the *composite genetic component* in pharmacological traits (27). This information is essential before decisions on further molecular pharmacogenomic work can be justified. Typically, heritability estimates are obtained using the twin method. Although twin studies are indeed very useful to establish the baseline heritability figures for common complex diseases, they may have limited applicability in pharmacological responses to drugs and other xenobiotics. Some of these limitations include difficulties in recruitment of twins, obtaining clinical outcome data in both twins (since the twin pairs may not suffer from the same disease at the same time) as well as the financial cost of twin investigations. To circumvent the difficulties associated with dissection of genetic components with the twin approach, a repeated-drug-administration (RDA) method has been earlier proposed wherein between- and within-subject variances in drug efficacy or safety are compared (27,52–55). A relatively larger between-subject variance as measured by the RDA analysis points to significance of hereditary factors in pharmacological variability. Recent applications of the RDA method demonstrate that genetics also plays a paramount role in pharmacological traits hitherto not subjected to pharmacogenomic analysis such as renal drug disposition (27,56).

2.3. Recognition of Molecular Heterogeneity in Human Diseases: The Need for New Drug Targets

The completion of the HGP 4 years ago provided the reference framework for the sequence of some 30,000–40,000 protein-coding genes in the human genome (57,58). At that time, a high-density map of the human genome consisting of 1.42 million single-nucleotide polymorphisms (SNPs) (now >3.7 million) have also been made available (59,60). These advances witnessed in parallel the development of high-throughput genomic technologies and the related infrastructure; this was favorably reflected in marked decreases in genotyping costs over the past several years. On the other hand, some of the alternative viewpoints consider the HGP more of an engineering triumph with the development of tools and technologies for genetic research

(24). For translation of the human genome sequence to biology, clinical medicine and science-based therapeutics, a second additional layer of complexity, namely, interindividual and population-to-population variability in the genome will need to be addressed. The latter issue has recently culminated in the launch of the International HapMap project to identify the SNPs and their patterns (haplotypes) on individual chromosomes in various human populations from Africa, Asia and Europe (61). In the present post-genomic era, the ultimate goal is to utilize the haplotype map of the human genome as a foundation for future genetic association studies of human diseases as well as drug efficacy and safety (62–66). A further important consideration is that genetic variants underlying variability in drug effects are unlikely to be limited to the coding elements of genes. Variants in regulatory regions are most likely to be implicated and it is quite feasible that non-coding, intronic elements, which influence gene transcription, have an important role as well. These considerations greatly broaden the scope of genetic influences that will need to be considered.

Why do we need to know the genetic basis of human diseases? And how does this information contribute to rational therapeutics? A recent biochemical classification of drug targets found that the largest group was comprised of receptors (45%) followed primarily by enzymes (28%), hormones and related factors (11%), ion channels (5%), and nuclear receptors (2%) (67). Overall, current drug targets across all therapeutic areas amount to only about 500 molecular targets (67,68). Considering the diversity of human diseases, there is no doubt that novel drug targets will be essential to develop drugs with improved efficacy. To this end, identification of the molecular genetic basis of diseases may have three fundamental contributions. First, knowledge of the genes will eventually help discern the identity of the corresponding proteins leading to disease and its clinical manifestations. In some cases, these proteins can serve as direct targets for therapeutic interventions by conventional small molecule drugs ($<$500 Da molecular weight). On the other hand, it is also reasonable to expect that not all disease-causing genes or their protein products will be druggable. A second alternative therapeutic strategy in such cases may involve targeting other components of the biological pathway(s) containing the genes associated with disease. Third, in the absence of molecular genetic corollaries of disease, it is noteworthy that entry points for most therapeutic interventions have thus far been at the protein level. The knowledge of a specific disease-causing gene or mutation may allow interventions further upstream in the biological cascade at the level of gene expression before the corresponding proteins of pathophysiological significance are actually synthesized. For example, libraries of small inhibitory RNA (siRNA) molecules are now being investigated in the pharmaceutical and biotechnology industries as potential therapeutic agents to silence the genes whose expression may predispose to disease (69).

A glance at recent advances in human genetics can be instructive to gain a balanced context on the drivers of pharmacogenomics in the near future. In rare monogenic diseases, a clear pattern of Mendelian inheritance can be discerned wherein human genetic variation in one or both copies of a single gene will predictably lead to clinical manifestations of the attendant disease (Fig. 2) (70). Through positional cloning approaches, more than 1400 genes for some 1200 Mendelian traits have been identified by the year 2003 (71). The majority of the mutations associated with monogenic diseases are in-frame amino-acid substitutions and nonsense codons (59%), deletions (22%) and insertions/duplications (7%) while only about 1% of the mutations linked to Mendelian traits were in regulatory regions of human genes (72). The remarkable success of the positional cloning strategy in monogenic diseases has not been uniformly reproducible upon application to multifactorial complex human diseases such as diabetes, schizophrenia and non-familial, sporadic forms of common cancers. A hallmark of complex diseases is that a multitude of genetic loci as well as the environment and life-style importantly contribute to disease risk. The multiplicity of contributory factors and their interactions result in non-Mendelian transmission of disease phenotypes in pedigrees (Fig. 2) (70). Further, genetic variations in regulatory regions of the genome that affect gene expression and translationally silent mutations are also likely to play a significant role in predisposition to multifactorial diseases (73,74). Hence, current nosologies based on clinical chemistry and symptoms reflect vastly divergent underlying molecular mechanisms leading to the final and ostensibly similar clinical syndromes among patients. This means that effective treatment of patient subpopulations typified by distinct molecular etiologies may require differential therapies with drugs that target a unique complement of biological targets (70,75). In the near future, we will therefore likely witness the emergence of genetic tests to assist and complement diagnosis of diseases with clinical nosology. To this end, a synergy between tests for genes associated with disease and drug effects can also be expected. For instance, the *APOE4* allele of the *APOE* gene is associated with predisposition and a lower age of onset for sporadic or late onset familial forms of Alzheimer's disease as well as poor response to acetylcholinesterase enzyme inhibitor tacrine (76,77).

2.4. Shift in Emphasis on Health Care Policy Toward Disease Prevention

The dictum in the traditional folklore of medicine has been the treatment and alleviation of acute symptoms of disease. Aforementioned advances in our understanding of the risk factors and molecular basis of human diseases are now paving the way to develop medicines that can prevent or slow the progression of disease phenomena. Schizophrenia and some of the neurodegenerative diseases such as Alzheimer's or Parkinson's disease

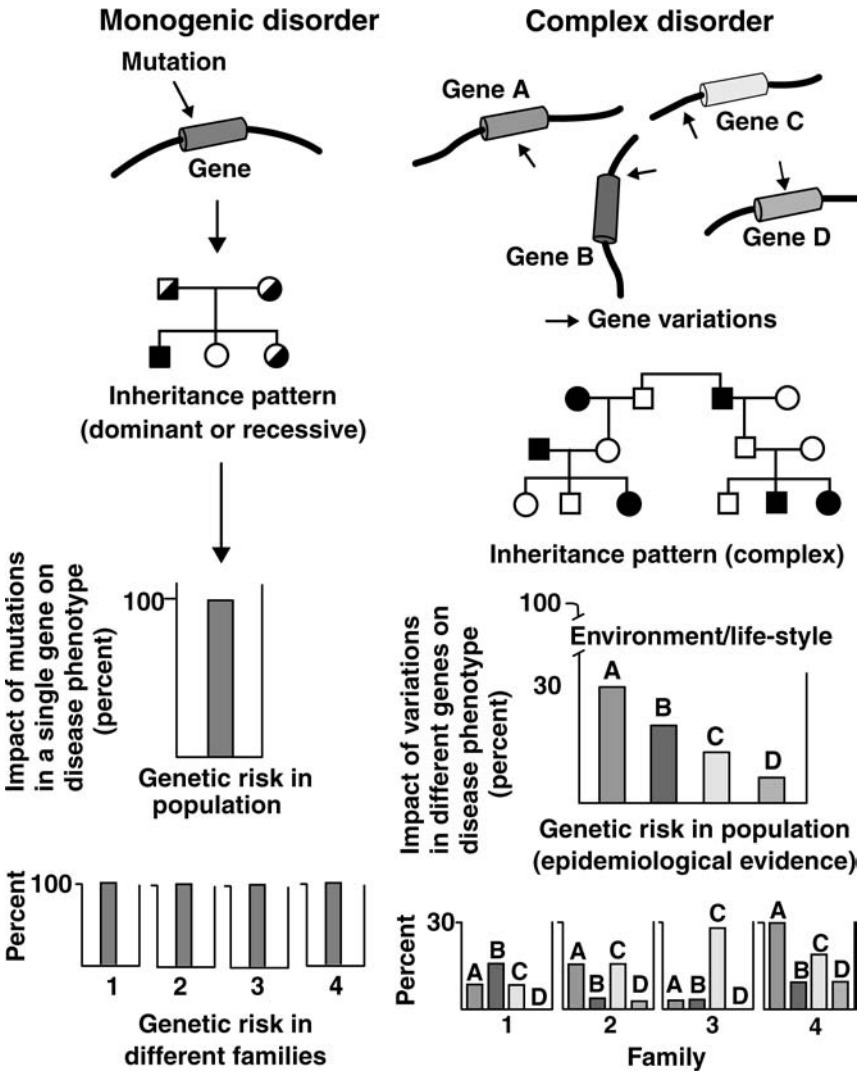


Figure 2 Inheritance of monogenic and complex (multifactorial) disorders. In monogenic diseases, mutations in a single gene are both necessary and sufficient to produce the clinical phenotype and to cause the disease. The impact of the gene on genetic risk for the disease is the same in all families. In complex disorders with multiple causes, variations in a number of genes encoding different proteins result in a genetic predisposition to a clinical phenotype. Pedigrees reveal no Mendelian inheritance pattern. Complex diseases subject to polygenic control may thus require multiple distinct drugs, or combination of drugs, directed to each complement of the causative genes and biological pathways. *Source*: From Ref. 70.

are being investigated in reference to neurodevelopmental theories to develop drugs that can halt the (patho)physiological processes leading to eventual clinical symptoms. The rise of statins and other lipid lowering drugs to prevent long-term cardiovascular morbidity and mortality is another example of preventative pharmacotherapies. It is thus noteworthy that the interest in personalized medicine does not happen in a vacuum and should be viewed within the larger framework of recent changes in health and therapeutic policies over the past decade.

Poised at the threshold of this upcoming and fundamental change in emphasis from treatment to prophylaxis in healthcare policy, pharmacogenomics will likely be shaped in the near future by particular attributes of the pharmacological interventions aimed at disease prevention. For instance, there will be an increasing demand to predict the long-term outcome and benefit-to-risk ratio of preventative therapies. Equally important, adherence to drug treatment by patients will need assurance beyond traditional physician visits and consultations, especially considering that such "future" patients will need to take drugs in the absence of acute symptoms and in the face of a probable disease risk in the far-too-distant future. It is in this very context that predictive pharmacogenomic tests will be essential for patients, physicians and insurers to make informed and rational decisions to subscribe to pharmacogenomically guided and customized therapies that may be offered at premium prices by drug manufacturers.

2.5. Asymmetry Between Genetics of Human Diseases and Pharmacological Phenotypes: An Impetus for Pharmacogenomic Research

The fundamental focus of human genetics is to establish the causal links between genes and phenotypes. Since its first introduction in late 1990s, the allure of pharmacogenomics has drawn a considerable number of human geneticists who have previously dealt with complex diseases. This importantly benefited and complemented the classical pharmacological approaches to questions of variability in drug effects. At the same time, a tendency has arisen to view pharmacological responses akin to disease phenotypes. There are however several fundamental differences unique to the field of pharmacology and the attendant phenotypes that may offer a relative advantage in application of genetic methodologies. It has long been known that most chronic human diseases initiate and progress over a considerable period of time before clinical findings become apparent. Thus, human genetics is essentially an observational science wherein molecular genetic markers are correlated with a disease process that is not experimentally controlled or reproducible for obvious ethical and practical reasons. A related corollary is that the scope of gene–environment interactions is often incalculable or difficult to estimate with adequate certainty. For all

the parallels between the disease traits and pharmacological phenotypes, the very difference between the *observational* nature of study designs in disease genetics and the ability to *experimentally* produce drug response phenotypes in a setting carefully controlled for environmental factors is perhaps one of the most salient tenets of pharmacology that may offer a marked opportunity to control for gene–environment interactions in pharmacogenomic investigations. For instance, it is feasible to quantify the baseline phenotypes prior to drug administration in humans, which can subsequently be subtracted from drug-induced phenotypes in biological systems. In theory, this can considerably facilitate correlative studies on human genetic variation and pharmacological phenotypes. By contrast, such experimental approaches are not applicable to the study of common multifactorial human diseases. This favorable asymmetry between disease genetics and pharmacogenomics is increasingly being recognized by scientists and clinicians with diverse backgrounds thereby coalescing interest in pharmacogenomics in various medical specialties in academia and the pharmaceutical industry.

2.6. Advances in Regulatory Science and Medicolegal Considerations

The transition from a new research idea to actual application in the clinic and drug development is a slow and arduous process. A recent search of the 2000 entries in the Physicians' Desk Reference (year 2003 electronic version) identified only 51 labels containing pharmacogenomic information (78). Moreover, presentations of the pharmacogenomic data in the labels were generally not in a form that could be readily applied in clinical practice (observations made by Dr. Lawrence Lesko) (78). Although it is difficult to estimate the number of drug labels that should appropriately contain data on host genetic make-up for rational dose titration or choice of medications, the latter example suggests that pharmacogenomic information is likely underrepresented in current drug labels. This is an important issue of relevance for adoption of pharmacogenomics in the clinic because it is not clear whether and to what extent research findings available as off-label scientific publications can be applied in routine medical practice with busy schedules. Still, drug labels do not always reflect the best available evidence (79). Recent announcement of a guidance document in March 2005 by the U.S. Food and Drug Administration (FDA) to encourage regulatory pharmacogenomics data submission by drug developers is a welcome development in this regard (80). Advances in regulatory science over the past three decades have been instructive on the critical role of regulatory guidance in widespread implementation of new technologies on practices in the pharmaceutical industry and clinical medicine. A case in point is the history of development of population pharmacokinetic approaches to explain variability in drug exposure. Despite the theoretical foundations established since

the 1970s and the availability of statistical software capable of conducting such advanced pharmacokinetic analyses, the routine use of population pharmacokinetics in clinical drug development has been considerably facilitated in part through recent guidelines developed by regulatory agencies (81). The present heightened public awareness on societal ramifications of the HGP and the attendant genomic technologies is increasingly prompting governments and policy makers to provide further guidance on pharmacogenomics. Moreover, regulatory consideration may soon be given to drugs already in the clinic (78). This should further encourage drug manufacturers to adopt pharmacogenomics, with the realization that regulatory review will be applicable at all phases of a medication's life-cycle, both prior and subsequent to introduction for human use. Seen in this light, it is conceivable that developments on the regulatory policy front will importantly shape and drive the application of pharmacogenomics technology in the very near future.

As information on the molecular basis of human diseases and individual variability in drug effects continues to accumulate, a threshold may soon be crossed beyond which the standard of care will include predictive pharmacogenomic testing. Advances in accessibility of pharmacogenomic technologies may thus challenge the traditional questions of ethics and litigation standards. In fact, the clinical significance of genetic variability in certain drug metabolizing enzymes such as thiopurine methyl transferase are already (TPMT) well-established, e.g., in relation to life-threatening toxicity associated with thiopurine anticancer drugs (82). In such cases where research findings increasingly impact patient care, demands by insurers and government payers as well as legal liability and litigation by patients who suffer from drug toxicity or treatment-failure may set a strong precedent for pharmacogenomic testing by physicians and the pharmaceutical industry. Looking further, it is tempting to suggest that physicians who attend focused continuing medical education courses on pharmacogenomics may presumably be subject to reduced premiums for malpractice insurance in the future. Such complex medicolegal dynamics between patients, insurers, healthcare providers, and drug manufacturers may eventually catalyze the adoption of pharmacogenomics as part of science-based therapeutics.

A recently released survey of American consumers ($n = 748$) and physicians ($n = 400$) suggests that nearly 50% of consumers would be agreeable to undergo genetic testing to determine which medication would be most appropriate for them and that they would be willing to pay more for an individually tailored prescription (83). About 80% of physicians and consumers, on the other hand, expressed views that genetically guided personalized medicine would have a favorable impact on the U.S. healthcare system (83). It will be of interest to also characterize consumer and physician

attitudes in other countries to determine the global patterns of acceptance of pharmacogenomics and the potential barriers to its implementation.

3. STRATEGIES FOR APPLICATION OF PHARMACOGENOMICS TO CUSTOMIZE THERAPY

3.1. Drug Discovery and Preclinical Development

Drug discovery has long relied on serendipitous observations. For drugs introduced in the 1950s and onward, the mode of drug action has been typically discovered after some period of use in the clinic. Over the past two decades, routine application of high-throughput screening (HTS) methods together with combinatorial chemistry and a rich array of chemical libraries significantly accelerated the entry of therapeutic candidates with clearly defined molecular targets to early phase clinical evaluation. Despite these advances, out of 1035 new drug applications approved in the 12-year period from 1989 to 2000 in the United States, a sizable fraction (54%) were differentiated from existing drugs primarily on the basis of dosage form, route of administration, or as a combination product with another active ingredient (84). Moreover, direct-to-consumer promotions of pharmaceutical products occasionally rely on claims of higher therapeutic "potency." An improvement in potency means that the same drug effect can be elicited at a lower drug concentration. This may favorably influence drug selectivity and safety in certain cases. The excessive emphasis on discovery of more potent compounds overlooks another basic currency of pharmacology: *efficacy*—the maximal effect that can be produced by a drug. This predicament is best exemplified by the moderate efficacy of current drugs in the clinic (Table 1). Efficacy is reflected in the asymptotic plateau of concentration–effect curves; it is determined by both the chemical structure of the drug and the biological attributes of the targeted receptor–effector systems (85). No matter how many structurally diverse hit and lead compounds are synthesized in early discovery, large volumes of compound libraries will not make up for the limited diversity in drug targets (only some 500 at present) thereby markedly constraining the upper limits of therapeutic drug effects in patients. For improvements in drug efficacy, advances made by combinatorial chemistry and HTS need to be complemented by the discovery of novel drug targets, or better characterization of the subtypes of existing targets (43,86). Genome-wide association studies of drug response or common human diseases may provide clues for discovery of novel drug targets as well as entry points for the development of first-in-class compounds with unique pharmacological mode(s) of action. In efforts for customized therapies, it is clear that the identification of pharmacogenomic biomarkers should start very early, preferably in drug

discovery and phase 2A/2B clinical studies for proof-of-concept (see also Sec. 3.2).

In the present post-genomic period, another favorable influence of the HGP in drug discovery will likely be seen in preclinical studies of global gene expression before and after drug treatment in animal species (87,88). Shared patterns of gene expression among drugs within the same therapeutic class or alternatively, drugs that produce a similar form of toxicity may pave the way to map the attendant biological networks essential for drug efficacy and safety. Further, identification of genes related to common human diseases may contribute towards better preclinical models of drug efficacy, an area that needs much improvement for rational scaling of data obtained in animals to humans.

In the ideal case, the existing HTS processes in drug discovery may soon be tailored to accommodate the specific subtypes of novel or existing drug targets. This can then provide insights on optimal therapeutic candidates prior to clinical trials and create niche subpopulations in whom drugs can be used with greater efficacy and optimal safety. Focused pharmacogenetic testing in the clinic will be essential before such customized drugs can be prescribed. A reverse and complementary approach is to choose the compounds that will bind with high affinity and functionally interact with all subtypes of a given novel drug target. This alternative strategy for drug discovery may alleviate the need for pharmacogenetic testing at point of care: the lead compounds are chosen or customized (instead of the patients), in this case, to accommodate all variants of a biological target. Conversely, these early individualization attempts in drug discovery may be constrained by limits of medicinal and combinatorial chemistry to synthesize compounds that will be versatile enough to bind and activate all drug target subtypes, while demonstrating an adequate safety, bioavailability and drug–drug interaction profile.

All these technical advances and promises of pharmacogenomic-guided drug discovery also come with attendant ethical and social policy challenges. For example, if drugs are discovered and developed only for the most common subtypes of drug targets, this may potentially create therapeutic orphan subpopulations—will there be financial or legislative mechanisms in place to ensure that these individual patients also benefit from novel and mechanism-oriented therapeutic interventions? When developed, genetically tailored medications will likely be available at premium prices. Who will take the responsibility for ethical promotion of such personalized drugs (89)?

It is interesting to note that among 1393 new drugs developed between 1975 and 1999, only 1.1% (16 drugs) was for tropical and other diseases prevalent in developing countries (90). While one-third of the world population live on less than US$2/day, and occasionally without access to essential drugs (91), it is foreseeable that some may question

how pharmacogenomics may contribute to discovery of drugs that will benefit the patients in the developing world (92). To this end, it is worthwhile to bear in mind that the genomes from numerous parasitic pathogens, bacteria and viruses have been sequenced and available for the research community (92). Hence, pharmacogenomics offers much hope for novel drug target discovery and development against infectious agents affecting world's populations. In addition, concepts and technologies developed in parallel to pharmacogenomics may be instrumental for identification of individuals who may be more sensitive or resistant to infectious organisms in developing countries. It is hoped that these promises offered by pharmacogenomics will materialize and will be put into practice with a global public health policy in mind (93). Finally, although pharmacogenomics will likely revolutionize how new drug candidates are tailored for individual patients, serendipity and keen scientific insights will still continue to play a role in the drug discovery process.

3.2. Early Phase Clinical Drug Development

Allometric scaling of preclinical data to humans is often limited by marked interspecies differences in drug disposition, drug targets as well as the inevitable biological contrasts between outbred human populations and the inbred laboratory animals with a homogenous genetic background. First-in-human studies and the subsequent phase 2A/2B proof-of-concept clinical trials of therapeutic candidates in early stages of drug development play a bridging role with drug discovery efforts before large scale and confirmatory phase three clinical studies can be rationalized. The principal aims of these early investigations in humans include estimation of pharmacokinetic variability, the optimal dose ranges and administration schedules, the risk for drug–drug interactions, and provision of evidence on proof-of-concept for drug efficacy (or the lack thereof) and mode of action. Typically, single and multiple ascending doses of a new molecular entity (NME) are administered to healthy volunteer subjects and carefully selected groups of patients with a relatively uniform and usually mild to moderate disease state. On the other hand, even though these studies collectively provide preliminary insights into presumed effects of NMEs in humans, the complete range of pharmacokinetic and pharmacodynamic variability within and among human populations is seldom available at the end of phase two drugs development. This predicament, along with the occasional disconnect between the efficacy of NMEs in animal models, and the clinic, have recently stimulated a surge for identification of early indicators or biomarkers of drug activity and mode of action (94,95).

Per definitions provided by the U.S. National Institutes of Health expert working group, a biological marker (biomarker) is "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a

therapeutic intervention'' (95). A surrogate endpoint represents a special subset of biomarkers that is intended to substitute for a clinical endpoint; this implies a stronger correlation between the surrogate biomarker and clinical outcome (e.g., therapeutic benefit, lack of efficacy, and toxicity). Insofar as the pharmacogenomic biomarkers are concerned, there is growing interest in characterizing changes in the expression of genes that encode known drug targets or other biological elements downstream from the immediate drug target–effector systems. A particular advantage in this regard is that the effects of drug treatment on expression of key pathways related to drug response may precede changes in clinical outcomes; this may alleviate the need for long-term clinical studies and expedite the drug development time frames (16). Moreover, genomic methodologies typically have a higher throughput than the conventional phenotypic or clinical endpoint-based measures of drug efficacy and safety. This may in effect provide a broader picture of drug effects in various pathophysiological pathways and in some cases, lead to discovery of hitherto unexpected additional mode(s) of drug action and indications for future therapeutic applications. It should be mentioned that all these exploratory and research-oriented applications of pharmacogenomics in early stage development contribute essentially to better characterization of NMEs and their effects in humans. The small study sample sizes in early phase clinical trials, however, are not sufficient to characterize the diagnostic sensitivity and specificity of genetic tests (see also Secs. 3.3 and 4.2). On the other hand, these early pharmacogenomic investigations, if judiciously analyzed and interpreted, can be invaluable for hypothesis-generation and the initial identification of genetic biomarkers that may later be utilized for customization of drug therapy.

Early stage clinical studies typically describe the pharmacokinetic variability associated with a NME but these estimates, in most cases, remain limited to the study samples (16). Broader views of pharmacokinetic variations within and among human populations can be obtained only if the mechanisms of variability in drug disposition are discerned. To this end, panel studies in patients representative of the population extremes (e.g., poor metabolizers vs. rapid or ultrarapid metabolizers) for a given drug disposition pathway may help to define the upper limit of contribution of individual CYP450 and Phase II drug metabolizing isozymes or drug transporters to pharmacokinetic variability in the general population (16). These panel investigations in genetically stratified samples may inform several critical drug development decisions including: (1) extent of variability in dosages required for an optimal response in the targeted population, (2) risk for competitive drug–drug interactions and the list of concomitant medications that may cause a pharmacokinetic interaction upon co-administration with the NME under clinical development, and (3) prioritizing pharmacokinetic studies in different populations (16,34,96).

Pharmacokinetic bridging-studies are usually conducted when regulatory drug approval is sought in various countries. These data provide guidance for registration of new drugs in different populations based on similarities or differences in drug exposure. Consider, e.g., that an NME is found to be metabolized by CYP2C19 in early phase clinical studies. Because CYP2C19 displays marked inter-ethnic differences in the frequency of poor metabolizers (3% in Caucasians and 20–24% in Asians), this would immediately point towards the need to conduct highly focused and comparative pharmacokinetic studies in Asia, North America, Europe, and possibly in other countries. Without knowledge of the precise mechanisms governing interindividual variability in drug exposure, such inter-ethnic differences in drug disposition remain undetected during drug development. If dose adjustments are not made based on genetic or phenotypic differences in CYP450 function early in clinical development, drugs may face the risk of withdrawal from human use after their introduction in the clinic.

A notable application of pharmacogenomics in early phase drug development, which extends beyond personalized medicine, is in proof-of-concept investigations. Phase 2A/2B trials enriched for patients with certain genetic subtypes of drug targets previously shown to confer an increased likelihood of response can facilitate decisions on whether and to what extent a NME is a viable therapeutic candidate. Conversely, an inadequate response to a NME in such enriched samples may serve as an early indication of possible therapeutic failure in the general patient population.

3.3. Late Stage Clinical Drug Development—Phase Three Registration Trials

Phase three clinical trials form the centerpiece for overall clinical drug development programs. These pivotal late stage clinical studies are typically conducted in large patient samples and play an important role in registration of therapeutic candidates and confirming their efficacy and safety. This is also the stage of drug development when genetic tests for customization of therapy should be validated with adequate attention to their diagnostic sensitivity and specificity. Any new information learned on subpopulations identified with pharmacogenomic tests at this phase of development can favorably influence the customization of drug labels. Despite investments and enthusiasm for discovery of novel drug targets and pharmacogenomic applications in early phase clinical trials, direct stratification of patients using genetic tests in advanced stages of drug development is still rare (97). To date, a great majority of the pharmacogenomic biomarker development efforts have taken place in studies that were primarily designed for other purposes: demonstration of drug efficacy and safety. The highly structured and defined time frames of the latter clinical trials may pose serious limitations for flexibility to accommodate statistical analysis and interpretation of

pharmacogenomic data. Also, the high degree of population variability that characterizes the samples recruited for large-scale drug trials adds a further level of complexity to the interpretation of pharmacogenetic data, because of potentially spurious results due. Additionally, the statistical power of the traditional efficacy or safety-oriented clinical investigations may not always be sufficient to identify and validate predictive genetic tests to tailor drug therapy; in cases where epistatic effects are sought (a very likely scenario) there is a high likelihood that most current studies will be vastly underpowered for pharmacogenomic analysis. To achieve personalized medicines in the clinic, commitment to prospective pharmacogenomic testing in phase three trials designed specifically for biomarker validation and development is essential.

Through identification of patient subpopulations that are more likely to display a therapeutic response, pharmacogenomics holds the potential to conduct highly focused clinical trials enriched with "responder" patients. It is anticipated that this may reduce the required sample size and the time to obtain regulatory approval for new drug candidates. If and when successful, pharmacogenomic-guided drug development will also require companion genetic tests to tailor drug dosages or the choice of medicines. It is therefore crucial to coordinate in parallel the timely development of pharmaceutical and diagnostic products (98).

With the anticipated discovery of genes underlying some of the common complex human diseases over the next decade, we may increasingly witness NMEs with new drug targets and mode of action. These drugs, lacking clinical precedence on the therapeutic value of the targeted novel biological pathways, also bring along a considerable degree of risk for unexpected treatment failure or drug toxicity in the clinic. If predictive pharmacogenomic tests are available, these projected risks and promises associated with such first-in-class drugs can meet with enthusiasm and acceptance by clinicians, patients and drug manufacturers. The recent introduction of Herceptin for the treatment of patients with breast cancer who over-express the *HER2* oncogene is a notable example in this regard (75,99). These critical drug development decisions can be further influenced by impact of genetic testing on the pharmaceutical patent life-cycle, the commercial promise of therapeutic candidates and the health policies in place for reimbursement of customized medicines and the companion pharmacogenomic diagnostic tests (23,98,100,101). Adoption of pharmacogenomics in drug development can be anticipated particularly in cases where severe drug toxicity is experienced in a small fraction of patients while the same drug may display adequate or remarkable efficacy in the majority of the patient population. Such therapeutic candidates can be brought to the clinic by genetic testing and identification of patients at risk for severe drug toxicity. Table 2 provides a summary of multidimensional criteria and hypothetical

Table 2 Multidimensional Criteria and Selected Examples That May Favor the Adoption of Pharmacogenomic Tests in Drug Development or at Point of Care in the Clinic

Drug class	Therapeutic alternatives	Clinical efficacy	Adverse Drug Reactions (ADRs)		Disease progression	Remaining patent exclusivity
			Severity	Prevalence		
First-in-class	None or limited alternatives	Low efficacy in the general population but high in a genetically defined subpopulation, or difficult to predict efficacy in preclinical models	Mild or moderate	Not an immediate consideration	Not an immediate consideration	Not an immediate consideration
First-in-class withdrawn post-approval due to toxicity, or hitherto failed in clinical development	None or limited alternatives	High	Serious morbidity or mortality	Low	Not an immediate consideration	At least 5 years or more
Any drug class	Not an immediate consideration	Time-to-response is as crucial as the extent of therapeutic effects	Not an immediate considera-tion	Not an immediate consideration	Only a narrow temporal window exists for drug inter-vention (e.g., Alzheimer's disease or adva-nced cancers refractory to conventional therapy)	Not an immediate consideration

(*Continued*)

Table 2 Multidimensional Criteria and Selected Examples That May Favor the Adoption of Pharmacogenomic Tests in Drug Development or at Point of Care in the Clinic (*Continued*)

Drug class	Therapeutic alternatives	Clinical efficacy	Adverse Drug Reactions (ADRs)		Disease progression	Remaining patent exclusivity
			Severity	Prevalence		
Any drug class	Not an immediate consideration	Not an immediate consideration	Not an immediate considera-tion	Not an immediate consideration	Not an immediate consideration	Patent expiry in the next 5 years—new patent application for novel therapeutic indications in subpopulations defined by genetic testing
Any drug class	Several alternative drugs are available—but all are limited by significant toxicity in a fraction of patients	High	Serious morbidity or mortality	Low to moderate	Not an immediate consideration	Therapeutic differentiation and new patent application in a subpopulation with improved benefit-to-risk ratio to circumvent the need for clinical monitoring of serious ADRs

Threshold values (low, moderate, or high) for efficacy or ADR prevalence were not defined in a numeric form as they may differ depending on, for instance, the unique attributes of the targeted disease, patient population and the health policies adopted at home countries.

but plausible scenarios whereby the adoption of pharmacogenomic-guided drug development and prescription in the clinic may be facilitated.

3.4. Applications Toward Drugs in the Clinic

A pivotal aim of pharmacogenomics is to provide the clinician with genetic tests that can be applied at relatively low cost in order to predict efficacy and adverse effects. The preceding sections have placed considerable emphasis on the anticipated contribution of pharmacogenomics to the development of new drugs. However, there exists a similarly great need to develop predictive tests that can be applied to existing agents. Antidepressant drugs are a highly illustrative example. Currently several classes of agents are in widespread use—these include specific serotonin uptake blockers, drugs with a putative action on brain norepinephrine as well as serotonergic systems, monoamine oxidase inhibitors and the older tricyclic antidepressants which are widely accepted to be highly effective but impose a significant adverse effects burden. The overall efficacy of all these drug classes is remarkably uniform with approximately 60% of patients meeting response criteria short of full remission and 10–15% less actually remitting. Some of these drugs are true ''blockbusters'' if one considers their annual sales and market share. It is entirely unclear whether patients who fail to respond to one class of drug will respond to another class and whether there are differences between different drug classes in their efficacy. Patients who do not respond are administered a sequential series of drugs with little information available to the clinician who wishes to make an informed decision based on rational considerations. When response or remission eventually does occur, it is quite possible that this is due to natural termination of the depressive episode, which is ultimately self-limiting in most but not all cases and not to the drug administered at the time. It is highly feasible that response/remission rates to the initial drug could be raised well beyond their current levels by the availability of genetic predictors of response (40). For this scenario to materialize prospective trials would need to be conducted and it is entirely unclear that the motivation exists for the pharmaceutical industry to support such studies when their targets are drugs that are no longer protected by patents. It is likely that intellectual property issues in this regard have not been sufficiently thought through and there may be possibilities that justify the investment but have not been considered.

4. BARRIERS

4.1. The Cost Factor and Access to Genomic Technologies

Despite the general enthusiasm and belief that pharmacogenomics will benefit drug therapy, the eventual decision to use genetic testing in the clinic may bear in part on cost-effectiveness analyses (102,103). This is not an unrealistic assumption because many of the newer therapies customized

by genetic tests are anticipated to be offered at premium prices to recover the research and development costs by the pharmaceutical manufacturers. In countries with a publicly funded health care and reimbursement system, cost-effectiveness analyses may particularly be a significant consideration. Similar economical evaluations have contributed in the past to acceptance of molecular genetic tests for the diagnosis of infectious pathogens such as *Chlamydia* and *Mycobacterium tuberculosis* as part of the microbiology services (103–105). Insofar as the monogenic variability in drug metabolism is concerned, for example, inpatients in a psychiatry unit who are at extremes for *CYP2D6* expression (poor and ultrarapid metabolizers) were found to incur, on average, $4000–$6000 higher healthcare costs when treated with drugs eliminated by CYP2D6 (106). By contrast, virtually no pharmacoeconomic data are available concerning the cost of multigenic variability in drug effects.

Veenstra et al. (107) recently proposed five basic requirements that may enhance the cost-effectiveness of a pharmacogenomic test: (1) provision of evidence that severe clinical or economic consequences can be avoided through the use of pharmacogenomics, (2) current methods of therapeutic drug monitoring are inadequate, (3) an unequivocal association between genotype and clinical phenotype, (4) availability of a rapid and relatively inexpensive genetic test, and (5) a relatively common genetic variant. It is foreseeable that the insurers and other third-party payers of pharmacogenomic tests as well as hospital formularies may soon demand pharmacoeconomic data concerning the value and clinical impact of predictive pharmacogenomic testing on hospitalization rates, quality of life and daily functioning of patients, well beyond the direct pharmacological effects of drugs in clinical trials.

It may be argued that the current trends to characterize an increasing number of genetic variants, including the enthusiasm for genome-wide pharmacogenomic association studies, may add to the cost of genetic testing for research purposes, or at point of care as a diagnostic tool. On the other hand, the unit cost of genotyping and other genomic technologies have decreased considerably over the past several years. Additionally, broader inquiries of the human genome recently uncovered regions or blocks characterized by high linkage disequilibrium (LD) interspersed with short segments of very low LD (recombination hotspots) (108,109). Within the high LD regions, there is limited haplotype diversity and the strong correlation between genetic markers creates redundancy in the informational value of each marker. This led to the recognition that common haplotypes within each block can be discerned with only a fraction of the markers, or a minimal set of SNPs named as haplotype tagging SNPs (htSNPs), present in a haplotype block (110). These advances may be advantageous to reduce the cost of pharmacogenomic studies further, especially in genome-wide association studies or when highly polymorphic candidate genes are being

investigated. By contrast, for diseases that may result from rare haplotypes or in studies of genomic regions with low LD, characterization of all genetic markers may still be necessary. Collectively, it is reasonable to anticipate in the near future that the bottleneck in clinical pharmacogenomic studies will shift towards the rate-limiting step of collection of accurate clinical phenotypic information and relational analyses and interpretations of genetic and clinical datasets.

4.2. Pharmacogenomic Testing at Point of Care: Technical Standards and Expectations for Diagnostic Applications

Technical barriers in pharmacogenomic research are increasingly being overcome by declining costs and increased throughput of genotyping (or gene expression) methodologies. To date, much emphasis in pharmacogenomics has been placed on SNPs and their characterization in clinical samples. For genetic testing at point of care to become a reality, a broader scope of human genetic variation will need to be captured, including small insertions/deletions, and nucleotide repeat polymorphisms in the genome. For instance, certain commonly occurring *CYP2D6* alleles are typified by insertions/deletion polymorphisms while the slow drug glucuronidation associated with the Gilbert's syndrome is a consequence of dinucleotide (TA) repeat polymorphism in the promoter region of *UGT1A1*. Attention to rare genetic variants will also be necessary particularly in cases where the test results inform critical decisions on choice of drug prescriptions or dosages. It is noteworthy that the required sensitivity and specificity of molecular genetic assays, in a diagnostic context, are markedly higher than the technical standards acceptable for pharmacogenomic research or biomarker discovery applications.

From a clinical standpoint, if there is a very toxic drug that can be prescribed by means of pharmacogenetic testing and exclusion of at risk patients, the only barrier between a patient and severe toxicity will be the pharmacogenetic test itself. Hence, in cases where the diagnostic sensitivity of the genetic test is not robust, a number of ethical and legal issues will readily emerge. Clinicians who are accustomed to high-throughput of diagnostic tests in clinical chemistry will understandably demand a comparable ease of interpretation of the test results, economic affordability and turn around times within several days or ideally, by the end of each patient's visit. These required standards for diagnostic pharmacogenomic tests at point of care are still not within reach in many countries and genetic testing largely remains restricted to specialized research laboratories or tertiary care medical centers.

4.3. Pharmacogenomics and Accelerated Drug Approval: Concerns for Long-Term Drug Safety

A key promise of pharmacogenomics for the pharmaceutical industry is accelerated drug approval in genetically stratified subpopulations or based on genomic surrogate markers of drug activity and efficacy. To be eligible for the accelerated approval program, the FDA however requires that "the medication must treat serious illnesses and show significant benefit over existing therapy for serious or life threatening illness, or provide benefit for serious or life threatening illness for which no therapy exists." The mandate for the latter program was established by the Prescription Drug User Fee Act of 1992 and the FDA Modernization Act of 1997 (111). The published literature over the past 5 years tends to adapt the view that pharmacogenomic biomarkers will markedly reduce the time and number of patients required for drug approval. This is indeed a reasonable expectation but some cautionary restraint is necessary to establish adequate post-approval pharmacovigilance procedures, particularly for drug safety. The standard drug approval process evaluates no more than 3000 patients and healthy volunteers combined; this sample size is able to confirm the primary therapeutic benefits and detect only the common ADRs of new drugs. Rare ADRs (e.g., $<1\%$) are typically detected after introduction of a new medication for routine clinical use. In fact, serious ADRs may be discovered as long as 36 years after regulatory approval (111). The concern for drug safety during the post-marketing phase is also supported by the observation that a significant number of drugs carry black box warnings on the label (9). With this in mind, it is plausible that pharmacogenomic-guided accelerated approval of drugs based on surrogate markers of efficacy in studies with smaller sample sizes may be fraught with safety issues following regulatory approval. In accordance with this, serious ADRs were reported for 79% of accelerated approval cancer or anti-HIV drugs compared with 25% of similar drugs that received standard approval between the years 1996 and 2002 (111). Taken together, this underscores the need to conduct confirmatory studies of efficacy and systematic post-marketing long-term safety evaluations for drugs that are developed by pharmacogenomic guidance. Consistency in definitions and collection of the data on safety endpoints should be planned and envisioned early in the drug development programs. This will ensure that both pre-registration and post-marketing safety data can be pooled to evaluate the broader population-based risks associated with customized therapies.

4.4. Clinical Barriers and Attention to Social, Legal, and Ethical Aspects

Studies that yield information relevant to pharmacogenetics might appear easy to perform. Ostensibly, the researcher simply needs to recruit a sample

of patients treated with a particular drug, obtain a DNA sample, rate the clinical effect of the drug, document adverse effects and then determine the relationship between genetic variants and the clinical phenotypes of interest. In practice, the situation is considerably more complex. In different populations the frequency of genetic variants can differ greatly, rendering stratification and population admixture major headaches in the interpretation of studies. Stratification and admixture may also exist within samples that are apparently homogeneous from the ethnic standpoint. These considerations require great care in the selection of study populations, the application of tests for admixture and the use of appropriate statistical procedures to account for population differences when samples from different populations are pooled (112,113). Studies that take interaction among genes (epistasis) into account, an important requirement when the genetic basis of a phenotype is polygenic, require large samples in order to have sufficient statistical power. Placebo effects are another important consideration. In the absence of a placebo control, apparent association of a gene with a positive therapeutic outcome could represent association with the placebo effect of the drug and not with a true drug effect, as might be the case for the highly studied association of response to antidepressant SSRI drugs with a polymorphism in the serotonin transporter. While the association has been supported by several studies (40) the only study to include a placebo group found an association with response among patients receiving placebo as well as patients receiving the active drug (114).

Definition of the clinical phenotype that is to be correlated with genetic variation may pose a significant challenge. Potential phenotypes include *response* to a drug (defined as improvement of symptoms to a predetermined degree), *remission* (defined as complete disappearance of the target disorder for a predetermined minimum time period) and also intermediate phenotypes such as *onset of therapeutic effect* and the *speed of response or remission*. It is conceivable that different genes may be implicated in each of these phenotypes particularly when the underlying genetic architecture is polygenic; there may also be varying degrees of overlap. Differentiating and applying these phenotypic definitions can be relatively straightforward, as in the case of cancer therapies but may be complex, as in psychiatric disorders, dementias, and other illness states where the etiology and pathology of the disorders is not known and the mechanisms of action of the drugs used to treat them not well established. Phenotypes may be defined categorically as presence or absence of response/remission or adverse effects or evaluated as continuous variables, the latter approach being more appropriate for analyzing genetic effects as QTLs. Consistent definition of the phenotype is a critical prerequisite for ensuring that studies are comparable and for facilitating meta-analysis which is a critical tool in evaluating the overall significance of ostensibly inconsistent results obtained by smaller, possibly under-powered studies. Thus, definition of the phenotype is a pivotal issue in pharmacogenetics. It needs to be addressed

prospectively in the design of studies in order that the required information be collected in the course of the study.

From the social, legal and ethical standpoints the acquisition and storage of DNA samples and genetic information are issues that greatly concern the public. There is no doubt that as with any confidential information regarding an individual, the potential for abuse of genetic information is considerable. The individual has the right to expect that privacy will be strictly maintained and that DNA and genetic information will be used only for the purpose that was intended and for which permission was given. On the other hand there is concern that the steps taken to safeguard the rights of individuals in regard to their DNA and genetic information may be excessive and may impede scientific inquiry and deny important potential benefits. This process has been termed "genetic exceptionalization" and it can be discerned in overly stringent rules and regulations that are applied only in the context of genetic research whereas, if justified, they should be applied to other situations as well. In many countries stringent conditions and safeguards that are applied to genetic research by internal review boards are not demanded of non-genetic projects that place the participant at the same or even greater risk.

Many of the ethical concerns raised by pharmacogenomics are shared by genetic research into disease predisposition. Others are specific. The Nuffield Council on Bioethics outlined and discussed some of the major issues (115). Their report identified four central areas of concern: The first is *information*. Since pharmacogenetic tests yield genetic information about individuals this raises issues of consent and confidentiality. The second area identified is *resource*. Pharmacogenetics may lower the cost of developing and delivering medicines but may also drive it up because of the need to incorporate pharmacogenetic testing at all pivotal stages and to design studies in accordance. The third concern is *equity*. Pharmacogenetics may significantly improve medical treatment for some people, but it may also result in more people falling into categories for which effective drugs are not developed, because of inadequate financial incentives to develop a drug that may be effective for only a small population, or for a large but economically poor population. The fourth general category identified by the Nuffield Council is *control*. Who should decide whether a patient takes a pharmacogenetic test? Should patients be entitled to a drug if they do not wish to take an associated test? In their report the Nuffield Council addresses a comprehensive series of specific questions that are subsumed by these general areas of concern. Among the issues addressed are informed consent, privacy and confidentiality, regulation of pharmacogenetic tests, re-instatement of withdrawn medicines based on pharmacogenetic information, allocation of resources, stratification and the development of new medicines, implications for racial groups, and pivotal aspects of the implementation of pharmacogenetics in clinical practice. The reader is referred to this authoritative report, further detailed consideration being beyond the scope of this chapter (115).

5. CONCLUSIONS AND FUTURE PERSPECTIVES

Medical therapeutics has long followed an empirical tradition based on average values of pharmacological effects in the population (27). Unfortunately, this approach does not lend itself to predictable and science-based therapeutics due to marked variations in drug effects among patients and populations. An alternative and preferred strategy to achieve optimal drug safety and efficacy is to elucidate the mechanisms underlying variability in therapeutic outcomes which can inform the rational choice of drugs and their dosages in individual patients or subpopulations.

Pharmacogenomics provides the necessary conceptual framework and technical infrastructure to identify the previously unaccounted host-specific genetic factors underlying common (or rare) medication side effects and therapeutic failure otherwise attributed to idiosyncratic reasons. In this regard, the initial methodological approach was comprised of hypothesis-driven candidate gene studies of drug disposition, safety, or mode of action. Over the past several years, genome-wide hypothesis-free clinical pharmacogenomic association studies and analysis of gene expression before and after drug treatment have been increasingly utilized for discovery of unprecedented biological pathways of relevance to pharmacology and human diseases. It is interesting to note that tangible examples of customized therapies with novel modes of action are still limited in the clinic. However, a lag period should normally be anticipated before any new technology or scientific paradigm bears fruits. New drug development usually takes 10–15 years and for therapeutic candidates with novel molecular targets identified through pharmacogenomic approaches, it would be reasonable to see the first examples in the clinic over the next 10-year period. By contrast, for drugs that are already available for clinical use, it should be feasible to develop genetically customized treatment guidelines within a relatively shorter time frame. It is difficult to verify the practical validity of this theoretical prediction since it is uncertain whether and to what extent the pharmaceutical industry will be willing to adopt pharmacogenomic testing for therapeutic agents in the clinic (23,100,101). Although patents provide protection for market exclusivity of a given pharmaceutical product, pharmacogenomic diagnostic tests for the same compound can be developed by more than one investigator or institution thereby limiting the economic promise of a genetic test. Because most drug effects are subject to polygenic control, it is foreseeable that a multitude of genetic tests may eventually be developed in various populations by different private or academic not-for-profit interest groups. Although this is certainly an advantageous situation for patients and consumers of pharmacogenomic tests, it may potentially decrease the interest on the part of the pharmaceutical industry to pursue customized therapies once a drug is available for human use. It remains to be seen how, and under which conditions, research funding will be

available for pharmacogenomic investigations involving drugs available in the clinic. As aptly stated by several investigators in the field, "the promise of pharmacogenomics is already here—the reality is getting closer (97)."

Customization of drug therapy by pharmacogenomics is an arduous but worthwhile task that requires commitment of substantial research and economic resources before tangible results at point of care can be obtained. During this process, it is crucial to adopt a longer term vision, beyond the immediate goal of obtaining regulatory approval, to enhance the entire life cycle, and quality of a medicinal product: i.e., both prompt and timely introduction of new drugs to patients as well as their sustainable use in the clinic, without post-registration withdrawal or black box warnings, should be taken into account as part of the evaluations on the overall success of pharmacogenomic-guided drug development programs.

There will be several additional and foreseeable challenges on the path to targeted therapies. Pharmacogenomic biomarkers will likely be population-specific; divergent sets of genes and epistatic interactions may influence drug effects in different populations. Another crucial consideration is the recognition that the role of genetics in pharmacology depends on the environment (temporal, geographic, or therapeutic) in which drugs are being administered. Thus, the genetic components in pharmacological variability are not physical constants; their magnitude can vary depending on the gene-environment interactions at the time of a pharmaceutical intervention. Drug effects that are apparently under strong genetic control in a certain therapeutic setting may be controlled entirely by environmental factors in another context. This means that environmental components of pharmacological variability, along with the attendant genetic factors, have to be identified in concert to develop unequivocal diagnostic genetic tests and customized therapies in the clinic. The broader philosophical questions surrounding gene patents, commercial genetic testing in the clinic and how best to bring capital, morality and knowledge into a productive, and ethical relationship still need to be resolved (115–123). Despite these challenges, the next 10 years will be an exciting and yet decisive period for pharmacogenomics. The gains will be achieved in small but significant increments and will favorably influence therapeutics as a science.

ACKNOWLEDGMENTS

Supported in part by the Southern California Institute for Research and Education, the Ontario Mental Health Foundation and by grants from the Israel Ministry of Science and the Israel Ministry of Health. Helpful discussions with Drs. Janice E. Graham (Department of Bioethics, Dalhousie University) and Azade Seyhan (Department of German, Comparative Literature and Philosophy, Bryn Mawr College) are gratefully acknowledged.

REFERENCES

1. Kurth JH. Pharmacogenetics—the horizon. Rev Gastroenterol Disord 2000; 3(suppl 1):S3–S8.
2. Reidenberg MM. Clinical pharmacology: the scientific basis of therapeutics. Clin Pharmacol Ther 1999; 66:2–8.
3. Holford NH, Sheiner LB. Kinetics of pharmacologic response. Pharmacol Ther 1982; 16:143–166.
4. Garrod AE. The incidence of alcaptonuria: a study in chemical individuality. Lancet 1902; 2:1616–1620.
5. Reidenberg MM. Individualization of drug therapy. Med Clin North Am 1974; 58:905–1162.
6. Lesko LJ, Woodcock J. Pharmacogenomic-guided drug development: regulatory perspective. Pharmacogenomics J 2002; 2:20–24.
7. Lerer B. Genes and psychopharmacology: exploring the interface. In: Lerer B, ed. Pharmacogenetics of Psychotropic Drugs. Cambridge: Cambridge University Press, 2002:3–17.
8. Manasco PK, Arledge TE. Drug development strategies. In: Rothstein MA, ed. Pharmacogenomics Social, Ethical, and Clinical Dimensions. Hoboken, NJ: Wiley, 2003:83–97.
9. Spear BB, Heath-Chiozzi M, Huff J. Clinical application of pharmacogenetics. Trends Mol Med 2001; 7:201–204.
10. Phillips KA, Veenstra DL, Oren E, Lee JK, Sadee W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. JAMA 2001; 286:2270–2279.
11. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998; 279:1200–1205.
12. Lazarou J, Pomeranz BH, Corey PN. Reply—Adverse drug reactions in hospitalized patients. JAMA 1998; 280:1743.
13. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. JAMA 1997; 277:301–306.
14. Kohn L, Corrigan J, Donaldson M, eds. To Err is Human: Building a Safer Health System. The Committee on Quality of Health Care in America, Institute of Medicine. Washington, DC: National Academy Press, 2000.
15. Sheiner LB. Clinical pharmacology and the choice between theory and empiricism. Clin Pharmacol Ther 1989; 46:605–615.
16. Albers LJ, Ozdemir V. Pharmacogenomic-guided rational therapeutic drug monitoring: conceptual framework and application platforms for atypical antipsychotics. Curr Med Chem 2004; 11:297–312.
17. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. Nat Rev Genet 2003; 4:937–947.
18. Kalow W. Pharmacogenetics and personalised medicine. Fundam Clin Pharmacol 2002; 16:337–342.
19. Reidenberg MM. Evolving ways that drug therapy is individualized. Clin Pharmacol Ther 2003; 74:197–202.

20. Ingelman-Sundberg M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. J Intern Med 2001; 250:186–200.
21. Goldstein DB. Pharmacogenetics in the laboratory and the clinic. N Engl J Med 2003; 348:553–556.
22. Hedgecoe A, Martin P. The drugs don't work: expectations and the shaping of pharmacogenetics. Soc Stud Sci 2003; 33:327–364.
23. Williams-Jones B, Corrigan OP. Rhetoric and hype: where's the ''ethics'' in pharmacogenomics? Am J Pharmacogenomics 2003; 3:375–383.
24. Terwilliger JD, Weiss KM. Confounding, ascertainment bias, and the blind quest for a genetic ''fountain of youth.''. Ann Med 2003; 35:532–544.
25. Johnson JA, Bootman JL. Drug-related morbidity and mortality. A cost-of-illness model. Arch Intern Med 1995; 155:1949–1956.
26. Physicians' Desk Reference. 54th ed. PDR Staff, eds. Greenwood Village, CO: Thomson Micromedex, 2000.
27. Ozdemir V, Kalow W, Tothfalusi L, Bertilsson L, Endrenyi L, Graham JE. Multigenic control of drug response and regulatory decision-making in pharmacogenomics: the need for an upper-bound estimate of genetic contributions. Curr Pharmacogenomics 2005; 3:53–71.
28. Vesell ES, Page JG. Genetic control of drug levels in man: antipyrine. Science 1968; 161:72–73.
29. Alexanderson B, Evans DA, Sjoqvist F. Steady-state plasma levels of nortriptyline in twins: influence of genetic factors and drug therapy. Br Med J 1969; 4:764–768.
30. Endrenyi L, Inaba T, Kalow W. Genetic studies of amobarbital elimination based on its kinetics in twins. Clin Pharmacol Ther 1976; 20:701–714.
31. Mahgoub A, Idle JR, Dring LG, Lancaster R, Smith RL. Polymorphic hydroxylation of debrisoquine in man. Lancet 1977; 2:584–586.
32. Eichelbaum M, Spannbrucker N, Steincke B, Dengler HJ. Defective *N*-oxidation of sparteine in man: a new pharmacogenetic defect. Eur J Clin Pharmacol 1979; 16:183–187.
33. Alvan G, Bertilsson L, Dahl ML, Ingelman-Sundberg M, Sjoqvist F. Moving toward genetic profiling in patient care: the scope and rationale of pharmacogenetic/ecogenetic investigation. Drug Metab Dispos 2001; 29:580–585.
34. Aklillu E, Hidestrand M, Ingelman-Sundberg M, Malmebo S, Westlind A. Recent progress in drug metabolism research. Drug News Perspect 2002; 15:528–534.
35. Smith RL. Introduction. In: Franklin RA, Parke DV, eds. Human Genetic Variations in Oxidative Drug Metabolism. Xenobiotica 1986; 16:361–365.
36. Meyer UA, Skoda RC, Zanger UM. The genetic polymorphism of debrisoquine/sparteine metabolism-molecular mechanisms. Pharmacol Ther 1990; 46:297–308.
37. Rawlins MD, Thomas SHL. Mechanisms of adverse drug reactions. In: Davies DM, Ferner RE, de Glanville H, eds. Davies's Textbook of Adverse Drug Reactions. 5th ed. London: Chapman & Hall Medical, 1998:40–64.
38. Furuta T, Ohashi K, Kamata T, Takashima M, Kosuge K, Kawasaki T, Hanai H, Kubota T, Ishizaki T, Kaneko E. Effect of genetic differences in omeprazole metabolism on cure rates for Helicobacter pylori infection and peptic ulcer. Ann Intern Med 1998; 129:1027–1030.

39. Evans WE, McLeod HL. Pharmacogenomics—drug disposition, drug targets, and side effects. N Engl J Med 2003; 348:538–549.
40. Lerer B, Macciardi F. Pharmacogenetics of antidepressant and mood-stabilizing drugs: a review of candidate-gene studies and future research directions. Int J Neuropsychopharmacol 2002; 5:255–275.
41. Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci USA 2000; 97:10483–10488.
42. Drazen JM, Yandava CN, Dube L, Szczerback N, Hippensteel R, Pillari A, Israel E, Schork N, Silverman ES, Katz DA, Drajesk J. Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. Nat Genet 1999; 22:168–170.
43. Arranz MJ, Munro J, Birkett J, Bolonna A, Mancama D, Sodhi M, Lesch KP, Meyer JF, Sham P, Collier DA, Murray RM, Kerwin RW. Pharmacogenetic prediction of clozapine response. Lancet 2000; 355:1615–1616.
44. Lerer B, Segman RH, Fangerau H, Daly AK, Basile VS, Cavallaro R, Aschauer HN, McCreadie RG, Ohlraun S, Ferrier N, Masellis M, Verga M, Scharfetter J, Rietschel M, Lovlie R, Levy UH, Meltzer HY, Kennedy JL, Steen VM, Macciardi F. Pharmacogenetics of tardive dyskinesia: combined analysis of 780 patients supports association with dopamine D3 receptor gene Ser9Gly polymorphism. Neuropsychopharmacology 2002; 27:105–119.
45. De Leon J. Glucuronidation enzymes, genes and psychiatry. Int J Neuropsychopharmacol 2003; 6:57–72.
46. Liston HL, Markowitz JS, DeVane CL. Drug glucuronidation in clinical psychopharmacology. J Clin Psychopharmacol 2001; 21:500–515.
47. Pirmohamed M, Park BK. Genetic susceptibility to adverse drug reactions. Trends Pharmacol Sci 2001; 22:298–305.
48. Ozdemir V, Shear NH, Kalow W. What will be the role of pharmacogenetics in evaluating drug safety and minimising adverse effects? Drug Saf 2001; 24:75–85.
49. Sellers EM, Tyndale RF. Mimicking gene defects to treat drug dependence. Ann NY Acad Sci 2000; 909:233–246.
50. Meyer UA, Gut J. Genomics and the prediction of xenobiotic toxicity. Toxicology 2002; 181–182:463–466.
51. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet 2002; 11:2463–2468.
52. Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. Pharmacogenetics 1998; 8:283–289.
53. Kalow W, Endrenyi L, Tang BK. Repeat administration of drugs as a means to assess the genetic component in pharmacological variability. Pharmacology 1999; 58:281–284.
54. Kalow W, Ozdemir V, Tang BK, Tothfalusi L, Endrenyi L. The science of pharmacological variability: an essay. Clin Pharmacol Ther 1999; 66:445–447.
55. Ozdemir V, Kalow W, Tang BK, Paterson AD, Walker SE, Endrenyi L, Kashuba AD. Evaluation of the genetic component of variability in CYP3A4

activity: a repeated drug administration method. Pharmacogenetics 2000; 10:373–388.
56. Leabman MK, Giacomini KM. Estimating the contribution of genes and environment to variation in renal drug clearance. Pharmacogenetics 2003; 13:581–584.
57. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001; 15:860–921.
58. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science 2001; 291:1304–1351.
59. The International SNP Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 409:928–933.
60. Jiang R, Duan J, Windemuth A, Stephens JC, Judson R, Xu C. Genome-wide evaluation of the public SNP databases. Pharmacogenomics 2003; 4:779–789.
61. The International HapMap Consortium. The International HapMap Project. Nature 2003; 426(6968):789–796.
62. Judson R, Salisbury B, Schneider J, Windemuth A, Stephens JC. How many SNPs does a genome-wide haplotype map require?. Pharmacogenomics 2003; 3:379–391.
63. Hoehe MR. Haplotypes and the systematic analysis of genetic variation in genes and genomes. Pharmacogenomics 2003; 4:547–570.
64. Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. Trends Genet 2003; 19:135–140.
65. Athanasiou MC, Malhotra AK, Xu C, Stephens JC. Discovery and utilization of haplotypes for pharmacogenetic studies of psychotropic drug response. Psychiatr Genet 2002; 12:89–96.
66. Malhotra AK. From pharmacogenetics to pharmacogenomics of psychotropic drug response. In: Lerer B, ed. Pharmacogenetics of Psychotropic Drugs. Cambridge: Cambridge University Press, 2002:21–35.
67. Drews J, Ryser S. The role of innovation in drug development. Nat Biotechnol 1997; 15:1318–1319.
68. Drews J. Drug discovery: a historical perspective. Science 2000; 287: 1960–1964.
69. Sioud M. Therapeutic siRNAs. Trends Pharmacol Sci 2004; 25:22–28.
70. Peltonen L, McKusick VA. Genomics and medicine. Dissecting human disease in the postgenomic era. Science 2001; 291:1224–1229.
71. Page GP, George V, Go RC, Page PZ, Allison DB. Are we there yet? Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. Am J Hum Genet 2003; 73:711–719.
72. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nat Genet 2003; 33(suppl):228–237.
73. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. Nat Rev Genet 2002; 3: 285–298.
74. Farrall M. Quantitative genetic variation: a post-modern view. Hum Mol Genet 2004; 13(suppl 1):R1–R7.

75. Ross JS, Fletcher JA, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Pusztai L, Bloom KJ. The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy. Oncologist 2003; 8:307–325.
76. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer disease meta-analysis consortium. JAMA 1997; 278:1349–1356.
77. Poirier J, Delisle MC, Quirion R, Aubert I, Farlow M, Lahiri D, Hui S, Bertrand P, Nalbantoglu J, Gilfix BM, Gauthier S. Apolipoprotein E4 allele as a predictor of cholinergic deficits and treatment outcome in Alzheimer disease. Proc Natl Acad Sci USA 1995; 92:12260–12264.
78. CDER, Food and Drug Administration, U.S. Department of Health and Human Services. Pharmacogenetics: improvement of existing drug treatments; drug interactions: metabolism and transport-based. Clinical Pharmacology Subcommittee, the Advisory Committee for Pharmaceutical Science Meeting, April 22–23, 2003.
79. Rajpal A, Reidenberg MM. Drug labeling should be kept current. Clin Pharmacol Ther 2003; 73:4–6.
80. CDER, Food and Drug Administration, U.S. Department of Health and Human Services. Guidance for Industry: Pharmacogenomic Data Submissions, March 22, 2005. Available at: http://www.fda.gov/cder/guidance/6400fnl.htm—Accessed May 9, 2005.
81. CDER, Food and Drug Administration, U.S. Department of Health and Human Services. Guidance for Industry: Population Pharmacokinetics, February 10, 1999. Available at: http://www.fda.gov/cder/guidance/1852fnl.pdf—Accessed May 9, 2005.
82. Weinshilboum R. Inheritance and drug response. N Engl J Med 2003; 348:529–537.
83. [No author listed]. Bulletin Board: Study finds overwhelming majority of American consumers and physicians back development of personalized medicine. Pharmacogenomics 2003; 4:13.
84. The National Institute for Health Care Management (NIHCM) Research and Educational Foundation. Changing Patterns of Pharmaceutical Innovation. May 28, 2002. Available at: http://www.nihcm.org/innovations.pdf—Accessed May 9, 2005.
85. Nies AS. Principles of therapeutics. In: Hardman JG, Limbird LE, Gilman AG, eds. Goodman and Gilman's the Pharmacological Basis of Therapeutics. 10th ed. New York: McGraw-Hill, 2001:45–66.
86. Chumakov I, Blumenfeld M, Guerassimenko O, et al. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proc Natl Acad Sci USA 2002; 99:13675–13680.
87. Wong ML, Licinio J. From monoamines to genomic targets: a paradigm shift for drug discovery in depression. Nat Rev Drug Discov 2004; 3:136–151.
88. Wong ML, O'Kirwan F, Hannestad JP, Irizarry KJ, Elashoff D, Licinio J. St John's wort and imipramine-induced gene expression profiles identify cellular functions relevant to antidepressant action and novel pharmacogenetic

candidates for the phenotype of antidepressant treatment response. Mol Psychiatry 2004; 9:237–251.
89. Quick JD, Hogerzeil HV, Rago L, Reggi V, de Joncheere K. Ensuring ethical drug promotion—whose responsibility? Lancet 2003; 362:747.
90. Trouiller P, Olliaro P, Torreele E, Orbinski J, Laing R, Ford N. Drug development for neglected diseases: a deficient market and a public-health policy failure. Lancet 2002; 359(9324):2188–2194.
91. Quick JD. Ensuring access to essential medicines in the developing countries: a framework for action. Clin Pharmacol Ther 2003; 73:279–283.
92. Pang T. Impact of pharmacogenomics on neglected diseases of the developing world. Am J Pharmacogenomics 2003; 3:393–398.
93. [No authors listed]. Timely messages for the South. Nature 2002; 415(6867):1.
94. Downing GJ. Enhancing pathways to therapeutic development with clinical biomarkers. Am J Geriatr Psychiatry 2002; 10:646–648.
95. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001; 69:89–95.
96. Bertilsson L, Dahl ML, Dalen P, Al-Shurbaji A. Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs. Br J Clin Pharmacol 2002; 53:111–122.
97. Sanders R. Molecular diagnostics and personalized medicine. Pharmacogenomics 2003; 4:541–545.
98. Eisenberg RS. Will pharmacogenomics alter the role of patents in drug development? Pharmacogenomics 2002; 3:571–574.
99. Leyland-Jones B. Maximizing the response to Herceptin therapy through optimal use and patient selection. Anticancer Drugs 2001; 12(suppl 4):S11– S17.
100. Shah J. Economic and regulatory considerations in pharmacogenomics for drug licensing and healthcare. Nat Biotechnol 2003; 21:747–753.
101. Service RF. Surviving the blockbuster syndrome. Science 2004; 303(5665):1796–1799.
102. Phillips KA, Van Bebber SL, Veenstra D, Sakowski J. The economics of pharmacogenomics. Curr Pharmacogenomics 2003; 1:277284.
103. Wedlund PJ, de Leon J. Pharmacogenomic testing: the cost factor. Pharmacogenomics J 2001; 1:171–174.
104. Drobniewski FA, Watterson SA, Wilson SM, Harris GS. A clinical, microbiological and economic analysis of a national service for the rapid molecular diagnosis of tuberculosis and rifampicin resistance in Mycobacterium tuberculosis. J Med Microbiol 2000; 49:271–278.
105. Fedricks DN, Relman DA. Application of polymerase chain reaction to the diagnosis of infectious diseases. Clin Infect Dis 1999; 29:475–488.
106. Chou WH, Yan FX, de Leon J, Barnhill J, Rogers T, Cronin M, Pho M, Xiao V, Ryder TB, Liu WW, Teiling P, Wedlund PJ. Extension of a pilot study: impact from the cytochrome P450 2D6 polymorphism on outcome and costs associated with severe mental illness. J Clin Psychopharmacol 2000; 2:246–251.
107. Veenstra DL, Higashi MK, Phillips KA. Assessing the cost-effectiveness of pharmacogenomics. AAPS PharmSci 2000; 2:E29.

108. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science 2002; 296: 2225–2229.
109. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 2001; 294:1719–1723.
110. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet 2001; 29:233–237.
111. Ladewski LA, Belknap SM, Nebeker JR, Sartor O, Lyons EA, Kuzel TC, Tallman MS, Raisch DW, Auerbach AR, Schumock GT, Kwaan HC, Bennett CL. Dissemination of information on potentially fatal adverse drug reactions for cancer drugs from 2000 to 2002: First results from the research on adverse drug events and reports project. J Clin Oncol 2003; 21:3859–3866.
112. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999; 65:220–228.
113. Lerer B, Macciardi F, Segman RH, Adolfsson R, Blackwood D, Blairy S, Del-Favero J, Dikeos D, Kaneva E, Lilli R, Massat I, Milanova V, Muir W, Noethen M, Oruc L, Petrova T, Papadimitriou GN, Rietschel M, Serretti A, Souery D, Van Gestel S, Van Broeckhoven C, Mendlewicz JM. Variability of 5-HT2C receptor cys23ser polymorphism among European populations and vulnerability to affective disorder. Mol Psychiatry 2001; 6:579–585.
114. Rausch JL, Johnson ME, Fei YJ, Li JQ, Shendarkar N, Hobby HM, Ganapathy V, Leibach FH. Initial conditions of serotonin transporter kinetics and genotype: influence on SSRI treatment trial outcome. Biol Psychiatry 2002; 51:723–732.
115. The Nuffield Council on Bioethics. Pharmacogenetics: ethical issues. September 20, 2003. Available at: http://www.nuffieldbioethics.org/filelibrary/pdf/pharmacogenetics_report.pdf—Accessed May 9, 2005.
116. Williams-Jones B, Burgess M. Social contract theory and just decision-making: Lessons from genetic testing for the BRCA mutations. Kennedy Institute of Ethics J 2004; 14:115–142.
117. Smart A, Martin P, Parker M. Tailored medicine: whom will it fit? The ethics of patient and disease stratification. Bioethics 2004; 18:322–342.
118. Webster A, Martin P, Lewis G, Smart A. Integrating pharmacogenetics into society: In search of a model. Nat Rev Genet 2004; 5:663–669.
119. Williams-Jones B, Graham JE. Actor-network theory: a tool to support ethical analysis of commercial genetic testing. New Genetics and Society 2003; 22:271–296.
120. Williams-Jones B. Knowledge commons or economic engine—What's a university for? J Medical Ethics 2005; 31:249–250.

121. Hedgecoe AM. Terminology and the construction of scientific disciplines: the case of pharmacogenomics. Sci Technol Human Values 2003; 28:513–537.
122. Williams-Jones B, Ozdemir V. Enclosing the 'Knowledge Commons': Patenting genes for disease risk and drug response at the university-industry interface. In: Ethics and Law of Intellectual Property. Current Problems in Politics, Science and Technology. Lenk C, Hoppe N and Andorno R. (eds.) (in press).
123. Weijer C, Miller PB. Protecting communities in pharmacogenetic and pharmacogenomic research. Pharmacogenomics J 2004; 4:9–16.

3

Pharmacogenetics of Drug Metabolism: Two Clinically Important Polymorphic Enzymes, CYP2D6 and TPMT

Georgios Panagiotidis and Leif Bertilsson

Division of Clinical Pharmacology, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

1. INTRODUCTION

The interindividual variation in the effect of drug treatment is pronounced. A major part of this is due to variation in drug metabolism, but also due to variation in drug targets such as receptors and transporters, which is evident from other chapters in this second edition of Pharmacogenomics. The rate of drug metabolism is regulated by environmental factors such as food intake and smoking habits, but genetic factors are also of utmost importance.

The pharmacogenetics of drug metabolism was established in the 1950s by the demonstration of the polymorphisms of plasma cholinesterase and *N*-acetyltransferase. The area of pharmacogenetics of drug metabolism has expanded tremendously and is now recognized in drug development and by drug regulatory authorities. This area of research has been reviewed extensively for example in the volume edited by Dr. Kalow: ''Pharmacogenetics of drug metabolism'' published in 1992 (1). There are many recent reviews to consult in this area (2–7). Table 1 gives same examples of firmly established polymorphic drug metabolizing enzymes and substrates. To cover the details of the individual polymorphic enzymes catalyzing drug metabolism, Refs. 1–7 are recommended. In the present review we will

Table 1 Some Examples of Firmly Established Polymorphic Drug Metabolizing Enzymes and Substrate Drugs

Polymorphic enzymes	Drug substrates
N-acetyltransferase 2 (NAT2)	Isoniazid, procainamide
Cytochrome P450 CYP2C9	S-Warfarin, losartan, phenytoin, tolbutamide, NSAIDs
Cytochrome P450 CYP2C19	Diazepam, omeprazole, proguanil
Cytochrome P450 CYP2D6	Most antidepressants and neuroleptics, antiarrhythmics
Thiopurine methyltransferase (TPMT)	6-Mercaptopurine, azathioprine

For individual references the reader is referred to recent books and reviews (1–6).

focus on two polymorphisms, which we consider both illustrative and clinically important, i.e., the cytochrome P450 CYP2D6 and the thiopurine methyltransferase (TPMT).

Until recently, polymorphisms in drug metabolism were discovered by phenotypic analysis detecting a bi-or trimodality of an index of enzymatic activity. This was the case for slow and rapid acetylators of isoniazid and extensive and poor metabolizers (EM and PM) of debrisoquine/sparteine. This area of research is a good example of what is now popularly called *Functional Genomics*. With the molecular genetics techniques available today single-nucleotide polymorphism (SNPs) are easily detected and later on their biochemical and clinical consequences may or may not be demonstrated. As an example of this, mutated alleles of *CYP1A2* (8,9) have recently been demonstrated, but their functional importance needs to be validated. A study by Özdemir et al. (10) strongly suggests a genetic component in the variability of the activity of the important enzymes CYP3A4 and CYP3A5. There is a clear polymorphism of the *CYP3A5* gene and the frequency of functional alleles is much higher in black Africans compared to Caucasians. Wong et al. (11) recently showed that the clearance of midazolam is higher in patients with the *CYP3A5*1/*3* than in *CYP3A5*3/*3* indicating an importance for the CYP3A5 enzyme. Similarly, we recently showed that the in vivo 3-hydroxylation of quinine is clearly related to the *CY3A5* genotype in Tanzanians (Mirghani et al., in preparation).

2. THE CYP2D6 POLYMORPHISM

2.1. The Discovery and Incidence of the PM Phenotype

British investigators (12,13) described in 1977 that the hydroxylation of the antihypertensive drug debrisoquine is polymorphic in nature. Independently Eichelbaum et al. in Germany (14) showed that the oxidation of sparteine

also is polymorphic. The metabolic ratios (MR; parent drug/metabolite) of the two drugs were closely correlated (15), showing that the same enzyme, now termed CYP2D6, is responsible for the two metabolic reactions.

The incidence of PMs of debrisoquine/sparteine has been investigated in many populations, in most of them with a fairly small number of subjects (16). Among 1011 Swedish Caucasians, we found 69 (6.3%) PMs of debrisoquine (Fig. 1) (17). This incidence is very similar to other European (16) and American (18) Caucasian populations. In collaboration with professor Lou and associates in Beijing, we showed that the incidence of PM among 695 Chinese was only 1.0% using the antimode MR = 12.6 established in Caucasian populations (Fig. 1) (17). A similar low incidence of PM has been shown in Japanese (18) and Koreans (19).

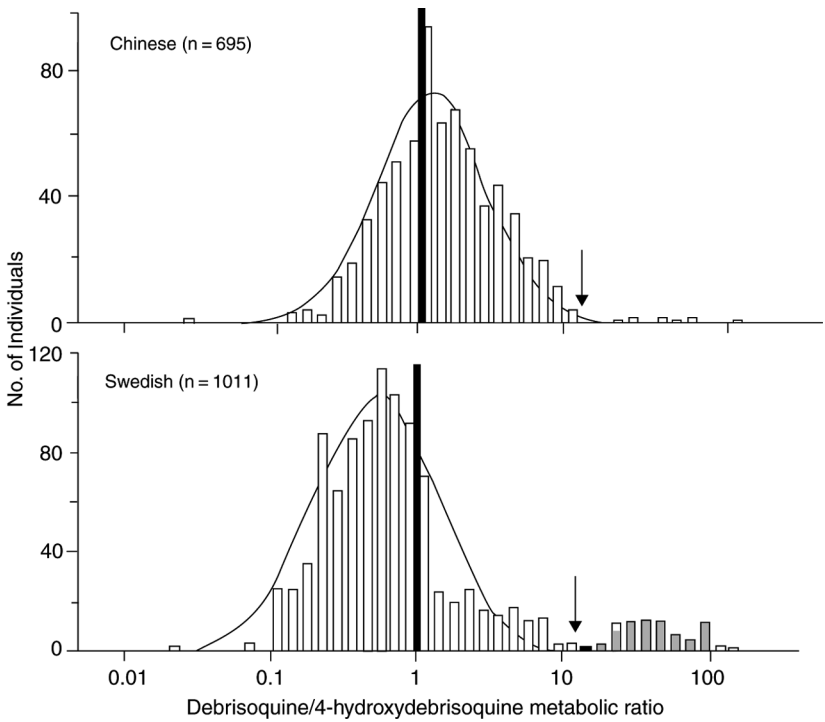


Figure 1 Distribution of the urinary debrisoquine/4-hydroxydebrisoquine metabolic ratio (MR) in 695 Chinese and 1011 Swedish healthy subjects. The *arrows* indicate MR = 12.6, the antimode between EM and PM established in Caucasians. A line is drawn at MR = 1. Most Chinese EM have MR > 1, while most Swedish EM have MR < 1. *Source*: From Ref. 17.

2.2. *CYP2D6* Alleles Causing Absent or Decreased Enzyme Activity

The gene encoding the CYP2D6 enzyme is localized on chromosome 22 (20). Using restriction fragment length polymorphism (RFLP) analysis and allele specific—polymerase chain reaction (PCR), three major mutant alleles were found in Caucasians (21–24). These are now termed *CYP2D6*3*, *CYP2D6*4,* and *CYP2D6*5* (25) (Table 2). In Swedish Caucasians the *CYP2D6*4* allele occurs with a frequency of 22% and accounts for more than 75% of the mutant alleles in this population (26). The *CYP2D6*4* allele is almost absent in Chinese and this is the reason for the low incidence of 1% PM in this population compared to 7% in Caucasians (17). The occurrence of the gene deletion (*CYP2D6*5*) is very similar, i.e., 4–6% (Table 2) in Sweden, China, and Zimbabwe. This shows that this is a very old mutation, which occurred before the separation of the three major races 100–150,000 years ago (2).

It is apparent in Figure 1 that the distribution of the MR of Chinese EMs is shifted to the right compared to Swedish EMs ($p < 0.01$) (17). Most Swedes have MR < 1, whereas the opposite is true for Chinese subjects. This shows that the mean rate of hydroxylation of debrisoquine is lower in Chinese EMs compared to Caucasian EMs (17). This right shift in MR in Orientals is due to the presence of a mutant *CYP2D6*10* allele at the high frequency of 51% in Chinese (27,28) (Table 2). The SNP C188T causes a Pro34Ser amino acid substitution giving an unstable enzyme with decreased catalytic activity (28). As shown in Figure 2 the presence of this C188T mutation causes the right shift among the investigated 152 Koreans (29).

Table 2 Frequency of Normal *CYP2D6*1* or **2* Alleles and Some Alleles Causing No or Deficient CYP2D6 Activity in Three Different Populations

CYP2D6 alleles	Functional mutation	Consequence	Allele frequency (%)		
			Swedish	Chinese	Zimbabwean
*1 or *2 (wt)			69	43	54
*3 (A)	A2637 del	Frame shift	21	0	0
*4 (B)	G1934A	Splicing defect	22	0–1	2
*5 (D)	Gene deletion	No enzyme	4	6	4
*10 (Ch)	C188T	Unstable enzyme	n.d.	51	6
*17 (Z)	C1111T	Reduced affinity	n.d.	n.d.	34

Abbreviations: n.d., not determined.
Source: From original publications (26–28,30) and reviews (2,3,7).

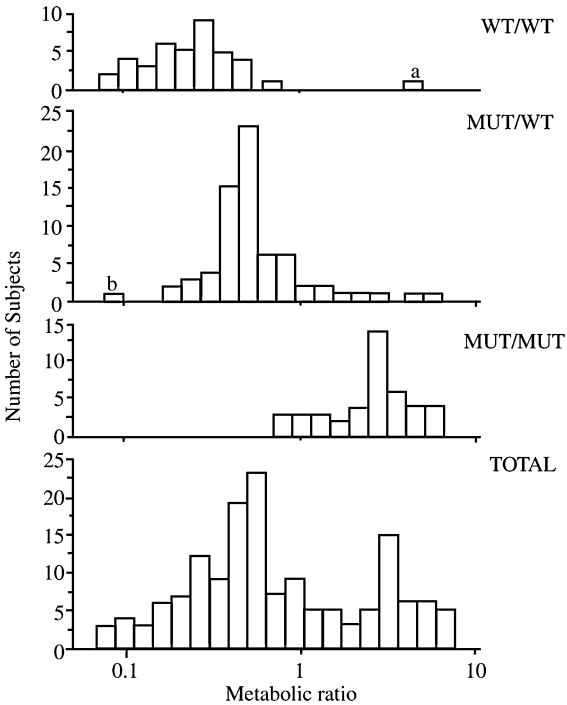


Figure 2 Distribution of the debrisoquine MR in three genotype groups related to the *CYP2D6*10* allele in 152 Korean subjects. Wt = *CYP2D6*1*(*or*2*) and mut = *CYP2D6*10*. *Source*: From Ref. 29.

The high frequency of this *CYP2D6*10* allele is similar in Chinese, Japanese, and Koreans.

Masimirembwa et al. (30) found a right shift of debrisoquine MR in black Zimbabweans similar to that found in Orientals. A mutated allele was subsequently identified and named *CYP2D6*17*, which encodes an enzyme with decreased debrisoquine hydroxylase activity. Among black Africans the frequency of this allele was found to be 34% in Zimbabweans (30; Table 2), 17% in Tanzanians (31), 28% in Ghanaians (32), and 9% in Ethiopians (33). This and many other studies demonstrate the genetic heterogeneity of populations in Africa. Wennerholm et al. (34) administered four different CYP2D6 substrates on separate occasions to Tanzanians with different genotypes. Subjects with the *CYP2D6*17/*17* had decreased rate of metabolism of debrisoquine and dextromethorphan, but normal metabolism of codeine and metoprolol. This demonstrates a changed substrate specificity of the *CYP2D6*17* encoded enzyme.

There are population specific *CYP2D6* alleles with the *4 in Caucasians with a C1934A mutation giving a splicing defect and thus no enzyme

is encoded. The *CYP2D6*10* and *CYP2D6*17* in Orientals and Africans, respectively, encode two different enzymes with decreased activity. In Caucasians and Orientals a close genotype and phenotype relationship has been demonstrated in several studies (26,28,29). However, in studies in Ethiopia (33), Ghana (32) as well as in Tanzania (31) a lower CYP2D6 activity in relation to genotype has been demonstrated indicating that environmental factors in Africa, e.g., infections or food intake are of importance in addition to genetic factors. Evidence for environmental influence on CYP2D6 catalyzed debrisoquine hydroxylation was demonstrated by comparing Ethiopians living in Ethiopia or in Sweden (35).

2.3. Gene Duplication, Multiduplication and Amplification as a Cause of Increased CYP2D6 Activity

The problem in treating PMs of debrisoquine with various drugs has been extensively discussed over the years since the discovery of the CYP2D6 polymorphism (15). Much less attention has been given to patients at the other extreme of the distribution of MR, i.e., ultrarapid hydroxylators. We described women with depression having an MR of debrisoquine of 0.07, who had to be treated with 500 mg of nortriptyline daily (36). This is three to five times higher than the recommended dose. Due to the fruitful collaboration with Prof. M. Ingelman-Sundbergs group, we could identify the molecular genetic basis for the ultrarapid metabolism in this patient and in another patient, who had to be treated with megadoses of clomipramine (37). These two patients had an *Xba*I 42 kb fragment containing two different functionally active *CYP2D6* genes in the *CYP2D* locus causing more enzyme to be expressed. The same year in 1993, a father and his daughter and son with 12 extra copies of the *CYP2D6* gene were described (37). This was the first demonstration of an inherited amplification of an active gene encoding a drug-metabolizing enzyme. These subjects were ultrarapid hydroxylators of debrisoquine with MR 0.01–0.02. The 12.1 kb fragment obtained by *Eco*RI RFLP analysis corresponds to the presence of a duplicated or multiduplicated *CYP2D6*2* gene (38). There are now also a few examples of duplicated *CYP2D6*1* and *CYP2D6*4* genes (39).

In Swedish Caucasians the frequency of subjects having duplicated/multiduplicated genes is about 1% (40). Going south in Europe the frequency increases being 3.6% in Germany (41), 7–10% in Spain (39,42), and 10% on Sicily in Italy (43). The frequency is as high as 29% in black Ethiopians (33) and 20% in Saudi Arabians (44). There is thus a European-African north–south gradient in the incidence of *CYP2D6* gene duplication. The high incidence among Ethiopians and Saudi Arabians indicates that the high incidence in Spain and Italy may have an ancestry in the Arabian conquest in the Mediterranean area (39). The high frequency of duplicated genes among Ethiopians might be the result of a dietary pressure favoring the preservation

of duplicated *CYP2D6* genes because of the ability of the enzyme to metabolize plant toxins including alkaloids (3,44).

Kawanish et al. (45) recently studied 81 depressed patients, who had failed to respond to antidepressant drugs, which are substrates of CYP2D6. With the hypothesis that there is an overrepresentation of ultrarapid metabolizers as a cause of non-response, *CYP2D6* gene duplication was analyzed. Of those 81 patients, eight had a gene duplication (9.9% and 95% confidence interval 3.4–16.4%) (45), which is higher than the 1% found in healthy Swedish subjects (40). These results suggest that the *CYP2D6* gene duplication is a possible factor that influences the development of persistence in depressed patients probably due to ultrarapid drug metabolism.

2.4. Metabolism of CYP2D6 Drug Substrates in Relation to Genotypes

Since the discovery of the CYP2D6 polymorphism in the 1970s almost 100 drugs have been shown to be substrates of this enzyme. Some of these drugs are shown in Table 3. The CYP2D6 substrates are all lipophilic bases. To study whether a drug is metabolized by CYP2D6 or not both in vitro and in vivo techniques may be employed. To establish the quantitative importance of this enzyme for the total metabolism of the drug, in vivo studies need to be performed. We will here give two examples with the tricyclic antidepressant nortriptyline and the neuroleptic haloperidol.

Nortriptyline was one of the first clinically important drugs to be shown to be metabolized by CYP2D6 (46,47). These early studies (prior to the era of genotyping) were performed in phenotyped panels of healthy subjects and the results have been confirmed in vivo in patients as well as

Table 3 Some Drugs Whose Metabolism Is Catalyzed by the CYP2D6 Enzyme, i.e., the Debrisoquine/Sparteine Hydroxylase

β-*Adrenoceptor blockers*	*Antidepressants*	*Neuroleptics*
Metoprolol	Amitriptyline	Haloperidol
Propranolol	Clomipramine	Perphenazine
Timolol	Desipramine	Risperidone
	Fluoxetine	Thioridazine
Antiarrhythmic drugs	Fluvoxamine	Zuclopenthixol
Encainide	Imipramine	
Flecainide	Mianserin	*Miscellaneous*
Perhexiline	Nortriptyline	Codeine
Propafenone	Paroxetine	Debrisoquine
Sparteine		Dextromethophan
		Phenformin
		Tramadol

in vitro using human liver microsomes and expressed enzymes. In a study by Dalén et al. (48), nortriptyline was given as a single oral dose to 21 healthy Swedish Caucasian subjects with different genotypes. As seen in Figure 3, there was a decrease in the plasma concentration of nortriptyline from subjects with 0 functional genes (five PMs with the *CYP2D6*4/*4* genotype) to those with 1 to 3 (gene duplication) functional genes (five in each group). The plasma concentrations of the parent drug were extremely low in one subject with 13 *CYP2D6* genes. This is the son in the family mentioned above (genotype *CYP2D6*2x13/*4*). The plasma concentrations of the formed metabolite 10-hydroxynortriptyline show the opposite pattern, i.e., highest concentrations in the subject with 13 genes and lowest in the PMs (Fig. 3, right). This study clearly shows the impact of the detrimental *CYP2D6*4* allele as well as the duplication/amplification of the *CYP2D6*2* gene on the metabolism of nortriptyline (48). A relationship between *CYP2D6* genotype and steady-state plasma concentration of nortriptyline and its hydroxy metabolite has been shown in Swedish depressed patients treated with the drug (49).

Using the same protocol as in the study of Dalén et al. in Caucasians (48), we investigated the influence of the Asian specific *CYP2D6*10* allele on the disposition of nortriptyline in Chinese subjects living in Sweden (50). Morita et al. (51) related the *CYP2D6*10* allele to steady-state plasma levels of nortriptyline and its metabolites in Japanese depressed patients. From these two studies it may be concluded that the Asian *CYP2D6*10* allele

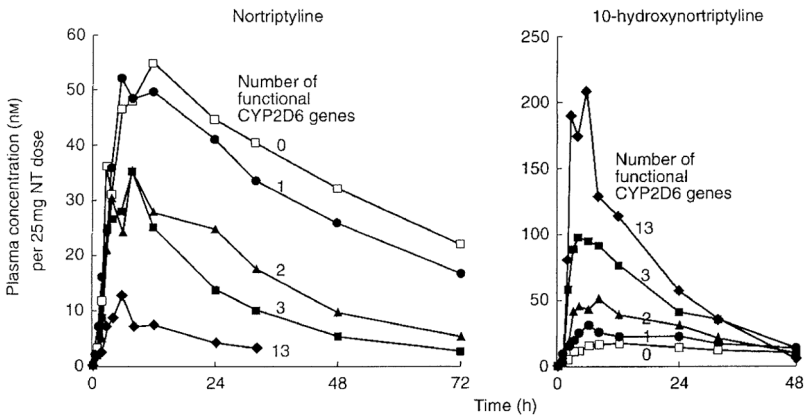


Figure 3 Mean plasma concentrations of nortriptyline (*left*) and 10-hydroxynortriptyline (*right*) in different genotype groups after a single oral dose of nortriptyline. The numerals close to the curves represent the number of functional *CYP2D6* genes in each genotype group. In groups with 0 to 3 functional genes, there were five subjects in each group. There was only one subject with 13 functional genes. *Source*: From Ref. 48.

encodes an enzyme with decreased activity to metabolize nortriptyline. This effect is less pronounced than the Caucasian specific *CYP2D6*4* allele, which encodes no enzyme at all. Genotyping of *CYP2D6* may be a tool to predict proper dosing of drugs such as nortriptyline in individual patients. It must, however, be remembered that there are population specific alleles.

Haloperidol will serve as a second example of an important drug substrate of CYP2D6. Llerena et al. (52) gave single oral doses of haloperidol to panels of six EMs and six PMs of debrisoquine. The PMs eliminated haloperidol significantly slower than EMs, the mean plasma half-life being longer (29.4 and 16.3 hr, respectively; $p < 0.01$) and the mean clearance lower (1.16 and 2.49 L hr^{-1} kg^{-1}, respectively; $p < 0.05$) (52). In a clinical study involving eight Caucasian patients with schizophrenia treated with depot haloperidol (as the decanoate), the dopamine D2 receptor occupancy was determined by positron emission tomography 1 and 4 weeks after intramuscular injection of the drug (53). One of the patients was genotypically a PM of debrisoquine. Of the group, he had the highest plasma concentration of haloperidol and also the highest D2 receptor occupancy.

Two studies from Hirosaki in Japan (54,55) have shown a relationship between increased haloperidol plasma concentrations and the presence of *CYP2D6*10* (and **5*) alleles in Japanese patients treated with oral doses of haloperidol. The dose used was 12 mg daily. In a study in Korea a relationship between haloperidol concentration and *CYP2D6* genotype was established in patients receiving less than 20 mg daily, but not in patients receiving higher doses (56). We believe that the high affinity-low capacity CYP2D6 is the predominant enzyme at low concentrations or low doses of haloperidol, while the low affinity-high capacity CYP3A4 becomes more important at higher doses. We could thus conclude that at least at low doses haloperidol is metabolized by CYP2D6. The metabolic pathway of haloperidol catalyzed by CYP2D6 is presently not known. Also other neuroleptics such as perphenazine, risperidone, thioridazine, and zuclopenthixol are metabolized by CYP2D6 (Table 3) (4).

2.5. Relationship Between CYP2D6 and Personality: An Endogenous Substrate in the Brain?

In early studies with healthy Swedish (57) and Spanish (58) Caucasian subjects, we found personality differences between EMs and PMs of debrisoquine. These differences have been confirmed in two studies published in 2004 (a) in depressed patients from New Zealand (59) and (b) in patients undergoing elective surgery in Malaysia (60). These four studies taken together suggest that CYP2D6 catalyzes the formation or elimination of endogenous substances important for certain functions in the central nervous system. The CYP2D6 is present in the brain, although with lower activity than in the liver (61). There seems to be an association (similar distribution) between dopamine transporter

and brain CYP2D6 (61,62) suggesting a role of the dopamine system as part of the CYP2D6—personality relationship as discussed by Llerena et al. (58). Recently, dopamine and serotonin have been found to be formed by CYP2D6 from tyramine (63) and 5-methoxytryptamine (64), respectively.

3. THE TPMT POLYMORPHISM

One of the most developed examples of clinical pharmacogenomics involves the polymorphism of thiopurine S-methyltransferase (TPMT) (EC 2.1.1.67).

The TPMT is a cytosolic enzyme whose precise physiological role is unknown. It catalyzes the S-methylation of the thiopurine agent azathioprine, 6-mercaptopurine, and 6-thioguanine using S-adenosine methionine as a methyl donor (65). Originally found in kidney and liver of rats and mice, it was subsequently shown to be present in most tissues including blood cells (66). Due to the correlation to TPMT activity in other tissues TPMT activity is measured in easily obtained erythrocytes (67). The TPMT activity is polymorphic and a trimodal distribution has been demonstrated in Caucasians (68). About one subject of 300 is homozygous for a defect *TPMT* allele with very low or absent enzyme activity. Eleven percent is heterozygous with an intermediate activity (68). The frequencies of loss of TPMT activity vary in different populations. The deficiency has been reported to occur to as low as 0.006–0.04% in Asian populations in contrast to the frequency of 0.3% in Caucasians (69).

The *TPMT* gene is located on chromosome 6 and includes 10 exons (70). *TPMT*3A*, the most common mutated allele, contains two point mutations in exons 7 (G460A and Ala154Thr) and 10 (A719G and Tyr240Lys). Two other alleles contain a single mutation, the first SNP (*TPMT*3B*) and the second SNP (*TPMT*3C*) (66). Aarbakke et al. (71) have reviewed the variant alleles of the *TPMT* gene and the relationship to TPMT deficiency. In Caucasians the *TPMT*3A* accounts for about 85% of mutated alleles, and in such populations the analysis of the known alleles may predict the phenotype (TPMT activity). In a Korean population the *TPMT*3A* was absent and the most common alleles was *TPMT*3C* (72,73). However, early investigations focused on allele-specific screening for only four alleles namely *TPMT*2*, *TPMT*3A*, *TPMT*3B*, and *TPMT*3C*, respectively. A restriction to selected alleles that are more frequent in one population means that important alleles may be overlooked in other populations. Due to the restricted methodology in the majority of studies investigating ethnic specific TPMT allele frequencies, continued studies in different populations involving full-gene sequencing or similar techniques seem necessary (74).

Azathioprine and 6-mercaptopurine are immunosuppressants that are used for the treatement of several conditions including immunological disorders and for prevention of acute rejection in transplant recipients. In Europe, azathioprine the precursor of 6-mercaptopurine has been the thio-

purine of choice in Inflammatory Bowel Disease (IBD), whereas in parts of North America 6-mercaptopurine is more commonly used. Furthermore, 6-mercaptopurine is more commonly used in acute lymphoblastic leukemia of childhood (69). Azathioprine is an imidazole derivative of, and metabolized non-enzymatically to, 6-mercaptopurine (Fig. 4). 6-Mercaptopurine is metabolized by several pathways one of which is catalyzed by TPMT and leads to inactive methyl-thiopurine metabolites. Other pathways catalyzed by several other enzymes leads to the active thioguanine nucleotides (6-TGN). The resulting 6-TGNs act as purine antagonists through their incorporation into the DNA molecule and subsequent prevention of DNA replication. The reduction in DNA replication suppresses various immunological functions in lymphocytes, T cells, and plasma cells (69). Numerous studies have been shown that TPMT-deficient patients are at very high risk of developing severe hematopoietic toxicity, if treated with conventional doses of thiopurines (75). At high concentrations, 6-TGNs may cause toxicity and bone marrow suppression (result of low TPMT activity). On the other hand low concentrations (result of high TPMT activity) may give an increased risk of therapeutic failure and due to other metabolite accumulation, such as 6-methylmercaptopurine nucleotides, also to liver toxicity (Fig. 4). Other less serious side effects of azathioprine are gastrointestinal symptoms such as nausea and vomiting. These side effects or azathioprine intolerance is not clearly associated to TPMT activity or metabolite levels. A number of

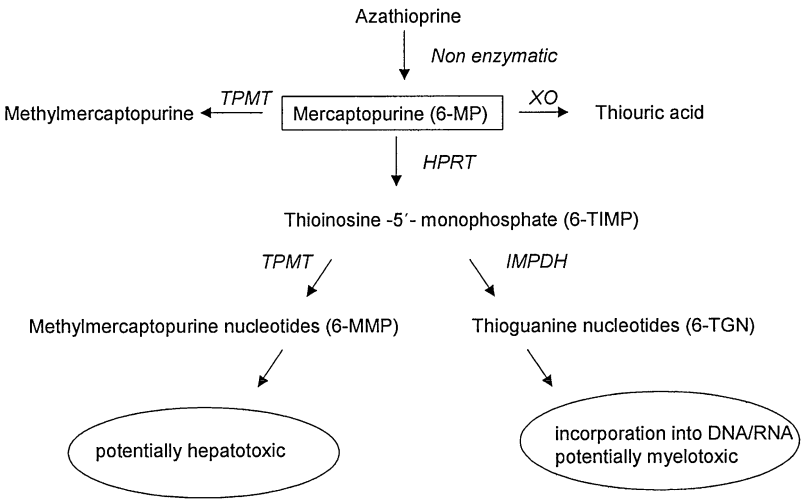


Figure 4 Pathway of thiopurine metabolism. *Abbreviations*: TPMT, thiopurine methyltransferase; XO, xanthineoxidase; HPRT, hypoxanthine guanine phosphoribosyltransferase; IMPDH, inosine monophosphate dehydrogenase.

studies have shown pre-treatment TPMT status testing to be cost-effective and a reliable way of predicting life-threatening bone marrow toxicity. Many authors, including us, are of the opinion that TPMT status testing should be incorporated in routine clinical practice to avoid severe adverse drug reactions and to adjust dosing in patients identified with intermediate TPMT activity. Another important issue part from avoiding adverse effects is of course the treatment effect. Several studies have shown a relationship between therapeutic effects and TPMT activity or 6-TGN concentrations in red blood cells. However, more clinical studies are needed to establish therapeutic intervals for the various conditions in which these drugs are used. So far most drug-effect studies are focused on 6-TGN concentrations. The 6-TGNs are a group of several metabolites. Other enzymes and metabolites are also involved in the complex metabolism of thiopurines. Thus, there might be other unknown factors involved in the metabolism and action of thiopurine drugs. The question is if there are other metabolites and enzymes with better correlation to treatment outcome to focus on in further studies?

In conclusion, low TPMT activity due to TPMT polymorphism can lead to severe myelosuppression in patients treated with thiopurines such as 6-mercaptopurine. The pre-treatment TPMT status of patients can be measured by phenotype (TPMT activity) or genotype. The TPMT genotype is, however, uncertain in view of the difficulties in interpreting the consequences of novel polymorphism detection and the chance of missing clinically relevant allelic variation in different racial groups. Furthermore, standard genotyping techniques cannot as yet, predict those individuals with very high TPMT activities who may not respond to standard doses of azathioprine or 6-mercaptopurine. Thus, further genotype–phenotype correlation studies are needed as well as further drug-effect studies where relevant metabolites are monitored.

4. FUTURE PERSPECTIVES ON THE PHARMACOGENETICS OF DRUG METABOLISM

The pronounced interindividual variation in the rate of drug metabolism has been known for many years. It was initially only of an academic interest, but today the pharmaceutical industry has to document the metabolism of a novel drug in development before registration. This is a requirement from the Food and Drug Administration and similar authorities. The knowledge of how a drug is metabolized and which enzymes are involved may help to predict drug–drug interactions and how fast an individual patient may metabolize a specific drug.

With the high troughput screening techniques available today pharmaceutical industries may screen for drug candidates, which are metabolized and/or are inhibitors of, e.g., CYP2D6. Such drugs are in the hands of

many companies not further developed. If this had been the case in the past, we would not have had many of the important drugs available today (Table 3). We see a danger in deleting potential drugs from development, if CYP2D6 is involved in its metabolism. The novel antimuscarinic drug tolterodine may serve as an example of a drug metabolized by CYP2D6, but not discontinued for development. It is now used as a valuable drug for the treatment of urinary incontinence. In early phase I studies, it was discovered that tolterodine is hydroxylated by CYP2D6, but the surrogate antimuscarinic effect on salivation was the same in EM and PM (76). The reason is that 5-hydroxy-tolterodine is an active metabolite and responsible for some of the effect in EM. After some well-planned studies (77), this drug was soon approved in both Europe and in the United States.

Phenotype analysis with, e.g., debrisoquine or sparteine may be performed to predict a proper dosage schedule before starting treatment with a CYP2D6 substrate drug. Phenotyping should be avoided during treatment as many drugs especially psychotropic drugs such as imipramine and levomepromazine may inhibit the test drug (78). Genotyping may of course be performed independent of ongoing drug treatment. In many cases a nice genotype–phenotype relationship has been established, but as pointed out above, interethnic differences must be considered. With the development of novel less expensive genotyping methods, the use of such techniques will increase in the future. The measurement of the active species of a drug in plasma, i.e., therapeutic drug monitoring is maybe the best tool to optimize prescription in the individual patient. We consider TPMT to be the most well-documented polymorphic drug metabolizing enzyme, which requires personalized medication. In this case the best is to determine TPMT activity in red blood cells, but active thioguanine nucleotides may be measured as an alternative. Genotyping might here be a valuable tool in the future.

The need for personalized medication has been discussed in many articles. One often quoted is by Lazarou et al. (79), who performed a meta-analysis of the incidence of adverse drug reactions (ADR). They found the extremely high incidence of serious ADRs of 6.7% and of fatal ADRs of 0.32% among hospitalized patients in the United States. This means that about 100,000 U.S. citizens die every year from drug side effects and about two millions get seriously ill because of drug intake. At least part of these patients might have had low activity of polymorphic drug metabolizing enzymes such as CYP2D6 or TPMT, but this can only be speculated upon. Partial support for such a speculation comes from a pilot study by de Leon et al. (80). They determined the *CYP2D6* genotype in 100 consecutive patients in a psychiatric hospital in Kentucky and found that 14% were PM, which is twice that of the U.S. population (7%). The patients with CYP2D6 deficiency also appeared more likely to experience side effects in response to CYP2D6 medications (80).

ACKNOWLEDGMENTS

The studies performed in the authors' laboratory were supported by the Swedish Research Council Medicine (3902), the National Institutes of Health (R01 GM 60548) and Karolinska Institutet.

REFERENCES

1. Kalow W. Pharmacogenetics of Drug Metabolism. New York: Pergamon Press Inc., 1992:1–897.
2. Bertilsson L, Dahl M-L, Dalén P, Sjurbaji AAL-. Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs. Br J Clin Pharmacol 2002; 53:111–122.
3. Ingelman-Sundberg M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. Trends Pharmacol Sci 2004; 25:193–200.
4. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999; 286:487–491.
5. Weinshilboum R. Inheritance and drug response. N Engl J Med 2003; 348: 529–537.
6. Wolf CR, Smith G, Smith RL. Clinical review. Pharmacogenetics BMJ 2000; 320:987–990.
7. Bertilsson L, Kalow W. Interethnic differences in drug disposition and effects. In: Pacifici GM, Pelkonen O, eds. Interindividual Variability in Human Drug Metabolism. London & New York: Taylor & Francis, 2001:15–74.
8. Sachse C, Brockmöller J, Bauer S, Roots I. Functional significance of C-A polymorphism in intron 1 of the cytochrome P450 *CYP1A2* gene tested with caffeine. Br J Clin Pharmacol 1999; 47:445–449.
9. Aklillu E, Carrillo JA, Makonnen E, Hellman K, Pitarque M, Bertilsson L, Ingelman-Sundberg M. Genetic polymorphism of CYP1A2 in Ethiopians affecting induction and expression. Characterization of novel haplotypes with SNPs in intron 1. Mol Pharmacol 2003; 64:659–669.
10. Özdemir V, Kalow W, Tang B-K, Paterson AD, Walker SE, Endrenyi L, Kashuba APM. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. Pharmacogenetics 2000; 10: 373–388.
11. Wong M, Balleine RL, Collins M, Liddle C, Clarke CL, Gurney H. CYP3A5 genotype and midazolam clearance in Australian patients receiving chemotherapy. Clin Pharmacol Ther 2004; 75:529–538.
12. Mahgoub A, Idle JR, Dring DG, Lancaster R, Smith RL. Polymorphic hydroxylation of debrisoquine in man. Lancet 1977; 2:584–586.
13. Tucker GT, Silas JH, Iyun AO, Lennard MS, Smith AJ. Polymorphic hydroxylation of debrisoquine in man. Lancet 1977; 2:718.
14. Eichelbaum M, Spannbrucker N, Steinke B, Dengler HJ. Defective *N*-oxidation of sparteine in man: a new pharmacogenetic defect. Eur J Clin Pharmacol 1979; 16:183–187.

15. Eichelbaum M, Bertilsson L, Säwe J, Zekorn C. Polymorphic oxidation of sparteine and debrisoquine. Related Pharmacogenetic entities. Clin Pharmacol Ther 1982; 31:184–186.
16. Eichelbaum M, Gross AS. The genetic polymorphism of debrisoquine/sparteine metabolism—clinical aspects. In: Kalow W, ed. Pharmacogenetics of drug metabolism. New York: Pergamon Press Inc, 1992:625–648.
17. Bertilsson L, Lou YQ, Du YL, Liu Y, Kuang T-Y, Liao XM, Wang KY, Reviriego J, Iselius L, Sjöqvist F. Pronounced differences between native Chinese and Swedish populations in the polymorphic hydroxylations of debrisoquine and S-mephenytoin. Clin Pharmacol Ther 1992; 51:388–397.
18. Nakamura K, Goto F, Ray WA, McAllister CB, Jacqz E, Wilkinson GB, Branch RA. Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. Clin Pharmacol 1985; 38:402–408.
19. Sohn D-R, Shin S-G, Park C-W, Kusaka M, Chiba K, Ishizaki T. Metoprolol oxidation polymorphism in a Korean population: comparison with native Japanese and Chinese populations. Br J Clin Pharmacol 1991; 32:504–507.
20. Eichelbaum M, Baur MP, Dengler HJ, Osikowska-Evers BO, Tieves G, Zekorn C, Rittner C. Chromosomal assignment of human cytochrome P450 (debrisoquine/sparteine type) to chromosome 22. Br J Clin Pharmacol 1987; 23:455–458.
21. Skoda RC, Gonzalez FJ, Demierre A, Meyer UA. Two mutant alleles of the human cytochrome P450 db1 gene (P450 II D1) associated with genetically deficient metabolism of debrisoquine and other drugs. Proc Natl Acad Sci USA 1988; 85:5240–5243.
22. Gaedigk A, Blum M, Gaedigk R, Eichelbaum M, Meyer UA. Deletion of the entire cytochrome P450 CYP2D6 gene as a cause of impaired drug metabolism in poor metabolizers of the debrisoquine/sparteine polymorphism. Am J Hum Genet 1991; 48:943–950.
23. Heim HM, Meyer UA. Genotyping of poor metabolisers of debrisoquine by allele-specific PCR amplification. Lancet 1990; 336:529–532.
24. Heim HM, Meyer UA. Evolution of a high polymorphic human cytochrome P450 gene cluster: CYP2D6. Genomics 1992; 14:49–58.
25. Daly AK, Brockmöller J, Broly F, Eichelbaum M, Evans WE, Gonzalez FJ, Huang J-D, Idle JR, Ingelman-Sundberg M, Ishizaki T, Jacqz-Aigrain E, Meyer UA, Nebert DW, Steen VM, Wolf CR, Zanger UM. Nomenclature for human *CYP2D6* alleles. Pharmacogenetics 1996; 6:193–201.
26. Dahl M-L, Johansson I, Porsmyr Palmertz M, Ingelman-Sundberg M, Sjöqvist F. Analysis of the CYP2D6 gene in relation to debrisoquin and desipramine hydroxylation in a Swedish population. Clin Pharmacol Ther 1992; 51:12–17.
27. Wang SL, Huang JD, Lai MD, Liu B-H, Lai M-L. Molecular basis of genetic variation in debrisoquine hydroxylation in Chinese subjects: polymorphism in RFLP and DNA sequence of CYP2D6. Clin Pharmacol Ther 1993; 53: 410–418.
28. Johansson I, Oscarsson M, Yue Q-Y, Bertilsson L, Sjöqvist F, Ingelman-Sundberg M. Genetic analysis of the Chinese CYP2D locus. Characterization

of variant CYP2D6 genes present in subjects with diminished capacity for debrisoquine hydroxylation. Mol Pharmacol 1994; 46:452–459.
29. Roh H-K, Dahl M-L, Johansson I, Ingelman-Sundberg M, Cha Y-N, Bertilsson L. Debrisoquine and S-mephenytoin hydroxylation phenotypes and genotype in a Korean population. Pharmacogenetics 1996; 6:441–447.
30. Masimirembwa C, Persson I, Bertilsson L, Hasler J, Ingelman-Sundberg M. A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a black African population: association with diminished debrisoquine hydroxylase activity. Br J Clin Pharmacol 1996; 42:713–719.
31. Wennerholm A, Johansson I, Massele AY, Jande M, Alm C, Aden Abdi Y, Dahl M-L, Ingelman-Sundberg M, Bertilsson L, Gustafsson LL. Decreased capacity for debrisoquine metabolism among black Tanzanians: analyses of the *CYP2D6* genotype and phenotype. Pharmacogenetics 1999; 9: 707–714.
32. Griese E-U, Asante-Poku S, Ofori-Adjei D, Mikus G, Eichelbaum M. Analysis of the *CYP2D6* gene mutations and their consequences for enzyme function in a West African population. Pharmacogenetics 1999; 9:715–723.
33. Aklillu E, Persson I, Bertilsson L, Johansson I, Rodriguez F, Ingelman-Sundberg M. Frequent distribution of ultrarapid metabolizers of debrisoquine in an Ethiopian population carrying duplicated and multiduplicated functional *CYP2D6* alleles. J Pharmacol Exp Ther 1996; 278:441–446.
34. Wennerholm A, Dandara C, Sayi J, Svensson J-O, Aden Abdi Y, Ingelman-Sundberg M, Bertilsson L, Hasler J, Gustafsson LL. The African-specific *CYP2D6***17* allele encodes an enzyme with changed substrate specificity. Clin Pharmacol Ther 2002; 71:77–88.
35. Aklillu E, Herrlin K, Gustafsson LL, Bertilsson L, Ingelman-Sundberg M. Evidence for environmental influence on CYP2D6-catalysed debrisoquine hydroxylation as demonstrated by phenotyping and genotyping of Ethiopians living in Ethiopia or in Sweden. Pharmacogenetics 2002; 12:375–383.
36. Bertilsson L, Åberg-Wistedt A, Gustafsson LL, Nordin C. Extremely rapid hydroxylation of debrisoquine—a case report with implication for treatment with nortriptyline and other tricyclic antidepressants. Ther Drug Monit 1985; 7:478–480.
37. Bertilsson L, Dahl M-L, Sjöqvist F, Åberg-Wistedt A, Humble M, Johansson I, Lundqvist E, Ingelman-Sundberg M. Molecular basis for rational megaprescribing in ultrarapid hydroxylators of debrisoquine. Lancet 1993; 341:63.
38. Johansson I, Lundqvist E, Bertilsson L, Dahl M-L, Sjöqvist F, Ingelman-Sundberg M. Inherited amplification of an active gene in the cytochrome P450 CYP2D-locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad Sci USA 1993; 90:11825–11829.
39. Bernal ML, Sinues B, Johansson I, McLellan RA, Wennerholm A, Dahl M-L, Ingelman-Sundberg M, Bertilsson L. Ten percent of North Spanish individuals carry duplicated or triplicated *CYP2D6* genes associated with ultrarapid metabolism of debrisoquine. Pharmacogenetics 1999; 9:657–660.
40. Dahl M-L, Johansson I, Bertilsson L, Ingelman-Sundberg M, Sjöqvist F. Ultrarapid hydroxylation of debrisoquine in a Swedish population. Analysis of the molecular genetic basis. J Pharmacol Exp Ther 1995; 274:516–520.

41. Sachse C, Brockmöller J, Bauer S, Roots I. Cytochrome P450 2D6 variants in a Caucasian population: allele frequencies and phenotypic consequences. Am J Hum Genet 1997; 60:284–295.
42. Agúndez JAG, Ledesma MC, Ladero JM, Benitez J. Prevalence of *CYP2D6* gene duplication and its repercussion on the oxidative phenotype in a white population. Clin Pharmacol Ther 1995; 57:265–269.
43. Scordo MG, Spina E, Facciolá G, Avenoso A, Johansson I, Dahl M-L. Cytochrome P450 2D6 genotype and steady state plasma levels of risperidone and 9-hydroxy risperidone. Psychopharmacology 1999; 147:300–305.
44. McLellan RA, Oscarson M, Seidegård JE, Evans DAP, Ingelman-Sundberg M. Frequent occurrence of *CYP2D6* gene duplication in Saudi Arabians. Pharmacogenetics 1997; 7:187–191.
45. Kawanish C, Lundgren S, Ågren H, Bertilsson L. Increased incidence of CYP2D6 gene duplication in patients with persistent mood disorders: ultrarapid metabolism of antidepressants as a cause of nonresponse. A pilot study. Eur J Clin Pharmacol 2004; 59:803–807.
46. Bertilsson L, Eichelbaum M, Mellström B, Säwe J, Schulz HU, Sjöqvist F. Nortriptyline and antipyrine clearance in relation to debrisoquine hydroxylation in man. Life Sci 1980; 27:1673–1677.
47. Mellström B, Bertilsson L, Säwe J, Schulz HU, Sjöqvist F. E- and Z-hydroxylation of nortriptyline in man—relationship to polymorphic hydroxylation of debrisoquine. Clin Pharmacol Ther 1981; 30:189–193.
48. Dalén P, Dahl M-L, Bernal Ruiz ML, Nordin J, Bertilsson L. 10-Hydroxylation of nortriptyline in Caucasians with 0, 1, 2, 3, and 13 functional CYP2D6 genes. Clin Pharmacol Ther 1998; 63:444–452.
49. Dahl M-L, Bertilsson L, Nordin C. Steady-state plasma levels of nortriptyline and its 10-hydroxy metabolite: relationship to the *CYP2D6* genotype. Psychopharmacology 1996; 123:315–319.
50. Yue Q-Y, Zhong Z-H, Tybring G, Dalén P, Dahl M-L, Bertilsson L, Sjöqvist F. Pharmacokinetics of nortriptyline and its 10-hydroxy metabolite in Chinese subjects of different CYP2D6 genotypes. Clin Pharmacol Ther 1998; 64:384–390.
51. Morita S, Shimoda K, Someya T, Yoshimura Y, Kamijima K, Kato N. Steady-state plasma levels of nortriptyline and its hydroxylated metabolites in Japanese patients: impact of *CYP2D6* genotype on the hydroxylation of nortriptyline. J Clin Psychopharmacol 2000; 20:141–149.
52. Llerena A, Alm C, Dahl M-L, Ekqvist B, Bertilsson L. Haloperidol disposition is dependent of debrisoquine hydroxylation phenotype. Ther Drug Monit 1992; 14:92–97.
53. Nyberg S, Farde L, Halldin C, Dahl M-L, Bertilsson L. D2 dopamine receptor occupancy during low-dose treatment with haloperidol decanoate. Am J Psychiatry 1995; 152:173–178.
54. Suzuki A, Otani K, Mihara K, Yasui N, Kaneko S, Inoue Y, Hayashi K. Effects of the CYP2D6 genotype on the steady-state plasma concentrations of haloperidol and reduced haloperidol in Japanese schizophrenic patients. Pharmacogenetics 1997; 7:415–418.
55. Mihara K, Suzuki A, Kondo T, Yasui N, Furukori H, Nagashima U, Otani K, Kaneko S, Inoue Y. Effects of the *CYP2D6*10* allele on the steady-state plasma

concentrations of haloperidol and reduced haloperidol in Japanese patients with schizophrenia. Clin Pharmacol Ther 1999; 65:291–294.
56. Roh H-K, Chung J-Y, Oh D-Y, Park C-S, Svensson J-O, Dahl M-L, Bertilsson L. Plasma concentrations of haloperidol are related to the CYP2D6 genotype at low, but not high doses of haloperidol in Korean schizophrenic patients. Br J Clin Pharmacol 2001; 52:265–271.
57. Bertilsson L, Alm C, de las Carreras C, Edman G, Schalling D, Widen J. Debrisoquine hydroxylation polymorphism and personality. Lancet 1989; I:555.
58. Llerena A, Edman G, Cobaleda J, Benitez J, Schalling D, Bertillsson L. Relationship between personality and debrisoquine hydroxylation capacity—suggestion of an endogenous neuroactive substrate or product of the cytochrome P4502D6. Acta Psychatr Scand 1993; 87:23–28.
59. Roberts RL, Luty SE, Mulder RT, Joyce PR, Kennedy MA. Association between cytochrome P450 2D6 genotype and harm avoidance. Am J Med Genet Part B Neuropsychiatr Genet 2004; 127B:90–93.
60. Gan SH, Ismail R, Wan Adnan WA, Zulmi W, Kumaraswamy N, Larmie ET. Relationship between types A and B personality and debrisoquine hydroxylation capacity. Br J Clin Pharmacol 2004; 57:785–789.
61. Kalow W, Tyndale RF. Debrisoquine/sparteine monooxygenase and other P450s in brain. In: Kalow W, ed. Pharmacogenetics of Drug Metabolism. New York: Pergamon Press, 1992:649–656.
62. Niznik HB, Tyndale RF, Sallee RF, Gonzalez FJ, Hardnick JP, Inaba T, Kalow W. The dopamine transporter and cytochrome P450IID1 (debrisoquine 4-hydroxylase) in brain: resolution and identification of two distinct [3H] GBR-12935 binding proteins. Arch Biochem Biophys 1990; 276:424–432.
63. Hiroi T, Imaoka S, Funae Y. Dopamine formation from tyramine by CYP2D6. Biochem Biophys Res Commun 1998; 249:838–843.
64. Yu AM, Idle JR, Byrd LG, Krausz KW, Kupfer A, Gonzales FJ. Regeneration of serotonin from 5-methoxytryptamine by polymorphic human CYP2D6. Pharmacogenetics 2003; 13:173–181.
65. Woodson LC, Ames MM, Selassie CD, Hansch C, Weinshilboum RM. Thiopurine methyltransferase: aromatic thiol substrates and inhibition by benzioc acid derivatives. Mol Pharmacol 1983; 24:471–478.
66. Krynetski EY, Evans WE. Genetic polymorphism of thiopurine S-methyltransferase: molecular mechanisms and clinical importance. Pharmacology 2000; 61:136–146.
67. Krynetski EY, Evans WE. Pharmacogenetics of cancer therapy: getting personal. Am J Hum Genet 1998; 63:11–16.
68. Weinshilboum RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. Am J Hum Genet 1980; 32:651–662.
69. Baker DE. Pharmacogenomics of azathioprine and 6-mercaptopurine in gastroenterologic therapy. Gastroenterol Disord 2003; 3:150–157.
70. Szumlanski C, Otterness D, Her C, Lee D, Brandriff B, Kelsell D. Thiopurine methyltransferase pharmacogenetics: human gene cloning and characterization of a common polymorphism. DNA Cell Biol 1996; 15:17–30.

71. Aarbakke J, Janka-Schaub G, Elion GB. Thiopurine biology and pharmacology. Trends Pharmacol Sci 1997; 18:3–7.
72. Lennard L. Clinical implications of thiopurine methyltransferase—Optimization of drug dosage and potential drug interactions. Ther Drug Monit 1998; 20: 527–531.
73. Otterness D, Szumlanski C, Lennard L, Klemetsdal B, Aarbakke J, Park-Hah JO, Iven H, Schmeigelow K, Branum E, Brien J O', Weinshilboum R. Human thiopurine methyltransferase pharmacogenetics: gene sequence polymorphisms. Clin Pharmacol Ther 1997; 62:60–73.
74. van Aken J, Schmedders M, Feuerstein G, Kollek R. Prospects and limits of pharmacogenetics. The thiopurine methyl transferase (TPMT) experience. Am J Pharmacogenomics 2003; 3:149–155.
75. McLeod HL, Yu J. Cancer pharmacogenomics: SNPs, chips, and the individual patient. Cancer Invest 2003; 21:630–640.
76. Brynne N, Dalén P, Alván G, Bertilsson L, Gabrielsson J. The influence of CYP2D6 polymorphism on the pharmacokinetics and dynamics of tolterodine. Clin Pharmacol Ther 1998; 63:529–539.
77. Brynne N. Consequences of CYP2D6 Polymorphism for the Disposition and Dynamics of Tolterodine. Ph.D. dissertation, Karolinska Institutet, Stockholm, 1998.
78. Brynne N, Böttiger Y, Hallén B, Bertilsson L. Tolterodine does not affect the human in vivo metabolism of the probe drugs caffeine, debrisoquine and omeprazole. Br J Clin Pharmacol 1999; 47:145–150.
79. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients. JAMA 1998; 279:1200–1205.
80. de Leon J, Barnhill J, Rogers T, Boyle J, Chou W-H, Wedlund PJ. Pilot study of the cytochrome P450-2D6 genotype in a psychiatric state hospital. Am J Psychiatry 1998; 155:1278–1280.

4

Receptors

Wendell W. Weber

Department of Pharmacology, University of Michigan, Ann Arbor, Michigan, U.S.A.

1. INTRODUCTION

Most responses of higher organisms to exogenous chemicals including drugs are receptor mediated. It follows that information about variant forms of receptor proteins is central to understanding human sensitivity to these substances. Although hereditary changes in drug metabolizing enzymes are foremost historically in the annals of pharmacogenetics, investigators inferred long ago that genetic changes in receptors probably accounted for several disorders of pharmacogenetic interest such as the long QT syndrome, malignant hyperthermia, and insulin resistance.

Structural diversity among receptors was first recognized by Ahlquist (1) who observed that different adrenergic antagonists imparted two distinct patterns of pharmacologic response within animal tissues. From tissue responses to agonists such as dichloroisoproterenol and antagonists such as propranolol, phenoxybenzamine, and phentolamine, he deduced the existence of α- and β-forms of adrenergic receptors. Other investigators proceeded to deduce the presence of more than one subtype of receptor in tissues such as the $\beta 1$ adrenergic receptor of cardiac tissue and the $\beta 2$ adrenergic receptor of bronchial smooth muscle by probing tissues with additional agonists and antagonists of different specificities and affinities.

Investigators relied on the properties of pharmacological probes for many years to identify and distinguish different types of receptors and

receptor subtypes, but they realized that the pharmacological approach by itself did not yield important details about the molecular basis of receptor heterogeneity. With the widespread adoption of molecular technologies that began in the 1980s that grew to encompass restriction analysis, dot blot and slot blot analysis, polymerase chain reaction, microarray analysis, and other techniques for manipulating DNA and RNA molecules, cloning, sequencing, and site-directed mutagenesis of genes became the norm in many laboratories.

Pharmacological approaches combined with molecular techniques have advanced the structural and functional analyses of receptor gene families at a rapid pace. The goal in receptor pharmacogenetics is to gain an understanding of how receptors modulate human drug response by studying the influence of variations in DNA sequences of the receptor genes and of important proteins they encode. Many receptor gene families have been scrutinized intensely during the past few years, and understanding of their structure and function has progressed. On the other hand, finding gaps in the interpretation of genetically polymorphic variations of receptor systems for lack of knowledge about the biological consequences of many polymorphisms is not surprising because the field is still in the discovery stage. Searches of the literature reveal (Sec. 1.1) the amount of information regarding genetically polymorphic receptor systems has grown so rapidly and so large that a comprehensive account of the pharmacogenetics for every receptor that has been studied is impractical. The purpose of this chapter is to convey an idea of the current state and future directions for research by focusing principally on one or two predominant features of pharmacogenetic interest for a few extensively studied receptor families.

1.1. Criteria for Selection of Receptor Polymorphisms

More than 17,000 citations for genetic studies of receptors and 8074 citations for receptor polymorphisms were found for 1975–2003 (Table 1). The fraction of human receptor genetic citations relative to citations for all species has risen steadily from about half ($363/694 = 0.52$) in 1986–1988 to about two-thirds ($3605/5342 = 0.67$) currently (2001–2003). The fraction of citations that deals with human receptor polymorphisms stood at about three-fourths those for all species in the mid-1980s ($170/229 = 0.74$), but now includes more than 90% that for all studies. The number of human receptor polymorphism citations has increased more than 16-fold ($2753/170 = 16.2$) during the same interval (1986–2003) (Table 2).

Several receptor-related traits that illustrate the depth and scope of knowledge of genetic variation inherent in human drug response have been selected for discussion. Discussions emphasize but do not treat exhaustively the genetic aspects, molecular basis, and significance of each trait selected.

Table 1 Medline Literature Citations on Receptor Genetics and Receptor Polymorphisms Inclusive of 1975–2003

	Medline textwords (tw) searched[a]			
	Receptor genetics		Receptor polymorphisms	
Citations	All species	Human	All species	Human
1975–1985	74		29	
1986–1988	694	363	229	170
1989–1991	1,232	722	443	364
1992–1994	1,997	1,231	785	687
1995–1997	3,076	1,987	1,324	1,207
1998–2000	4,792	3,259	2,358	2,206
2001–2003	5,342	3,605	2,935	2,753
Total 1986–2003	17,133	11,167	8,074	7,387

[a]The combinations of (receptor$ and genetic$). tw. and (receptor$ and polymorphi$). tw. were searched on the Medline database from 1975 to 2003. From 1986 to 2003 searches were made at 3-year intervals.

1.2. Highlights of Receptor Pharmacogenetics

Highlights attending receptor pharmacogenetics research are presented chronologically in Table 3. It is evident from the record that the advent and widespread application of recombinant DNA technology has greatly

Table 2 Five years (1999–2003) of Medline Literature Citations for the Most Prevalent Polymorphic Human Receptors of Pharmacogenetic Interest

Receptor	1999	2000	2001	2002	2003	$\sum(1999-2003)$
Nuclear (*cytoplasmic*)						
Vitamin D	61	73	63	78	39	314
Androgen	44	49	69	55	36	253
Estrogen	28	43	42	63	34	210
Sub total	133	165	174	196	109	777
Cell surface						
Dopamine (D2)	12	22	23	16	23	96
β_2-adrenergic	3	15	12	18	8	56
Angiotensin II	18	4	13	12	5	52
Serotonin 5-HT2A	7	15	13	9	6	50
β_3-adrenergic	8	2	10	21	4	45
Subtotal	48	58	71	76	46	299
Total	181	223	245	272	155	1,076
Nuclear/cell surface	2.8	2.8	2.4	2.6	2.4	2.6

Table 3 A Chronology of Human Receptor Pharmacogenetics

Year	Event	References
1960	Malignant hyperthermia identified in a human kindred	2
1964	Coumarin anticoagulant resistance identified in a human kindred	3
1977	APL[a] attributed to balanced translocation t (15;17)	4
1985	Insulin receptor gene cloned and mapped to chromosome 19p 13.2–13.3	5
1985	Glucocorticoid receptor gene cloned	6
1986	Estrogen receptor gene (ESR1) cloned and mapped to chromosome 6q25.1	7–9
1987	β2-Adrenoceptor cloned, sequenced and mapped	10
	Mineralocorticoid receptor gene cloned	11
	Retinoic acid receptor gene cloned	12
1988	Vitamin D receptor gene clone and sequenced	13
	Vitamin D resistant rickets associated with mutant Vitamin D receptor	14
	Estrogen receptor gene (ESR1) genomic structure determined	15
	Androgen receptor gene cloned; androgen insensitivity associated with mutant androgen receptor	16 and 17
	Severe insulin resistance associated with mutant insulin receptor	18 and 19
	Retinoic acid receptor localized to chromosome 17q21	20
1990	FcεRI receptor (α-chain) mapped to chromosome 1q23	21
1991	Malignant hyperthermia associated with mutant ryanodine receptor	22,23
	Glucocorticoid resistance associated with mutant glucocorticoid receptor	24
	LQTI (Romano-Ward) syndrome associated with mutant potassium channel alleles mapped to chromosome 11	25
	APL[a] associated with ATRA-responsive chimeric PML-RARα gene[a]	26 and 27
1992	Vasopressin resistance due to mutant AVPR2 receptor on X chromosome	28
	FcεRI receptor (β-chain) mapped to chromosome 11q 13	29
	Angiotensin II type 1 receptor cloned, sequenced, and mapped to chromosome 3q21–25	30
1993	β_2 adrenoceptor mutations first identified	31
1994	Human AH receptor localized to chromosome 7p21–P15	32
	Paradoxical response to antiandrogens tied to CAG repeats in androgen receptor	33

(*Continued*)

Table 3 A Chronology of Human Receptor Pharmacogenetics (*Continued*)

Year	Event	References
	Estrogen resistance identified with mutant estrogen receptor	34
	Vitamin D receptor gene localized to chromosome 12q12-q22 (see 1999)	35
	LQT3 gene mapped to 3p21–24	36
1995	Nocturnal asthmatic phenotype associated with β_2AR polymorphism	37
	Sulfonylurea receptor cloned, associated with hyperinsulin secretion maps to chromosome 11p15.1	38
1996	Asthmatic phenotype associated with FcεRIβ subunit mutation	39
	Resistance to HIV-induced infection associated with CCR5Δ32 allele in adults	40–42
1997	Vitamin D receptor gene genomic structure determined	43
	High CYP1A1 activity associated with AH receptor polymorphism	44
	Mineralocorticoid resistance associated with mutant mineralocorticoid receptor	45
	Gram-negative (LPS) shock associated with mutant *Tlr4* receptor	46
	ESR2 cloned and mapped to chromosome 14q22–24	47
1999	Vitamin D receptor gene localized on chromosome 12cenq12	48
2000	Resistance to HIV-induced infection associated with CCR2-64I allele in children, and not to the CCR5Δ32 allele	49

[a]Acute promyelocytic leukemia.

accelerated the identification of receptors and receptor subtypes and has greatly refined their structural and functional analysis.

Many altered drug responses are associated with aberrant forms of receptors. Functional modifications in the abundance, affinity, or stability of receptors identified as pharmacological mechanisms that cause such variability are summarized in Tables 4 and 5.

1.3. Classification of Receptors

Receptors are traditionally subdivided into nuclear receptors and cell surface receptors. The two groups of receptors are considered separately in the following discussions.

2. NUCLEAR RECEPTORS

The superfamily of regulatory proteins that interacts with hormones such as cortisol, the sex steroids estradiol and testosterone, thyroxine and with

Table 4 Nuclear Receptors of Pharmacogenetic Interest

Receptor (Chromosome)	Ligand(s)	Response	Molecular basis
Glucocorticoid (5q-q32) (54)	Dexamethasone	Glucocorticoid resistance	Arg641Val missense mutation creates receptor with decreased affinity for dexamethasone (24)
	Generalized compensated glucocorticoid resistance	Increased cortisol secretion	Splice site deletion creates reduced receptor abundance (55)
	Dexamethasone	Decreased dexamethasone potency	Ile729Val missense mutation creates receptor with decreased affinity for dexamethasone (56)
	Endogenous cortisol; dexamethasone	Cushing's syndrome	Ile559Asp missense mutation creates an aberrant receptor with reduced abundance, an inability to bind dexamethasone, and a dominant negative effect on gene transcription (57)
	Dexamethasone	Increased cortisol suppression and insulin sensitivity	Asp363Ser missense mutation creates higher sensitivity to dexamethasone (58)
Mineralocorticoid (4q31.2) (50)	Aldosterone	Mineralocorticoid resistance	Inactivating mutations (ΔG 1226, ΔT 1597, C/T 1831 stop, Δ A intron splice site) create receptors that lead to dominant forms of pseudohypoaldosteronism (PHA1) (45)
	Aldosterone	Early onset hypertension accelerated by pregnancy	Ser810Lys missense mutation creates receptors with constitutive activity in absence of steroid, but normal activation by aldosterone (53,45)

Androgen (Xq11-q13) (17)	Endogenous testosterone and dihydrotestosterone	Androgen insensitivity syndromes. Impaired response to exogenous testosterone	Missense mutations, large deletions, and frameshift mutations create receptors with absent or reduced hormone binding, increased lability, and failure of up-regulation (59,60)
	Hormonal therapy	Hormone refractory prostate tumors	Somatic missense mutations (e.g. Thr877Ala in exon 1) create tumors refractory to hormonal therapy (60–63)
	Antiandrogens: cyproterone acetate, andronon, hydroxyflutamide	Paradoxical response to antiandrogens	Thr868Ala missense mutation in the ligand binding domain creates a receptor that responds to low doses of estrogens and progestens, and has an increased affinity for antiandrogenic drugs (64,65)
	Flutamide	Paradoxical response to flutamide	Hypothesize that polymorphism in glutamine (CAG) repeat length creates a receptor that affects prostate tumor response to anticancer agents (33,66)
Estrogen (6q25.1) (9)	Transdermal ethinyl estradiol	Estrogen insensitivity. Impaired response to exogenous estrogen	Replacement of Arg 157 (CGA) codon with a premature stop codon (TGA) creates receptor lacking both the DNA binding and hormone binding domains (34,67)
	4-hydroxytamoxifen (4OHT); ICI 164,384	Reduced estrogen-dependent transcriptional activation	Substitution of L543A/L544A or M547A/L548A creates a receptor less responsive to estradiol but strongly stimulated by 4OHT and ICI 164,384 (68)
	None	Occurs in cell line (MCF-7.2A) that developed estrogen independent growth	Creates cell with 4–5 copies of the wild-type receptor; acts as dominant negative receptor (69)

(*Continued*)

Table 4 Nuclear Receptors of Pharmacogenetic Interest (*Continued*)

Receptor (Chromosome)	Ligand(s)	Response	Molecular basis
	Estradiol; ICI 164,384; RU 54,876; *trans*-hydroxytamoxifen	Inverted ligand activity	Leu540Gln mutant creates a receptor that interprets antiestrogens as estrogens and vice versa (70)
	Estradiol; tamoxifen, ICI 164,384	Constitutive transactivation	Tyr537Asn creates a receptor that displays constitutive transactivation (71)
	Estradiol; raloxifene; keoxifene; ICI 182,780	Antiestrogen keoxifene increases estrogenic activity but antiestrogen 182,780 maintains antagonistic activity	Asp351Tyr mutation occurs in tamoxifen-stimulated breast tumor; changes pharmacology of non-steroidal antiestrogen (keoxifene) (72,73)
Retinoic acid (t 15;17) (4)	All *trans*-retinoic acid (ATRA)	ATRA induces remission of newly diagnosed acute promyelocytic leukemia	Chimeric receptor (PML-RARα) blocks differentiation of APL cells by complexing with corepressors. ATRA promotes complex dissociation of this complex to relieve the block (74)
Vitamin D (nuclear) (12cen-q12) (48)	1,25 Dihydroxyvitamin D_3 (calcitriol)	Vitamin D-resistant rickets in children. Osteoporosis in adults	A variety of mutations, mainly in the DNA-binding domain and to a lesser extent in the ligand-binding domain, create a receptor that is non-functional or hyporesponsive to the vitamin D_3 (75)
Arylhydrocarbon (AH) [(7p16) (32,76)]	Halogenated and polyaromatic hydrocarbons	Activation of chemicals to genotoxins, teratogens and carcinogens, and detoxification of toxic/carcinogenic xenobiotics	Ligand activates a battery of genes coding for enzymes involved in drug and toxicant metabolism. Arg554Lys polymorphism creates a receptor associated with high inducible CYP1A1 (44,76,77)

Table 5 Cell Surface Receptors of Pharmacogenetic Interest

Receptor (Chromosome)	Response	Associated with	Molecular basis
Potassium channel *KVLQT1* α-subunit (1p155) (114)	LQT1 syndrome	Spontaneous and drug-induced ventricular arrhythmias. Induced by drugs including antiarrhythmics (quinidine), antipsychotics (chlorpromazine), psychotropics (tricylic antidepressants), macrolide antibiotics (clathromycin), H1 antihistamines (terfenadine, astemizole) (113)	At least eleven different heterozygous (Romano-Ward) and homozygous (JL-N) missense and deletions mutations create defective ion channel receptors (115)
Potassium channel *HERG* α-subunit (7q35–36) (114)	LQT2 syndrome		At least six missense, intragenic and splicing mutations create defective ion channel receptors (116)
Sodium channel *SCN5A* (3p21–24) (114)	LQT3 syndrome; Brugada variant of LQT3 syndrome		At least three missense and intragenic activating mutations create defective ion channel receptors (117)
Ankyrin–B (4q25–27) (118)	LQT4 syndrome		Loss of function E1425G missense mutation creates a receptor that leads to altered Ca2+ signaling and provides a rationale for arrhythmias (118)
Potassium channel *minK KVLQT1* β-subunit (21q22.1) (119)	LQT5 syndrome		At least two missense mutations create defective ion channel receptors (119)

(*Continued*)

Table 5 Cell Surface Receptors of Pharmacogenetic Interest (*Continued*)

Receptor (Chromosome)	Response	Associated with	Molecular basis
Potassium channel *MiRP1* (KCNE2) β-subunit (21q22.1) (120)	LQT6 syndrome		At least three missense mutations create defective ion channel receptors (120)
Sulfonylurea (11p15.1) (38,121)	Decreased secretion of insulin and tolbutamide-stimulated insulin	Familial hyperinsulinism and hypoglycemia of infancy, and NIDDM	Splice site and missense mutations in non-coding and coding regions create defective receptors (121,122)
β_2-Adrenergic (5q31–32) (10)	Bronchial hyperresponsiveness. Patients with the Gly16 and Arg16 polymorphism exhibit different responses to albuterol	Asthma	Two common polymorphisms (Arg16Gly; Gln27Glu); and one rare polymorphism (Thr164Ile) in the coding region create receptors whose abundance and function are altered in persons homozygous for the polymorphism (123,124)
FcεRI α (1q23) (21)	FcεRI receptor is located on mast cells and other effector cells. Exposure to allergen causes formation of IgE-receptor complex which degranulates mast cells and initiates inflammation	Atopic individuals that produce excess amounts of IgE are predisposed to asthma, hay fever and eczema	Three SNPs identified in the coding region, but their function is unknown (125)

FcεRI beta (11q13) (126)			A missense mutation (Glu237Gly) of the β-subunit of the receptor creates a receptor associated with bronchial hyper-responsiveness. The mechanism of this response is not clear (39,127,128)
CCR5Δ32 (3p21) (40–42)	HIV-induced infection	Resistance to HIV-induced infection	32 basepair deletion of the coding region creates a severely truncated, non-functional receptor incapable of binding HIV-suppressive β-chemokines (40–42)
Toll-like (9q32–33) (129)	Gram-negative septic shock	Susceptibility to endotoxic (LPS) shock from Gram-negative bacteria	Missense and null variants known in C3H/HeJ and C57BL/10ScCr mice (46,130). In humans, Asp299Gly and Thr399Ile, are common mutations associated with blunted responses to inhaled LPS (131)
Angiotensin II type 1 (3q21–25) (30)	Vasoconstrictive and salt-conserving actions of the renin angiotensin system	Essential hypertension and other cardiovascular disorders	A common genetic variation (A1166C) in the 3′ untranslated tract at the AT1 locus is associated with a small effect on blood pressure that may predispose individuals to essential hypertension (132,133). No consistent association of genetic variation in AT1 has been observed to predispose individuals to myocardial infarction, ventricular hypertrophy, and microvascular complications (134)

drugs such as dexamethasone, the steroidal contraceptives, certain vitamins, and triiodothyronine is prototypical of nuclear receptors. Glucocorticoid, mineralocorticoid, progesterone, androgen, and estrogen receptors (ER) form the human steroid receptor subfamily. The Ah receptor, a receptor that interacts with exogenous planar aromatic toxins and carcinogens such as dioxin (2,3,7,8-tetrachlorobenzo-*p*-dioxin, or TCDD) and benzo[a]pyrene, though distinct from the steroid receptor, is another member of the nuclear receptor superfamily.

Nuclear hormone receptor genes comprise six (A–F) distinct domains: an N-terminal domain (A/B) that codes for activating transcription; a central, DNA-binding (DBD, C) domain containing two zinc fingers required for insertion into the hormone response elements in the promoters of hormone responsive genes; a hinge (D) domain; a ligand-binding (LBD, E) domain and the binding site for heat shock proteins; and, in the case of estrogen receptors, a C-terminal (F) domain (Fig. 1). Of these the DBD and LBD are evolutionarily conserved and are encoded by exons 2–3 and 4–8, respectively. Functionally, nuclear receptors are direct signal transduction systems that in their ligand-bound form bind to hormone response and enhancer elements of the genome to alter transcription rates of quiescent genes. In their unliganded state, these receptor proteins are transcriptionally inactive and are distributed in the cytoplasm associated with several chaperone proteins (hsp90 and hsp70) and several other receptor-related proteins. After hormone binding, the receptor undergoes conformational changes, dissociates from the heat shock proteins, homodimerizes to assume transactivating function, and translocates to the nucleus where it binds to hormone responsive elements in the promoter regions of target genes to enhance or repress transcription of the linked gene. Despite a great deal of elegant

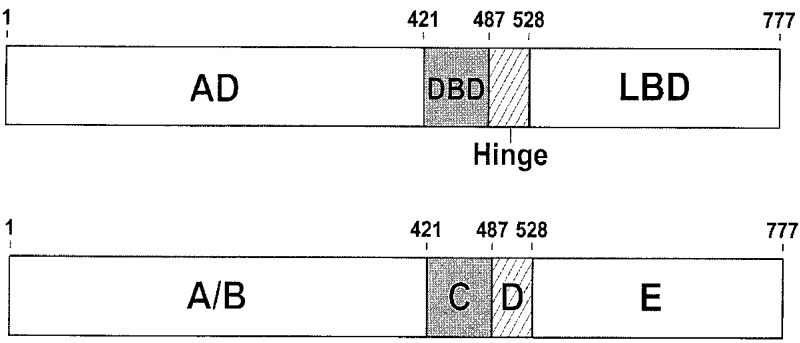


Figure 1 Domain map of the steroid receptor. The numbers correspond to residues of the human glucocorticoid α receptor at the boundaries of the regions identified in the schematic diagram. *Source*: PubMed Primary accession number P04150.

research, however, the exact sequence of events prior to the entry of the receptor into the nucleus is not yet fully understood.

Mutations in each of the three domains mentioned above can potentially disrupt gene transcription and lead to hormone resistance or other hormonally related disorders (Table 4). Information about glucocorticoid, mineralocorticoids, androgens, estrogens, retinoic acid, vitamin D, and the AH receptor is summarized in Table 4. Mineralocorticoid, estrogen, and AH receptors are discussed in greater detail below.

2.1. Mineralocorticoid Receptors and Early Onset Hypertension

The mineralocorticoid receptor (MR) mediates the sodium-retaining effects of aldosterone in the kidney, salivary glands, sweat glands, and colon. The human MR gene was cloned in 1987 and bears structural and functional kinships to the glucocorticoid receptor (11). It consists of 984 amino acids, spans 60–90 kb on chromosome 4q31.2 (50,51) and contains 10 exons including two exons (1α and 1β) that encode different 5′ untranslated sequences whose expression is controlled by two different promoters (52).

A missense mutation (MRS810L) in the ligand binding domain of the MR has been shown recently to cause an autosomal dominant form of early onset hypertension (53,45). This mutation was discovered in a 15-year-old boy with severe hypertension and tests of his family revealed 11 of 23 relatives had the mutation, all of whom had been diagnosed with severe hypertension before they were 20 years old. The mutation results in constitutive MR receptor activity in the absence of added steroid, but normal activation by aldosterone. However, progesterone and other clinically used corticoid antagonists such as spironolactone lacking 21-hydroxyl groups become agonists.

Since progesterone levels usually increase 100-fold during pregnancy, reaching concentrations of 500 nM, the investigators thought it likely that pregnant females with the S810L isoform of the MR receptor might develop severe hypertension. Two MR carriers of the family mentioned above had undergone five pregnancies, all of which were complicated by exacerbation of hypertension. The absence of proteinuria, edema, or neurological changes excluded preeclampsia.

Molecular modeling and site-directed mutagenesis demonstrated that the L810 residue in helix 5 of the ligand binding MR domain makes a new van der Waals interaction with A773 in helix 3 that substitutes for the interaction of the 21-hydroxy group with helix 3. This interaction eliminates the requirement for the 21-OH group of aldosterone to interact with N770 in helix 3. Effective therapy to correct this disorder has not been developed, but since every other steroid hormone receptor as well as several other nuclear hormone receptors have leucine–alalnine or methionine–glycine pairs at homologous positions in helixes 5 and 3, the modeling suggests a general approach to creation of steroid receptor antagonists (78).

2.2. Estrogen Receptors and Their Role in Estrogen Resistance and Pharmacotherapy

Estrogens are the key modulators of cell growth and differentiation, particularly, those associated with reproduction and mammary gland development. In the late 1980s, the ER was cloned (7), localized to chromosome 6q25.1 (8), and its genomic organization was determined (15). Early studies indicated the ER was the principal pathway by which responses to estrogens and antiestrogens are mediated in normal and cancerous target cells and was associated with progression of cancer of the breast, endometrium, and prostate. We have learned more recently that estrogen signaling is mediated through pathways controlled by two receptors, ERα and ERβ (also named ESR1 and ESR2), and that aberrant forms of ERα and ERβ could be responsible for disturbances of these pathways.

Current knowledge of naturally occurring variant and mutant ERs rests on a large and rapidly growing body of molecular evidence (79,80). In contrast to the highly mutable androgen receptor for which several hundred cases of partial or complete androgen sensitivity have been reported, complete absence of the ER has been reported only once (34), suggesting that mutations of this receptor severely compromise fetal survival. The reported case (Tables 3 and 4) is explained by a point mutation that results in a severely truncated, functionally inert receptor lacking both the DNA- and hormone-binding domains and complete insensitivity to continued estrogen therapy.

A number of earlier studies drew attention to unusual responses to estrogens and antiestrogens as described in Table 4 (68–73). As many of these investigations preceded the discovery of the second ER (ERβ), their conclusions may now need to be reassessed.

In 1995 ERβ was cloned from rat prostate, and homologs for humans (81) and mice (82) were cloned soon thereafter. This appears to be the first instance of a steroid hormone receptor existing in two isoforms. The ERα contains 595 amino acids while the predicted size of ERβ protein is somewhat smaller at 530 residues (83), but their structures are more complex than originally supposed. In addition to exons 1–8, the genes for ERα and ERβ contain additional exons, including several untranslated exons upstream of exon, and various ''intronic'' exons. There are also eight untranslated 5′-exons upstream of exon 1, and an ''intronic'' exon in ERα, and the presence of multiple coding exon 8s in ERβ (84).

The ERα and ERβ are genetically distinct. Both are differentially expressed in many tissues prompting the suggestion that they regulate different sets of genes. The ERα occurs in uterus, prostate stroma, ovarian thecal cells, and bone, while ERβ is found in breast stroma, sympathetic ganglia, colon, prostate epithelium, ovarian granulosa cells, dorsal raphe, and bone marrow, whereas both isoforms are present in breast epithelium

and brain (85). The fact that uterine tissue was used for most early cloning efforts probably explains why ERα was cloned first and why it occupies a dominant position in the literature. The ERα does dominate in a few tissues, but ERβ is more generally expressed and may play an important physiological role in many other tissues (86).

The occurrence of ER receptors in two forms (87) may also explain the broad spectrum of estrogen pharmacology and some of the paradoxical effects of estrogen-like drugs. The two forms are not only genetically distinct, but are located on different chromosomes: ERα on chromosome 6q25.1 and ERβ on chromosome 14q22–24 (47) (Tables 3 and 4), and differ in their distribution in normal tissue and cancerous tissue (88). Functionally, their transcriptional activity may oppose each other (e.g., ERα activates while ERβ inhibits AP-1 activity), and their interactions with antiestrogens differ (e.g., the selective estrogen modulator (SERM), tamoxifen, has greater antagonist activity through ERβ than ERα, but this relationship is reversed for another SERM, raloxifene). Most variant ERα RNAs that occur in primary breast cancer are alternative spliced forms, and a similar situation exists for ERβ, but the clinical significance of these observations is obscure (89). In contrast, point mutations of ERα are relatively sparse in primary disease, but do occur more frequently in metastatic disease. One study found a common somatic point mutation (A908G) of the ERα gene in 20 (34%) of 59 early premalignant breast lesions (90). The mutation substitutes an Arg for a Lys at residue 303 within exon 4 at the border between the hinge and LBD regions of the receptor. As this gain-of-function mutation confers maximal cell growth-stimulatory response and enhanced binding of the TIF-2 coactivator on ERα at extremely low, subphysiological levels of estrogen, it provides an explanation for the ER's increased estrogen responsiveness. The altered receptor could significantly accelerate tumor progression in early breast disease, but a large clinical study will be necessary to assess whether genetic analysis for the A908G ERα mutation in premalignant lesions would benefit preventive measures.

The ER agonists have become standard treatment for hormone replacement therapy of post-menopausal women while ER antagonists are widely employed as therapeutic agents in breast cancer treatment, but their usage appears to be associated with serious adverse effects, many of which are ER-related. The most serious side effects of ER agonists are enhanced risks of venous thrombosis, and cancer of the breast and endometrium (91). Therapy with ER antagonists is associated with endometrial cancer, cardiovascular symptoms (hot flushes and profuse sweating), CNS disturbances (insomnia and anxiety), sexual dysfunction (vaginal dryness, dyspareunia, and psychosexual symptoms) and the development of estrogen-dependent tumors (15,92). Despite their common use in medical therapy, further studies are needed to clarify the effect of genetic variation on response to these drugs.

The effects of ERα on cardiovascular disease are complex and contradictory (91), but the recent prospective long term (27-year follow-up) case-control study of 1739 subjects study of Shearman et al. (93) and Hopkins and Brinton (94) revealed the risk of cardiovascular disease associated with a common ERα receptor genotype, the *ESR1 c.454–397T>C* allelic variant is substantially increased. The CC genotype of *ESR1 c.454–397T>C* conferred a 2.0 greater risk of major atherosclerotic disease, and a 3.0-fold greater risk of myocardial infarction, compared to individuals of the CT or TT genotypes. These results are restricted to men because too few women had events to study them separately. The frequency of the CC genotype was 20% in the entire sample, but was 24% in persons with the broadest definition of atherosclerosis, 31% in those with major atherosclerotic cardiovascular disease, and 37% in those with myocardial infarction. The association was statistically significant for the latter two endpoints. Shearman's study supports and extends the Helsinki Sudden Death Study, which found a 10.6-fold greater risk of coronary artery disease among 300 white Finnish male autopsy cases (95).

The discovery of ERβ has raised more questions about estrogen regulation than there are answers (96). What we know regarding ERβ expression comes mainly from a limited number of RNA samples so the ultimate contribution of ERβ protein to the outcome of breast cancer is as yet unclear, and its application to predict treatment response must await examination in clinical trials. However, considerable progress is evident in developing ligands with ERα-selective agonist activity, with ERα-selective antagonist activity, with ERβ-selective agonist activity, and ERβ-selective antagonist/ERα-selective agonist activity (97). These ligands retain their ER-subtype-selectivity and their agonistic or antagonistic character at numerous diverse gene sites. They should be useful in elucidating the biological functions of ERα and ERβ, and contribute to drug discovery.

2.3. Arylhydrocarbon Hydroxylase Receptors and Halogenated and Polycyclic Hydrocarbon Toxicity

The Ah receptor (AhR) system regulates arylhydrocarbon hydroxylase (AH) activity, now renamed CYP1A1 activity (Table 1; Ref. 98), which is responsible for the metabolism of numerous exogenous chemicals that are high on the list of injurious environmental pollutants. The activation of this system is of pharmacogenetic interest because halogenated hydrocarbons such as TCDD (2,3,7,8-tetrachlorodibenzo-*p*-dioxin, dioxin), a by product of the synthesis of agent orange, and benzo[a]pyrene, a polycyclic hydrocarbon of cigarette smoke, cause a variety of toxic effects in animals and potentially in humans exposed to them.

In 1973, Kellermann et al. first drew attention to the genetic variability of AH activity in reporting a trimodal distribution of activity of lymphocytes

obtained form healthy volunteers and treated in culture with the polyaromatic hydrocarbon, 3-methylcholanthrene by Kellermann et al. (99). Other investigators have not been able to reproduce this finding (100), but they have demonstrated in twin studies that heritability of AHH inducibility in twin studies was probably controlled by a single gene or a few genes (101,102), and that a positive correlation exists between inducibility of AH activity (of lymphocytes) and primary lung cancer (103).

The AhR was discovered in mice and much of the initial research to define its biochemical and molecular mechanisms of action was performed in this animal model (104); more recently other models have been used (105). The studies in mice indicate that the AhR gene is located on mouse chromosome 12, is constitutively expressed in many cell types and regulated in a tissue-cell, developmental- and chemical-specific manner (77). Over the years, studies to define the biological effects of 3-methylcholanthrene, benzpyrene, TCDD, and related toxicants have focused on the induction of CYP1A1 activity.

Exposure to a ligand (such as TCDD) enhances (i.e., induces) expression of the AhR gene and activates a diverse battery of other genes including CYP1A1. Variability in Ah activity has been studied extensively in animal models, and it is clear that AhR induction by arylhydrocarbons is central to understanding the pharmacogenetics of nuclear receptors (106–108). This topic is more appropriately dealt with in detail elsewhere (see Chapter 2d, Variability of induction processes, A. Okey).

Polymorphism in the human AH receptor gene (Arg554Lys) was first detected in Japanese subjects, but this polymorphism was not associated with a change in AH activity, or with susceptibility to lung cancer (109). This mutation imparts a high transactivating activity in the mouse (110). Daly et al. (44) have confirmed the presence of the Arg554Lys polymorphism in Caucasians, and shown it to be associated with highly inducible CYP1A1 activity. The Arg-coded and Lys-coded alleles were both observed relatively frequently in Japanese, 0.57 and 0.43, respectively, but were quite different in Caucasians, 0.12 and 0.88, respectively. The Arg554Lys mutation plus two additional mutations (Val570Ile and Pro571Ser) have been reported in exon 10 of the coding region of the human AHR (106). None of these three SNPs seem to be related to any human phenotype of TCDD toxicity, but Wong et al. (111) note that a combination of Arg554Lys and Val570Ile expressed together in Hepa-1 Group B mutant cells failed to support loss of CYP1A1 induction despite the ability of each variant to bind TCDD and to promote binding of the ligands Ahr–Arnt complex to dioxin responsive elements. However, this combination of mutations is rare in humans and appears to be confined to persons of African descent. Another study found another variant (Met786Val) in exon 10 in the course of screening a small sample of French lung cancer patients. This variant coexisted with high CYP1A1 in these subjects but since its frequency was so low

(0.005), a larger study is necessary to confirm its phenotypic effect (112). Cauchi et al. (112) also reported an allelic variant in the promoter at position 157; it has a frequency of 0.25 but it does not seem to exert a phenotypic effect on CYP1A1 inducibility.

Allele frequencies appear to vary widely among diverse ethnic populations though only small numbers of subjects in a few populations have been reported (106). For example, allele frequencies for the Arg554Lys polymorphism range from 0.09 to 0.14 in Native American, French Canadian and Canadian Inuit subjects to 0.32 in subjects of Chinese descent to 0.40–0.57 in subjects of African descent. Limited information suggests that polymorphisms at codons 517 and 570 may be confined to subjects of African descent, but studies of larger populations are necessary to verify this.

So far, comparatively few genetic variations have been identified in the human AHR, and even fewer have been unequivocally established as having a functional effect on AHH activity. Those that have been detected reside primarily in exon 10, which encompasses the main portion of the transactivation domain of the receptor. Among these, the Arg554Lys polymorphism is the most widely studied, but the effects of this polymorphism on phenotype are complicated and confusing. Much of the difficulty is believed to stem from imperfections in various phenotyping assays (106) or from gender differences (112). The AHR system is believed to be a key player in the elimination of polycyclic hydrocarbons in the human environment, but whether it may also be implicated, as a regulatory factor in homeostasis and development is unclear. Such a role may imply the presence of an intrinsic ligand for AHR but identification of such a ligand would be difficult until an intrinsic role for this receptor system other than the regulation for xenobiotic metabolic enzymes has been clarified (107). The fact that so few functional mutations have been detected in the human AH receptor suggests that lack of a functional receptor may severely compromise fetal survival as proposed for the ER.

3. CELL SURFACE RECEPTORS

Cell surface receptors, the second major class of receptor proteins, account for the actions of the great majority of biological responses to exogenous chemicals because of their capacity to bind biogenic amines, protein and polypeptide hormones, autocoids, neurotransmitters, and environmental chemicals. The receptors for these substances are divided into three subgroups: (1) the ion channels and ion transporters; (2) those that act by an enzymatic cascade involving a second messenger such as the adenylyl cyclase cascade or the phosphoinositol cascade; (3) those whose action depends on an integral enzyme activity.

The ion channel/transporter group is represented by several receptors associated with the long QT syndrome, and the sulfonylurea receptor that is

associated with infantile hyperinsulinism and non-insulin dependent diabetes mellitus (NIDDM). The second messenger group is represented by the β_2-adrenergic and the CCR5 receptors. The insulin receptor is the prototypical member of the third group of receptors; but this receptor is not discussed because the pharmacogenetics of this receptor has been considered at length elsewhere (113). The toll (*Tlr4*) receptor is discussed, but this receptor family is not assigned to any of the three groups of cell surface receptors because the mechanism of transmembrane signaling has not been reported. Information regarding variant forms of these receptors is summarized in Table 5. The β_2-adrenergic, CCR5 and toll (*Tlr4*) are discussed in greater detail.

3.1. β_2 Adrenergic Receptor Genes and Asthma

Genetic analysis in families suggests that asthma is influenced by only a few genes each with a moderately strong effect, rather than many genes of small effect (135). Two receptor genes, the β-adrenergic receptor and the high affinity receptor for IgE (FcεRI-β), have been studied extensively, and the β_2 adrenergic receptors (β_2ARs) will be considered here. An effect of the Il-4 receptor (on chromosome 16) on atopy and serum IgE levels is also known. A gain-of-function mutation has also been identified in the α subunit of the Il-4 receptor, but its frequency and significance to the predisposition of human subjects to asthma is not yet established (136).

The β_2ARs have been the target of investigation since the β-adrenergic receptor theory of atopy was set forth some 30 years ago. The β_2 adrenergic receptors have been extensively studied and are prototypical of the G-protein-coupled receptor. The β_2ARs mediate the actions of catecholamines in various tissues. The gene is intronless and is located on chromosome 5q31–32. Like other members of the GPCR superfamily, they possess an extracellular amino terminus, seven transmembrane-spanning domains, three extracellular and three intracellular loops, and an intracellular carboxy terminus. Three of four non-synonymous mutations in the coding region of the receptor (Arg16Gly, Gln27Glu, and Thr164Ile) (123) alter the abundance and function of the receptor (137). Ethnic differences in allelic frequencies are significant. The allelic frequencies in Caucasians, Blacks, and Asians of Gly16 are 0.61, 0.50, and 0.40 and for Gln27 are 0.57, 0.73, and 0.80, respectively. The Ile164 polymorphism is uncommon, occurring in the heterozygous state in ≈6% of the population; persons homozygous for this polymorphism have not been identified.

The frequencies of each polymorphism are the same for normal adults and asthmatics, so that these primary variations in the structure of the receptor are not considered a major cause of asthma (138). However, the Gly16 form of the receptor undergoes agonist-promoted downregulation while Glu27 displays resistance to such downregulation. Additionally,

a clustering of the Gly16 polymorphism was noted in more severe asthmatics (123), and this polymorphism is associated with nocturnal asthma (37). On the other hand, the Glu27 form of the receptor protects against bronchial hyperreactivity (139). The Giy16Arg and Gln27Glu polymorphisms may thus act as modifiers of the asthmatic phenotype, but as they are linked, larger population studies will be necessary to determine the relative importance of each one to asthma (137,140).

The effect of polymorphic variation on the clinical response of asthmatic patients to β-adrenergic agonists and corticosteroids has been extensively studied (141–145), but these studies are rather inconsistent and not easily compared because different drugs were tested under different circumstances, and the measures of therapeutic effect achieved varied from one study to another. Several studies find that asthmatic patients who are homozygous for the Arg16 form of the receptor are more likely to respond favorably to albuterol (salbutamol) than patients who are homozygous for the Gly16 form (141,142,144). For instance, Martinez et al. (142) found that homozygotes for Arg16 were 5.3 times and heterozygotes for Arg16 were 2.3 times more likely to respond to albuterol than homozygotes for Gly16, whereas Lipworth et al. (145) found that neither of these polymorphisms influenced the response to isoproterenol or to formerterol protection against methacholine. Additionally, two studies on the response of asthmatic patients to inhaled corticosteroids (used on demand) showed that this mode of treatment afforded equal protection of patients regardless of the β_2-receptor polymorphism (143,145).

Receptors with Ile at position 164 exhibit sustained, markedly dysfunctional properties with altered high affinity binding and decreased coupling to the stimulatory G protein in transfected cells (31,137), but the rarity of this polymorphism has so far precluded its study in the homozygous state or in situations of clinical interest.

Drysdale's recent study, the first of its kind, assessed the effect of combinations of SNPs (haplotypes) on response to the drug, albuterol (146). Variation in 13 SNPs could be organized into 12 haplotypes, and investigation of these haplotypes in several discrete populations (Caucasian, African-American, Asian, and Hispanic-Latino) revealed >20-fold differences among the frequencies of the four major haplotypes. Mean responses to bronchodilator therapy with albuterol of individuals with the five most common β_2AR haplotypes varied by greater than twofold, but were unrelated to individual SNPs.

3.2. AIDS Susceptibility and the CCR5 Coreceptor

Initial characterization of HIV-1 revealed the presence of the β-chemokine receptor, CCR5 (also designated CC-CKR-5 and CKR5) was required on the surface of mature human helper T lymphocytes for attachment and

infectivity of R5 viruses such as HIV-1 (40–42). Meanwhile, the search for other viral entry factors disclosed another chemokine CXC receptor (CXCR4) that was crucial for cellular entry of HIV-1 (147), and that CXCR4 and CCR5 were coreceptors for this process. Additional chemokine receptors can facilitate the entry of HIV-1 variants in vitro, but since all clinical isolates to date use CCR5, CXCR4, or both for entry (148), CCR5 (for R5 viruses) and CXCR4 (for X4 viruses) are regarded as the primordial binding sites for HIV-1 (and many other viruses) (149).

The CCR5 gene encodes a seven-transmembrane, G-protein coupled, cell surface receptor that binds several HIV-suppressive β-chemokines (150) and naturally occurring mutations that alter expression or function of this gene are clinically significant. The 32-bp deletion mutant of CCR5 (CCR5Δ32 allele) that occurs disproportionately among persons who are frequently exposed to HIV-1 (40–42,151) explains in part the genetic basis for resistance to HIV-induced disease. The defective gene is severely truncated and not expressed at the cell surface. Persons homozygous for the CCR5Δ32 allele are highly protected from contracting HIV infection, and progression of AIDS is slower in individuals who are heterozygous for the deletion or carry a partial deletion of gene (42). Limited data suggest that rare cases of HIV-1 infection in persons homozygous for the CCR5Δ32 allele may be caused by X4 viruses (152).

The CCR5Δ32 allele occurs with a frequency of 0.10 in Caucasian and Western European populations, reaching approximately 0.20 in some populations. The CCR5Δ32 allele is present throughout Europe, the Middle East and the Indian subcontinent (153) and it dominates all other sequence variations observed in Caucasian Americans (154). This allelic variant is absent among native African, American Indian, and East Indian ethnic groups while allelic frequencies vary from 0% to 14% occurs across Eurasia. The occurrence of this geographic cline and its recent emergence (estimated at 700–2000 years ago) has prompted speculation that a strong selective event such as an epidemic of a pathogen that, like HIV-1, uses CCR5, has driven its frequency upward in ancestral Caucasian populations (155). A recent study in mice that carry a homozygous deletion of CCR5 opposes this hypothesis (156) and an alternative hypothesis proposes epidemics of smallpox may provide a better explanation for this cline (157).

The common CCR5 promoter polymorphism at position –2459 (A/G) has also been associated with a slower rate of AIDS progression by Salkowitz and colleagues (158). The mechanism by which this polymorphism alters the course of HIV-1 infection is not clear, but since the G/G, A/G, and A/A promoter genotypes were associated with low, medium, and high viral propagation, and have a similar relationship with CCR5 densities, they concluded that the CCR5-2459 promoter polymorphism determines CCR5 expression and predicts the magnitude of HIV-1 propagation in vitro. They

believe that this polymorphism may be exploited to reduce CCR5 expression that will suppress and/or prevent HIV-1 infection (158).

Currently, the potential of single-chain antibody fragments for intracellular applications, termed "intrabodies," is being explored as another alternative to alter CCR5 expression (159) similar in principle to methods of gene inactivation that target DNA or mRNA such as antisense, zinc finger proteins, targeted gene disruption, and RNA interference. Steinberger et al. (160) have proposed the introduction of a specific "intrabody" (intracellular antibodies) into hematopoietic stem cells to produce a pool of cells that is refractory to CCR5-dependent HIV-1 infection by functional deletion of the CCR5 receptor. Their strategy is founded on the observation that an antibody, ST6, recognizes a unique sequence in the first extracellular domain of CCR5 that is conserved in human and non-human primates. Intracellular expression of the ST6 antibody with an endoplasmic reticulum (ER)-retention signal is superior to other chemokines (such as RANTES) in blocking CCR5 surface expression and in preventing cell–cell fusion events. Aided by improved technologies for manipulating antibody genes, and gene delivery that stably transduces human hematopoietic stem cells, they hope to exploit this system as a tool for drug discovery.

The protective effects of CCR5Δ32, one of the few highly consistent, reproducible associations between polymorphisms and disease (Table 3 in Ref. 161), has been demonstrated only in adults and is less well defined in children. Among children, carriers of a CCR2-64I allele ($n = 93$) progressed significantly more slowly to AIDS than wild-type CCR2 homozygotes ($n = 289$; $p = 0.007$) prolonging the median AIDS-free survival time from 38 months in CCR2 wild type homozygotes to 87 months in CCR2-64I carriers. Among 27 patients who died during the study, only 1 carried a CCR2-64I allele. In contrast, the CCR5Δ32 did not modify the rate of HIV-1 transmission or disease progression in perinatally infected children (49). Further studies are needed to elucidate the reason for the beneficial effect of the genetic variation in CCR2.

Current treatment regimens for HIV-1 infection usually include two nucleoside analogues and either a protease inhibitor or a non-nucleoside reverse-transcriptase inhibitor in an attempt to avert the development of resistance (162). Some investigators believe that drugs that inhibit entry of HIV virions into target cells may have advantages over drugs that inhibit steps in the viral life cycle after infection has occurred (149). A number of cell entry inhibitors are currently being evaluated, most of which are at various early stages of development (149). Among these is a humanized monoclonal anti-CD4 antibody called TNX-355, and another agent, enfuvirtide. Despite numerous failures, the prospects for development of new inhibitors of viral cell entry continue to be pursued.

3.3. The *Tlr4* Receptor and Endotoxic Shock

Endotoxin is an abundant component of gram-negative bacteria that evokes fever, shock and other disturbances in mammals (163). The endotoxic principle is a lipopolysaccharide (LPS) comprised of a toxic lipid moiety (lipid A) and several other non-toxic moieties of highly variable structure. The LPS is of pharmacogenetic interest because gram-negative infection annually claims tens of thousands of lives in the United States alone, and genetic studies suggest that variation in the recognition of this toxin may be important for containment and eradication of a gram-negative infection.

The *Tlr4* represents an ancient self-defense immune response, and it appears to be the only gene in the LPS critical region. Mice with *Tlr4* mutations, fruit flies with Toll defects, and perhaps humans with *Tlr*-related problems are all susceptible to gram-negative infections. Mammals respond by developing gram-negative shock. Some 40 years ago, identification of LPS-resistant mice in the inbred strain A mouse (164) provided the first clue to the genetic basis of this trait. Because mutation of a single-gene entirely ablated LPS responses, the existence of a single pathway for responses to LPS was suggested. Cloning of the responsible gene from mice that exhibited defective LPS signaling (130), and further evidence in endotoxin-resistant mice indicated that LPS and Tlr4 were the same gene (46,165). Taken together, these results provide convincing evidence that Tlr4 is the gene product of the LPS locus on mouse chromosome 4, and when defective, this locus results in hyporesponsiveness to LPS in mice.

While the mouse studies were in progress, five human toll-like receptors TLR1-5 were cloned (129) including the human homologue of the mouse *Tlr4* gene (166). In mammals, at least 10 members of the Toll superfamily of receptors are known (167). The TLR genes were found to be located on human chromosomes 4 (TLRs 1–3), 9 (TLR 4), and 1 (TLR 5), and further refinement showed TLR4 mapped to human chromosome 9q32–33.

Except for a single anecdotal case (168), no human models of endotoxin hyporesponsiveness have been described. In addition to septic shock, however, exposure to endotoxin in the human environment is associated with the development and progression of various airway disorders including asthma (169,170). Arbour et al. (131) showed that variability in TLR4 may underlie the variability in airway responsiveness and found two common co-segregating missense mutations (Asp299Gly and Thr399Ile) in exon 4 that affected the extracellular domain of the receptor, and were associated with a blunted response to inhaled LPS in three populations. The allelic frequency of the Asp299Gly substitution was 6.6% in the Iowan study population, 7.9% in a control Iowan population, and 3.3% in the parental chromosomes of the CEPH population. The Asp299Gly mutation interrupted the TLR4-mediated LPS hyporesponsive phenotype but the Thr399Ile did not. This study provided the first direct evidence that a sequence polymorphism in TLR4 is

associated with an endotoxin hyporesponsive phenotype in humans. But because not all subjects who were hyporesponsive to LPS had mutations in TLR4, these mutations and conversely, and vice versa, these mutations are thought to act in concert with other genetic changes or acquired factors to influence the response to LPS.

The *Tlr4* has emerged as a specific conduit for the LPS response, and many relatively uncommon variants have been observed, but a single *Tlr4* allelic receptor variant predominates. Such receptors may also play a role in airway disease in addition to their role in Gram-negative endotoxic shock.

4. SUMMARY

Naturally occurring genetically determined changes that occur in nuclear and cell surface receptors can often explain the ineffectiveness of therapeutic agents, adverse or paradoxical drug responses, sensitivity to various foodstuffs, and susceptibility (or resistance) to infectious disease, and suggest new strategies of therapy and prevention. The first phase of the Human Genome Project has defined the consensus sequence for the human genome (171,172). The second phase of the project is moving steadily toward discovery of common patterns of DNA sequence variation in the major populations of the world (173), and recently, the international HapMap consortium has initiated the arduous task of defining the associations between these patterns and human disease (174).

Exhaustive coverage of the pharmacogenetics of receptors cannot be accomplished in such an abbreviated survey. With this limitation in mind, about a dozen receptors that illustrate the depth and scope of genetic variation inherent in human drug response were selected that cover the full range of structural and functional types of nuclear and cell surface receptors encountered in the biology of human drug response. Discussion emphasizes the molecular genetic basis and pharmacogenetic significance of important traits associated with mineralocorticoid, estrogen, and arylhydroxylase nuclear receptors, and the β2 adrenergic, CCR5, and Tlr4 cell surface receptors.

ABBREVIATIONS

AIDS	Autoimmune deficiency disease
AHR, AhR or Ahr	Arylhydrocarbon hydroxylase receptor in humans and mice, respectively
APL	Acute promyelocytic leukemia
ATRA	All *trans* retinoic acid
CAR	A member of the nuclear receptor family
C3H/HeJ	An inbred mouse strain
C57BL/	An inbred mouse strain
CCR2	HIV receptor

CCR5, CC-CKR-5, CKR5	Alternative names for HIV receptor
CCR5Δ32	Deletion variant of CCR5 HIV receptor
CYP1A1	A P450 enzyme
D_3	Active form of vitamin D
DBA	An inbred mouse strain
ERα and ERβ	Different isoforms of the estrogen receptor
ESR1 and ESR2	Alternative names for ERα and ERβ, respectively
FcεRI-β	High affinity receptor for IgE
GR	Glucocorticoid receptor
HERG	A channel locus associated with the Long QT syndrome
HIV-1	Human immunodeficiency virus type 1
HPA	Hypothalamic–pituitary–adrenal axis
ICI 164,384	An antiestrogen
IL-1	Interleukin 1
KVQT1	A channel locus associated with the Long QT syndrome
LPS	liposaccharide, the endotoxic principle
LQT 1,2,3,4 & 6	Gene loci associated with the Long QT syndrome
LQTS	Long QT syndrome
MR	Mineralocorticoid receptor
Mink	A channel locus associated with the Long QT syndrome
MiRP1	A channel locus associated with the Long QT syndrome
NIDDM	Non-insulin-dependent diabetes mellitus
PHHI	Persistent hyperinsulinemic hypoglycemia of infancy
PLZF	A transcription factor implicated in acute promyelocytic leukemia
PML	A form of the retinoic acid receptor
PPAR	A member of the nuclear receptor family
PXR	A member of the nuclear receptor family
RXR	A form of the retinoic acid receptor
SERMs	Selective estrogen receptor modulators
SCN5A	A channel locus associated with the Long QT syndrome
TCDD	Dioxin, 2,3,7,8-tetrachlorodibenzo-dioxin
Tlr4 (mouse) or TLR4 (human)	Receptor implicated in endotoxic shock

ACKNOWLEDGMENTS

My thanks to James Rae, Patrick Murphy, and Adam Kuszak for their criticisms of portions of the manuscript concerning estrogen receptors, the nuclear receptors, and the CCR5 receptors and AIDS susceptibility.

REFERENCES

1. Ahlquist RP. A study of adrenotropic receptors. Am J Physiol 1948; 153: 586–600.
2. Denborough MA, Forster JF, Lovell RR, Maplestone PA, Villiers JD. Anaesthetic deaths in a family. Br J Anaesth 1962; 34:395–396.
3. O'Reilly RA, Aggeler PM, Hoag MS, Leong LS, Kropatkin ML. Hereditary transmission of exceptional resistance to coumarin anticoagulant drugs. The first reported kindred. N Engl J Med 1964; 271:809–815.
4. Rowley JD, Golomb HM, Dougherty C. 15/17 Translocation, a consistent chromosomal change in acute promyelocytic leukaemia. Lancet 1977; 1(8010): 549–550.
5. Yang-Feng TL, Francke U, Ullrich A. Gene for human insulin receptor: localization to site on chromosome 19 involved in pre-B-cell leukemia. Science 1985; 228(4700):728–731.
6. Hollenberg SM, Weinberger C, Ong ES, Cerelli G, Oro A, Lebo R, Thompson EB, Rosenfeld MG, Evans RM. Primary structure and expression of a functional human glucocorticoid receptor CDNA. Nature 1985; 318(6047): 635–641.
7. Green S, Walter P, Kumar V, Krust A, Bornert JM, Argos P, Chambon P. Human oestrogen receptor CDNA: sequence, expression and homology to V-Erb-A. Nature 1986; 320(6058):134–139.
8. Gosden JR, MIddleton PG. Rout D Localization of the human oestrogen receptor gene to chromosome 6q24-Q27 by in situ hybridization. Cytogenet Cell Genet 1986; 43(3–4):218–220.
9. Menasce LP, White GR, Harrison CJ, Boyle JM. Localization of the estrogen receptor locus (ESR) to chromosome 6q25.1 by FISH and a simple post-FISH banding technique. Genomics 1993; 17(1):263–265.
10. Kobilka BK, Dixon RA, Frielle T, Dohlman HG, Bolanowski MA, Sigal IS, Yang-Feng TL, Francke U, Caron MG, Lefkowitz RJ. CDNA for the human beta 2-adrenergic receptor: a protein with multiple membrane-spanning domains and encoded by a gene whose chromosomal location is shared with that of the receptor for platelet-derived growth factor. Proc Natl Acad Sci USA 1987; 84(1):46–50.
11. Arriza JL, Weinberger C, Cerelli G, Glaser TM, Handelin BL, Housman DE, Evans RM. Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. Science 1987; 237:268–275.
12. Petkovich M, Brand NJ, Krust A, Chambon P. A human retinoic acid receptor which belongs to the family of nuclear receptors. Nature 1987; 330(6147): 444–450.
13. Baker AR, McDonnell DP, Hughes M, Crisp TM, Mangelsdorf DJ, Haussler MR, Pike JW, Shine J, O'Malley BW. Cloning and expression of full-length CDNA encoding human vitamin D receptor. Proc Natl Acad Sci USA 1988; 85(10):3294–3298.
14. Hughes MR, Malloy PJ, Kieback DG, Kesterson RA, Pike JW, Feldman D, O'Malley BW. Point mutations in the human vitamin D receptor gene associated with hypocalcemic rickets. Science 1988; 242(4886):1702–1705.

15. Ponglikitmongkol M, Green S, Chambon P. Genomic organization of the human oestrogen receptor gene. EMBO J 1988; 7(11):3385–3388.
16. Chang CS, Kokontis J, Liao ST. Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors. Proc Natl Acad Sci USA 1988; 85(19):7211–7215.
17. Lubahn DB, Joseph DR, Sar M, Tan J, Higgs HN, Larson RE, French FS, Wilson EM. The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. Mol Endocrinol 1988; 2(12):1265–1275.
18. Kadowaki T, Bevins CL, Cama A, Ojamaa K, Marcus-Samuels B, Kadowaki H, Beitz L, McKeon C, Taylor SI. Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance. Science 1988; 240:787–790.
19. Yoshimasa Y, Seino S, Whittaker J, Kakehi T, Kosaki A, Kuzuya H, Imura H, Bell GI, Steiner DF. Insulin-resistant diabetes due to a point mutation that prevents insulin proreceptor processing. Science 1988; 240(4853):784–787.
20. Mattei MG, Petkovich M, Mattei JF, Brand N, Chambon P. Mapping of the human retinoic acid receptor to the Q21 band of chromosome 17. Hum Genet 1988; 80(2):186–188.
21. Le Coniat M, Kinet JP, Berger R. The Human genes for the alpha and gamma subunits of the mast cell receptor for immunoglobulin E are located on human chromosome band 1q23. Immunogenetics 1990; 32(3):183–186.
22. Gillard EF, Otsu K, Fujii J, Khanna VK, De Leon S, Derdemezi J, Britt BA, Duff CL, Worton RG, MacLennan DH. A substitution of cysteine for arginine 614 in the ryanodine receptor is potentially causative of human malignant hyperthermia. Genomics 1991; 11(3):751–755.
23. Otsu K, Khanna VK, Archibald AL, MacLennan DH. Cosegregation of porcine malignant hyperthermia and a probable causal mutation in the skeletal muscle ryanodine receptor gene in backcross families. Genomics 1991; 11(3): 744–750.
24. Hurley DM, Accili D, Stratakis CA, Karl M, Vamvakopoulos N, Rorer E, Constantine K, Taylor SI, Chrousos GP. Point mutation causing a single amino acid substitution in the hormone binding domain of the glucocorticoid receptor in familial glucocorticoid resistance. J Clin Invest 1991; 87(2):680–686.
25. Keating M, Dunn C, Atkinson D, Timothy K, Vincent GM, Leppert M. Consistent linkage of the long-QT syndrome to the harvey Ras-1 locus on chromosome 11. Am J Hum Genet 1991; 49(6):1335–1339.
26. de The H, Lavau C, Marchio A, Chomienne C, Degos L, Dejean A. The PML-RAR alpha fusion MRNA generated by the T(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. Cell 1991; 66(4): 675–684.
27. Kakizuka A, Miller WH Jr, Umesono K, Warrell RP Jr, Frankel SR, Murty VV, Dmitrovsky E, Evans RM. Chromosomal translocation T(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML. Cell 1991; 66(4):663–674.
28. Birnbaumer M, Seibold A, Gilbert S, Ishido M, Barberis C, Antaramian A, Brabet P, Rosenthal W. Molecular cloning of the receptor for human antidiuretic hormone. Nature 1992; 357(6376):333–335.

29. Cookson WO, Young RP, Sandford AJ, Moffatt MF, Shirakawa T, Sharp PA, Faux JA, Julier C, Nakumuura Y, Nakumura Y. Maternal inheritance of atopic IgE responsiveness on chromosome 11q. Lancet 1992; 340(8816):381–384.
30. Curnow KM, Pascoe L, White PC. Genetic analysis of the human type-1 angiotensin II receptor. Mol Endocrinol 1992; 6(7):1113–1118.
31. Green SA, Cole G, Jacinto M, Innis M, Liggett SB. A polymorphism of the human beta 2-adrenergic receptor within the fourth transmembrane domain alters ligand binding and functional properties of the receptor. J Biol Chem 1993; 268(31):23116–23121.
32. Le Beau MM, Carver LA, Espinosa R III, Schmidt JV, Bradfield CA. Chromosomal localization of the human AHR locus encoding the structural gene for the Ah receptor to 7p21- >P15. Cytogenet Cell Genet 1994; 66(3):172–176.
33. Schoenberg MP, Hakimi JM, Wang S, Bova GS, Epstein JI, Fischbeck KH, Isaacs WB, Walsh PC, Barrack ER. Microsatellite mutation (CAG24→18) in the androgen receptor gene in human prostate cancer. Biochem Biophys Res Commun 1994; 198(1):74–80.
34. Smith EP, Boyd J, Frank GR, Takahashi H, Cohen RM, Specker B, Williams TC, Lubahn DB, Korach KS. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. N Engl J Med 1994; 331(16):1056–1061.
35. Pedeutour F, Merscher S, Durieux E, Montgomery K, Krauter K, Clevy JP, Barcelo G, Kucherlapati R, Gaudray P, Turc-Carel C. Mapping of the 12q12-Q22 region with respect to tumor translocation breakpoints. Genomics 1994; 22(3):512–518.
36. Jiang C, Atkinson D, Towbin JA, Splawski I, Lehmann MH, Li H, Timothy K, Taggart RT, Schwartz PJ, Vincent GM. Two long QT syndrome loci map to chromosomes 3 and 7 with evidence for further heterogeneity. Nat Genet 1994; 8(2):141–147.
37. Turki J, Pak J, Green SA, Martin RJ, Liggett SB. Genetic polymorphisms of the beta 2-adrenergic receptor in nocturnal and nonnocturnal asthma. Evidence that Gly16 correlates with the nocturnal phenotype. J Clin Invest 1995; 95(4): 1635–1641.
38. Glaser B, Chiu KC, Liu L, Anker R, Nestorowicz A, Cox NJ, Landau H, Kaiser N, Thornton PS, Stanley CA. Recombinant mapping of the familial hyperinsulinism gene to an 0.8 CM region on chromosome 11p15.1 and demonstration of a founder effect in Ashkenazi Jews. Hum Mol Genet 1995; 4(5):879–886.
39. Hill MR, Cookson WO. A new variant of the beta subunit of the high-affinity receptor for immunoglobulin E (Fc epsilon RI-beta E237G): associations with measures of atopy and bronchial hyper-responsiveness. Hum Mol Genet 1996; 5(7):959–962.
40. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 1996; 86(3):367–377.
41. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG,

Doms RW, Vassart G, Parmentier M. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 1996; 382(6593):722–725.
42. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, Goedert JJ, Buchbinder SP, Vittinghoff E, Gomperts E, Donfield S, Vlahov D, Kaslow R, Saah A, Rinaldo C, Detels R, O'Brien SJ. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia growth and development study, multicenter AIDS cohort study, multicenter hemophilia cohort study, San Francisco City Cohort, ALIVE study. Science 1996; 273(5283):1856–1862.
43. Miyamoto K, Kesterson RA, Yamamoto H, Taketani Y, Nishiwaki E, Tatsumi S, Inoue Y, Morita K, Takeda E, Pike JW. Structural organization of the human vitamin D receptor chromosomal gene and its promoter. Mol Endocrinol 1997; 11(8):1165–1179.
44. Daly AK, Fairbrother KS, Smart J. Recent advances in understanding the molecular basis of polymorphisms in genes encoding cytochrome P450 enzymes. Toxicol Lett 1998; 102–103:143–147.
45. Geller DS, Rodriguez-Soriano J, Vallo Boado A, Schifter S, Bayer M, Chang SS, Lifton RP. Mutations in the mineralocorticoid receptor gene cause autosomal dominant pseudohypoaldosteronism type I. Nat Genet 1998; 19(3):279–281.
46. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, Birdwell D, Alejos E, Silva M, Galanos C, Freudenberg M, Ricciardi-Castagnoli P, Layton B, Beutler B. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science 1998; 282(5396):2085–2088.
47. Enmark E, Pelto-Huikko M, Grandien K, Lagercrantz S, Lagercrantz J, Fried G, Nordenskjold M, Gustafsson JA. Human estrogen receptor beta-gene structure, chromosomal localization, and expression pattern. J Clin Endocrinol Metab 1997; 82(12):4258–4265.
48. Taymans SE, Pack S, Pak E, Orban Z, Barsony J, Zhuang Z, Stratakis CA. The human vitamin D receptor gene (VDR) is localized to region 12cen-Q12 by fluorescent in situ hybridization and radiation hybrid mapping: genetic and physical VDR map. J Bone Miner Res 1999; 14(7):1163–1166.
49. Mangano A, Kopka J, Batalla M, Bologna R, Sen L. Protective effect of CCR2-64I and not of CCR5-delta32 and SDF1-3′A in pediatric HIV-1 infection. J Acquir Immune Defic Syndr 2000; 23(1):52–57.
50. Morrison N, Harrap SB, Arriza JL, Boyd E, Connor JM. Regional chromosomal assignment of the human mineralocorticoid receptor gene to 4q31.1. Hum Genet 1990; 85(1):130–132.
51. Fan YS, Eddy RL, Byers MG, Haley LL, Henry WM, Nowak NJ, Shows TB. The human mineralocorticoid receptor gene (MLR) is located on chromosome 4 at Q31.2. Cytogenet Cell Genet 1989; 52(1–2):83–84.
52. Zennaro MC, Keightley MC, Kotelevtsev Y, Conway GS, Soubrier F, Fuller PJ. Human mineralocorticoid receptor genomic structure and identification of expressed isoforms. J Biol Chem 1995; 270(36):21016–21020.
53. Geller DS, Farhi A, Pinkerton N, Fradley M, Moritz M, Spitzer A, Meinke G, Tsai FT, Sigler PB, Lifton RP. Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. Science 2000; 289(5476):119–123.

54. Francke U, Foellmer BE. The glucocorticoid receptor gene is in 5q31-Q32 [corrected]. Genomics 1989; 4(4):610–612.
55. Karl M, Lamberts SW, Detera-Wadleigh SD, Encio IJ, Stratakis CA, Hurley DM, Accili D, Chrousos GP. Familial glucocorticoid resistance caused by a splice site deletion in the human glucocorticoid receptor gene. J Clin Endocrinol Metab 1993; 76(3):683–689.
56. Malchoff DM, Brufsky A, Reardon G, McDermott P, Javier EC, Bergh CH, Rowe D, Malchoff CD. A mutation of the glucocorticoid receptor in primary cortisol resistance. J Clin Invest 1993; 91(5):1918–1925.
57. Karl M, Lamberts SW, Koper JW, Katz DA, Huizenga NE, Kino T, Haddad BR, Hughes MR, Chrousos GP. Cushing's disease preceded by generalized glucocorticoid resistance: clinical consequences of a novel, dominant-negative glucocorticoid receptor mutation. Proc Assoc Am Physicians 1996; 108(4):296–307.
58. Huizenga NA, Koper JW, de Lange P, Pols HA, Stolk RP, Burger H, Grobbee DE, Brinkmann AO, de Jong FH, Lamberts SW. A polymorphism in the glucocorticoid receptor gene may be associated with and increased sensitivity to glucocorticoids in vivo. J Clin Endocrinol Metab 1998; 83(1): 144–151.
59. Tincello DG, Saunders PT, Hodgins MB, Simpson NB, Edwards CR, Hargreaves TB, Wu FC. Correlation of clinical, endocrine and molecular abnormalities with in vivo responses to high-dose testosterone in patients with partial androgen insensitivity syndrome. Clin Endocrinol (Oxford) 1997; 46(4): 497–506.
60. Gottlieb B, Trifiro M, Lumbroso R, Pinsky L. The androgen receptor gene mutations database. Nucleic Acids Res 1997; 25(1):158–162.
61. Taplin ME, Bubley GJ, Shuster TD, Frantz ME, Spooner AE, Ogata GK, Keer HN, Balk SP. Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. N Engl J Med 1995; 332(21):1393–1398.
62. Tilley WD, Buchanan G, Hickey TE, Bentel JM. Mutations in the androgen receptor gene are associated with progression of human prostate cancer to androgen independence. Clin Cancer Res 1996; 2(2):277–285.
63. Chen CP, Chen SR, Wang TY, Wang W, Hwu YM. A frame shift mutation in the DNA-binding domain of the androgen receptor gene associated with complete androgen insensitivity, persistent mullerian structures, and germ cell tumors in dysgenetic gonads. Fertil Steril 1999; 72(1):170–173.
64. Veldscholte J, Ris-Stalpers C, Kuiper GG, Jenster G, Berrevoets C, Claassen E, van Rooij HC, Trapman J, Brinkmann AO, Mulder E. A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens. Biochem Biophys Res Commun 1990; 173(2):534–540.
65. Veldscholte J, Berrevoets CA, Brinkmann AO, Grootegoed JA, Mulder E. Anti-androgens and the mutated androgen receptor of LNCaP cells: differential effects on binding affinity, heat-shock protein interaction, and transcription activation. Biochemistry 1992; 31(8):2393–2399.
66. Barrack ER. Androgen receptor mutations in prostate cancer. Mt Sinai J Med 1996; 63(5–6):403–412.

67. Sudhir K, Chou TM, Chatterjee K, Smith EP, Williams TC, Kane JP, Malloy MJ, Korach KS, Rubanyi GM. Premature coronary artery disease associated with a disruptive mutation in the estrogen receptor gene in a man. Circulation 1997; 96(10):3774–3777.
68. Mahfoudi A, Roulet E, Dauvois S, Parker MG, Wahli W. Specific mutations in the estrogen receptor change the properties of antiestrogens to full agonists. Proc Natl Acad Sci USA 1995; 92(10):4206–4210.
69. Pink JJ, Wu SQ, Wolf DM, Bilimoria MM, Jordan VC. A novel 80 kDa human estrogen receptor containing a duplication of exons 6 and 7. Nucleic Acids Res 1996; 24(5):962–969.
70. Montano MM, Ekena K, Krueger KD, Keller AL, Katzenellenbogen BS. Human estrogen receptor ligand activity inversion mutants: receptors that interpret antiestrogens as estrogens and estrogens as antiestrogens and discriminate among different antiestrogens. Mol Endocrinol 1996; 10(3): 230–242.
71. Zhang QX, Borg A, Wolf DM, Oesterreich S, Fuqua SA. An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. Cancer Res 1997; 57(7):1244–1249.
72. Levenson AS, Catherino WH, Jordan VC. Estrogenic activity is increased for an antiestrogen by a natural mutation of the estrogen receptor. J Steroid Biochem Mol Biol 1997; 60(5–6):261–268.
73. Levenson AS, Jordan VC. The key to the antiestrogenic mechanism of raloxifene is amino acid 351 (Aspartate) in the estrogen receptor. Cancer Res 1998; 58(9):1872–1875.
74. Lin RJ, Egan DA, Evans RM. Molecular genetics of acute promyelocytic leukemia. Trends Genet 1999; 15(5):179–184.
75. Malloy PJ, Pike JW, Feldman D. The vitamin D receptor and the syndrome of hereditary 1,25-dihydroxyvitamin D-resistant rickets. Endocr Rev 1999; 20(2): 156–188.
76. Micka J, Milatovich A, Menon A, Grabowski GA, Puga A, Nebert DW. Human Ah receptor (AHR) gene: localization to 7p15 and suggestive correlation of polymorphism with CYP1A1 Inducibility. Pharmacogenetics 1997; 7(2): 95–101.
77. Garrison PM, Denison MS. Analysis of the murine AhR gene promoter. J Biochem Mol Toxicol 2000; 14(1):1–10.
78. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. Cell 2001; 104(4):545–556.
79. Korach KS, Couse JF, Curtis SW, Washburn TF, Lindzey J, Kimbro KS, Eddy EM, Migliaccio S, Snedeker SM, Lubahn DB, Schomberg DW, Smith EP. Estrogen receptor gene disruption: molecular characterization and experimental and clinical phenotypes. Recent Prog Horm Res 1996; 51:159–186.
80. Murphy LC, Leygue E, Dotzlaw H, Douglas D, Coutts A, Watson PH. Oestrogen receptor variants and mutations in human breast cancer. Ann Med 1997; 29(3):221–234.
81. Mosselman S, Polman J, Dijkema R. ER Beta: identification and characterization of a novel human estrogen receptor. FEBS Lett 1996; 392(1):49–53.

82. Tremblay GB, Tremblay A, Copeland NG, Gilbert DJ, Jenkins NA, Labrie F, Giguere V. Cloning, chromosomal localization, and functional analysis of the murine estrogen receptor beta. Mol Endocrinol 1997; 11(3):353–365.
83. Fuqua SA, Schiff R, Parra I, Moore JT, Mohsin SK, Osborne CK, Clark GM, Allred DC. Estrogen receptor beta protein in human breast cancer: correlation with clinical tumor parameters. Cancer Res 2003; 63(10):2434–2439.
84. Hirata S, Shoda T, Kato J, Hoshi K. Isoform/Variant MRNAs for sex steroid hormone receptors in humans. Trends Endocrinol Metab 2003; 14(3):124–129.
85. Gustafsson JA. What pharmacologists can learn from recent advances in estrogen signalling. Trends Pharmacol Sci 2003; 24(9):479–485.
86. Gustafsson JA. Estrogen receptor beta—a new dimension in estrogen mechanism of action. J Endocrinol 1999; 163(3):379–383.
87. Paech K, Webb P, Kuiper GG, Nilsson S, Gustafsson J, Kushner PJ, Scanlan TS. Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. Science 1997; 277(5331):1508–1510.
88. Lippman SM, Brown PH. Tamoxifen prevention of breast cancer: an instance of the fingerpost. J Natl Cancer Inst 1999; 91(21):1809–1819.
89. Fuqua SA. The role of estrogen receptors in breast cancer metastasis. J Mammary Gland Biol Neoplasia 2001; 6(4):407–417.
90. Fuqua SA, Wiltschke C, Zhang QX, Borg A, Castles CG, Friedrichs WE, Hopp T, Hilsenbeck S, Mohsin S, O'Connell P, Allred DC. A hypersensitive estrogen receptor-alpha mutation in premalignant breast lesions. Cancer Res 2000; 60(15):4026–4029.
91. Herrington DM. Role of estrogen receptor-alpha in pharmacogenetics of estrogen action. Curr Opin Lipidol 2003; 14(2):145–150.
92. Splawski I, Shen J, Timothy KW, Lehmann MH, Priori S, Robinson JL, Moss AJ, Schwartz PJ, Towbin JA, Vincent GM, Keating MT. Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. Circulation 2000; 102(10):1178–1185.
93. Shearman AM, Cupples LA, Demissie S, Peter I, Schmid CH, Karas RH, Mendelsohn ME, Housman DE, Levy D. Association between estrogen receptor alpha gene variation and cardiovascular disease. J Am Med Assoc 2003; 290(17):2263–2270.
94. Hopkins PN, Brinton EA. Estrogen receptor 1 variants and coronary artery disease: shedding light into a murky pool. J Am Med Assoc 2003; 290(17): 2317–2319.
95. Lehtimaki T, Kunnas TA, Mattila KM, Perola M, Penttila A, Koivula T, Karhunen PJ. Coronary artery wall atherosclerosis in relation to the estrogen receptor 1 gene polymorphism: an autopsy study. J Mol Med 2002; 80(3): 176–180.
96. Katzenellenbogen BS, Korach KS. A new actor in the estrogen receptor drama—enter ER-beta. Endocrinology 1997; 138(3):861–862.
97. Harrington WR, Sheng S, Barnett DH, Petz LN, Katzenellenbogen JA, Katzenellenbogen BS. Activities of estrogen receptor alpha- and beta-selective ligands at diverse estrogen responsive gene sites mediating transactivation or transrepression. Mol Cell Endocrinol 2003; 206(1–2):13–22.

98. Nebert DW, Petersen DD, Puga A. Human AH locus polymorphism and cancer: inducibility of CYP1A1 and other genes by combustion products and dioxin. Pharmacogenetics 1991; 1(2):68–78.
99. Kellermann G, Luyten-Kellermann M, Shaw CR. Genetic variation of aryl hydrocarbon hydroxylase in human lymphocytes. Am J Hum Genet 1973; 25(3):327–331.
100. Paigen B, Minowada J, Gurtoo HL, Paigen K, Parker NB, WardE, Hayner NT, Bross ID, Bock F, Vincent R. Distribution of aryl hydrocarbon hydroxylase inducibility in cultured human lymphocytes. Cancer Res 1977; 37(6):1829–1837.
101. Atlas SA, Vesell ES, Nebert DW. Genetic control of interindividual variations in the inducibility of aryl hydrocarbon hydroxylase in cultured human lymphocytes. Cancer Res 1976; 36(12):4619–4630.
102. Paigen B, Ward E, Steenland K, Houten L, Gurtoo HL, Minowada J. Aryl hydrocarbon hydroxylase in cultured lymphocytes of twins. Am J Hum Genet 1978; 30(5):561–571.
103. Kouri RE, McKinney CE, Slomiany DJ, Snodgrass DR, Wray NP, McLemore TL. Positive correlation between high aryl hydrocarbon hydroxylase activity and primary lung cancer as analyzed in cryopreserved lymphocytes. Cancer Res 1982; 42(12):5030–5037.
104. Nebert DW, Robinson JR, Niwa A, Kumaki K, Poland AP. Genetic expression of aryl hydrocarbon hydroxylase activity in the mouse. J Cell Physiol 1975; 85(2 Pt 2 suppl1):393–414.
105. Pohjanvirta R, Wong JM, Li W, Harper PA, Tuomisto J, Okey AB. Point mutation in intron sequence causes altered carboxyl-terminal structure in the aryl hydrocarbon receptor of the most 2,3,7,8-tetrachlorodibenzo-p-dioxin-resistant rat strain. Mol Pharmacol 1998; 54(1):86–93.
106. Harper PA, Wong JY, Lam MS, Okey AB. Polymorphisms in the human AH receptor. Chem Biol Interact 2002; 141(1–2):161–187.
107. Mimura J, Fujii-Kuriyama Y. Functional role of AhR in the expression of toxic effects by TCDD. Biochim Biophys Acta 2003; 1619(3):263–268.
108. Ma Q, Lu AY. Origins of individual variability in P4501A induction. Chem Res Toxicol 2003; 16(3):249–260.
109. Kawajiri K, Watanabe J, Eguchi H, Nakachi K, Kiyohara C, Hayashi S. Polymorphisms of human Ah receptor gene are not involved in lung cancer. Pharmacogenetics 1995; 5(3):151–158.
110. Ko HP, Okino ST, Ma Q, Whitlock JP Jr. Transactivation domains facilitate promoter occupancy for the dioxin-inducible CYP1A1 gene in vivo. Mol Cell Biol 1997; 17(7):3497–3507.
111. Wong JM, Okey AB, Harper PA. Human aryl hydrocarbon receptor polymorphisms that result in loss of CYP1A1 induction. Biochem Biophys Res Commun 2001; 288(4):990–996.
112. Cauchi S, Stucker I, Solas C, Laurent-Puig P, Cenee S, Hemon D, Jacquet M, Kremers P, Beaune P, Massad-Massade L. Polymorphisms of human aryl hydrocarbon receptor (AhR) gene in a french population: relationship with CYP1A1 inducibility and lung cancer. Carcinogenesis 2001; 22(11): 1819–1824.

113. Weber WW. Pharmacogenetics. New York: Oxford University Press, 1997:344.
114. Splawski I, Timothy KW, Vincent GM, Atkinson DL, Keating MT. Molecular basis of the long-QT syndrome associated with deafness. N Engl J Med 1997; 336(22):1562–1567.
115. Priori SG, Barhanin J, Hauer RN, Haverkamp W, Jongsma HJ, Kleber AG, McKenna WJ, Roden DM, Rudy Y, Schwartz K, Schwartz PJ, Towbin JA, Wilde A. Genetic and molecular basis of cardiac arrhythmias; impact on clinical management. Study group on molecular basis of arrhythmias of the working group on arrhythmias of the european society of cardiology. Eur Heart J 1999; 20(3):174–195.
116. Curran ME, Splawski I, Timothy KW, Vincent GM, Green ED, Keating MT. A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. Cell 1995; 80(5):795–803.
117. Chen Q, Kirsch GE, Zhang D, Brugada R, Brugada J, Brugada P, Potenza D, Moya A, Borggrefe M, Breithardt G, Ortiz-Lopez R, Wang Z, Antzelevitch C, O'Brien RE, Schulze-Bahr E, Keating MT, Towbin JA, Wang Q. Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. Nature 1998; 392(6673):293–296.
118. Mohler PJ, Schott JJ, Gramolini AO, Dilly KW, Guatimosim S, duBell WH, Song LS, Haurogne K, Kyndt F, Ali ME, Rogers TB, Lederer WJ, Escande D, Le Marec H, Bennett V. Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. Nature 2003; 421(6923):634–639.
119. Bianchi L, Shen Z, Dennis AT, Priori SG, Napolitano C, Ronchetti E, Bryskin R, Schwartz PJ, Brown AM. Cellular dysfunction of LQT5-MinK mutants: abnormalities of IKs, IKr and trafficking in long QT syndrome. Hum Mol Genet 1999; 8(8):1499–1507.
120. Abbott GW, Sesti F, Splawski I, Buck ME, Lehmann MH, Timothy KW, Keating MT, Goldstein SA. MiRP1 forms IKr potassium channels with HERG and is associated with cardiac arrhythmia. Cell 1999; 97(2):175–187.
121. Thomas PM, Cote GJ, Wohllk N, Haddad B, Mathew PM, Rabl W, Aguilar-Bryan L, Gagel RF, Bryan J. Mutations in the sulfonylurea receptor gene in familial persistent hyperinsulinemic hypoglycemia of infancy. Science 1995; 268(5209):426–429.
122. Thomas PM, Wohllk N, Huang E, Kuhnle U, Rabl W, Gagel RF, Cote GJ. Inactivation of the first nucleotide-binding fold of the sulfonylurea receptor, and familial persistent hyperinsulinemic hypoglycemia of infancy. Am J Hum Genet 1996; 59(3):510–518.
123. Reihsaus E, Innis M, MacIntyre N, Liggett SB. Mutations in the gene encoding for the beta 2-adrenergic receptor in normal and asthmatic subjects. Am J Respir Cell Mol Biol 1993; 8(3):334–339.
124. Green SA, Turki J, Innis M, Liggett SB. Amino-terminal polymorphisms of the human beta 2-adrenergic receptor impart distinct agonist-promoted regulatory properties. Biochemistry 1994; 33(32):9414–9419.
125. Shikanai T, Silverman ES, Morse BW, Lilly CM, Inoue H, Drazen JM. Sequence variants in the FcepsilonRI alpha chain gene. J Appl Physiol 2002; 93(1):37–41.

126. Sandford AJ, Shirakawa T, Moffatt MF, Daniels SE, Ra C, Faux JA, Young RP, Nakamura Y, Lathrop GM, Cookson WO. Localisation of atopy and beta subunit of high-affinity IgE receptor (Fc Epsilon RI) on chromosome 11q. Lancet 1993; 341(8841):332–334.
127. Shirakawa T, Li A, Dubowitz M, Dekker JW, Shaw AE, Faux JA, Ra C, Cookson WO, Hopkin JM. Association between atopy and variants of the beta subunit of the high-affinity immunoglobulin E receptor. Nat Genet 1994; 7(2):125–129.
128. van Herwerden L, Harrap SB, Wong ZY, Abramson MJ, Kutin JJ, Forbes AB, Raven J, Lanigan A, Walters EH. Linkage of high-affinity IgE receptor gene with bronchial hyperreactivity, even in absence of atopy. Lancet 1995; 346(8985):1262–1265.
129. Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF. A family of human receptors structurally related to drosophila toll. Proc Natl Acad Sci USA 1998; 95(2):588–593.
130. Poltorak A, Smirnova I, He X, Liu MY, Van Huffel C, McNally O, Birdwell D, Alejos E, Silva M, Du X, Thompson P, Chan EK, Ledesma J, Roe B, Clifton S, Vogel SN, Beutler B. Genetic and physical mapping of the Lps locus: identification of the toll-4 receptor as a candidate gene in the critical region. Blood Cells Mol Dis 1998; 24(3):340–355.
131. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat Genet 2001; 25(June):187–191.
132. Bonnardeaux A, Davies E, Jeunemaitre X, Fery I, Charru A, Clauser E, Tiret L, Cambien F, Corvol P, Soubrier F. Angiotensin II type 1 receptor gene polymorphisms in human essential hypertension. Hypertension 1994; 24(1): 63–69.
133. Kainulainen K, Perola M, Terwilliger J, Kaprio J, Koskenvuo M, Syvanen AC, Vartiainen E, Peltonen L, Kontula K. Evidence for involvement of the type 1 angiotensin ii receptor locus in essential hypertension. Hypertension 1999; 33(3):844–849.
134. Wang JG, Staessen JA. Genetic polymorphisms in the renin-angiotensin system: relevance for susceptibility to cardiovascular disease. Eur J Pharmacol 2000; 410(2–3):289–302.
135. Cookson W. The alliance of genes and environment in asthma and allergy. Nature 1999; 402(6760 suppl):B5–B11.
136. Hershey GK, Friedrich MF, Esswein LA, Thomas ML, Chatila TA. The association of atopy with a gain-of-function mutation in the alpha subunit of the interleukin-4 receptor. N Engl J Med 1997; 337(24):1720–1725.
137. Weir TD, Mallek N, Sandford AJ, Bai TR, Awadh N, Fitzgerald JM, Cockcroft D, James A, Liggett SB, Pare PD. Beta2-adrenergic receptor haplotypes in mild, moderate and fatal/near fatal asthma. Am J Respir Crit Care Med 1998; 158(3):787–791.
138. Liggett SB. The pharmacogenetics of beta2-adrenergic receptors: relevance to asthma. J Allergy Clin Immunol 2000; 105(2 Pt 2):S487–S492.

139. Hall IP, Wheatley A, Wilding P, Liggett SB. Association of Glu 27 beta 2-adrenoceptor polymorphism with lower airway reactivity in asthmatic subjects. Lancet 1995; 345(8959):1213–1214.
140. Hall IP. Beta 2 adrenoceptor polymorphisms: are they clinically important? Thorax 1996; 51(4):351–353.
141. Tan S, Hall IP, Dewar J, Dow E, Lipworth B. Association between beta 2-adrenoceptor polymorphism and susceptibility to bronchodilator desensitisation in moderately severe stable asthmatics. Lancet 1997; 350(9083):995–999.
142. Martinez FD, Graves PE, Baldini M, Solomon S, Erickson R. Association between genetic polymorphisms of the beta2-adrenoceptor and response to albuterol in children with and without a history of wheezing. J Clin Invest 1997; 100(12):3184–3188.
143. Hancox RJ, Sears MR, Taylor DR. Polymorphism of the beta2-adrenoceptor and the response to long-term beta2-agonist therapy in asthma. Eur Respir J 1998; 11(3):589–593.
144. Lima JJ, Thomason DB, Mohamed MH, Eberle LV, Self TH, Johnson JA. Impact of genetic polymorphisms of the beta2-adrenergic receptor on albuterol bronchodilator pharmacodynamics. Clin Pharmacol Ther 1999; 65(5):519–525.
145. Lipworth BJ, Hall IP, Tan S, Aziz I, Coutie W. Effects of genetic polymorphism on ex vivo and in vivo function of beta2-adrenoceptors in asthmatic patients. Chest 1999; 115(2):324–328.
146. Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci USA 2000; 97(19):10483–10488.
147. Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional CDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 1996; 272(5263):872–877.
148. Clapham PR, McKnight A. Cell surface receptors, virus entry and tropism of primate lentiviruses. J Gen Virol 2002; 83(Pt 8):1809–1829.
149. Kilby JM, Eron JJ. Novel therapies based on mechanisms of HIV-1 cell entry. N Engl J Med 2003; 348(22):2228–2238.
150. Fauci AS. Host factors and the pathogenesis of HIV-induced disease. Nature 1996; 384(6609):529–534.
151. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, Kang S, Ceradini D, Jin Z, Yazdanbakhsh K, Kunstman K, Erickson D, Dragon E, Landau NR, Phair J, Ho DD, Koup RA. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 1996; 2(11): 1240–1243.
152. Michael NL, Nelson JA, KewalRamani VN, Chang G, O'Brien SJ, Mascola JR, Volsky B, Louder M, White GC, Littman DR, Swanstrom R, O'Brien TR. Exclusive and persistent use of the entry coreceptor CXCR4 by human immunodeficiency virus type 1 from a subject homozygous for CCR5 delta32. J Virol 1998; 72(7):6040–6047.
153. Ansari-Lari MA, Liu XM, Metzker ML, Rut AR, Gibbs RA. The extent of genetic variation in the CCR5 gene. Nat Genet 1997; 16(3):221–222.

154. Carrington M, Kissner T, Gerrard B, Ivanov S, O'Brien SJ, Dean M. Novel alleles of the chemokine-receptor gene CCR5. Am J Hum Genet 1997; 61(6): 1261–1267.
155. Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, Huttley GA, Allikmets R, Schriml L, Gerrard B, Malasky M, Ramos MD, Morlot S, Tzetis M, Oddoux C, di Giovine FS, Nasioulas G, Chandler D, Aseev M, Hanson M, Kalaydjieva L, Glavac D, Gasparini P, Dean M. Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 1998; 62(6):1507–1515.
156. Mecsas J, Franklin G, Kuziel WA, Brubaker RR, Falkow S, Mosier DE. Evolutionary genetics: CCR5 mutation and plague protection. Nature 2004; 427(6975):606.
157. Galvani AP, Slatkin M. Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. Proc Natl Acad Sci USA 2003; 100(25):15276–15279.
158. Salkowitz JR, Bruse SE, Meyerson H, Valdez H, Mosier DE, Harding CV, Zimmerman PA, Lederman MM. CCR5 promoter polymorphism determines macrophage CCR5 density and magnitude of HIV-1 propagation in vitro. Clin Immunol 2003; 108(3):234–240.
159. Jendreyko N, Popkov M, Beerli RR, Chung J, McGavern DB, Rader C, Barbas CF III. Intradiabodies, bispecific, tetravalent antibodies for the simultaneous functional knockout of two cell surface receptors. J Biol Chem 2003; 278(48):47812–47819.
160. Steinberger P, Andris-Widhopf J, Buhler B, Torbett BE, Barbas CF III. Functional deletion of the CCR5 receptor by intracellular immunization produces cells that are refractory to CCR5-dependent HIV-1 infection and cell fusion. Proc Natl Acad Sci USA 2000; 97(2):805–810.
161. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med 2002; 4(2):45–61.
162. Clavel F, Hance AJ. HIV drug resistance. N Engl J Med 2004; 350(10): 1023–1035.
163. Parnas J. Peter ludwig panum: great danish pathologist and discoverer of endotoxin. Dan Med Bull 1976; 23(3):143–146.
164. Heppner G, Weiss DW. High susceptibility of strain A mice to endotoxin and endotoxin-red blood cell mixtures. J Bacteriology 1965; 90(3):696–703.
165. Hoshino K, Takeuchi O, Kawai T, Sanjo H, Ogawa T, Takeda Y, Takeda K, Akira S. Cutting edge: toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. J Immunol 1999; 162(7):3749–3752.
166. Medzhitov R, Preston-Hurlburt P, Janeway CA Jr. A human homologue of the drosophila toll protein signals activation of adaptive immunity. Nature 1997; 388(6640):394–397.
167. Beutler B, Poltorak A. The sole gateway to endotoxin response: how LPS was identified as Tlr4, and its role in innate immunity. Drug Metab Dispos 2001; 29(4 Pt 2):474–478.
168. Kuhns DB, Long Priel DA, Gallin JI. Endotoxin and IL-1 hyporesponsiveness in a patient with recurrent bacterial infections. J Immunol 1997; 158(8):3959–3964.

169. Michel O, Kips J, Duchateau J, Vertongen F, Robert L, Collet H, Pauwels R, Sergysels R. Severity of asthma is related to endotoxin in house dust. Am J Respir Crit Care Med 1996; 154(6 Pt 1):1641–1646.
170. Schwartz DA, Thorne PS, Yagla SJ, Burmeister LF, Olenchock SA, Watt JL, Quinn TJ. The role of endotoxin in grain dust-induced lung disease. Am J Respir Crit Care Med 1995; 152(2):603–608.
171. Collins FS, McKusick VA. Implications of the human genome project for medical science. J Am Med Assoc 2001; 285(5):540–544.
172. Peltonen L, McKusick VA. Genomics and medicine. Dissecting human disease in the postgenomic era. Science 2001; 291(5507):1224–1229.
173. Guttmacher AE, Collins FS. Welcome to the genomic era. N Engl J Med 2003; 349(10):996–998.
174. The International HapMap Consortium. The international HapMap project. Nature 2003; 426(18/25 Dec):789–796.

5

Pharmacogenetics of Drug Transporters

Catia Marzolini, Rommel G. Tirona, and Richard B. Kim

Division of Clinical Pharmacology, Vanderbilt University, Nashville, Tennessee, U.S.A.

1. INTRODUCTION

The translocation of drugs across biological membranes has generally been considered to be predominantly a passive process. However, recent advances relating to the cloning and expression of individual transporters have revealed the presence of a wide variety of transporters capable of drug uptake as well as efflux transport (Table 1) thus providing a mechanistic basis for carrier-mediated transport. Uptake transporters act by facilitating the translocation of drugs into cells. Their driving force is mainly provided by an exchange or cotransport of intracellular and/or extracellular ions (i.e., Na^+, H^+, or HCO_3^-), therefore, their activity is dependent on a sustained electrochemical gradient (1). Included within this class of transporters are members of the organic anion transporting polypeptide (OATP) family, organic anion transporter (OAT), organic cation transporter (OCT), novel organic cation transporter (OCTN), and peptide transporter (PepT) family. By contrast, efflux transporters function to export drugs from the intracellular to the extracellular milieu. Efflux carriers are mainly represented by the ATP-binding cassette (ABC) family of transmembrane transporters. Their driving force is provided by the hydrolysis of ATP so that substrates can be pumped against steep concentration gradients (2). Included within this class of transporters are members of the multidrug resistance proteins (MDR), multidrug resistance-related protein (MRP) family, bile salt export

Table 1 Summary of Drug Transporters

Transporter	Gene symbol	Chromosome	Tissues	Polarity	Ref. seq./ Gene bank
OATP-A	SLCO1A2	12p12	B, K, I	AP	NM_005075
OATP-B	SLCO2B1	11q13	B, L, K, I	BL	NM_007256
OATP-C	SLCO1B1	12q	L	BL	NM_006446
OATP-D	SLCO3A1	15q26	Ub	nd	NM_013272
OATP-E	SLCO4A1	20q13.33	Ub	nd	NM_016354
OATP-F	SLCO1C1	12p12.3	B	nd	NM_017435
OATP-H	SLCO4C1	5q21	K	BL	NM_180991
OATP-I	SLCO6A1	5q21	T	nd	NM_173488
OATP-J	SLCO5A1	8q13.2	nd	nd	NM_030958
OATP8	SLCO1B3	12p12	L	BL	NM_019844
PGT	SLCO2A1	3q21	Ub	nd	NM_005630
OAT1	SLC22A6	11q13.1-q13.2	K, B	BL	NM_004790
OAT2	SLC22A7	6p21.2-p21.1	L, K	BL	NM_006672
OAT3	SLC22A8	11q11.7	K, B	BL	NM_004254
OAT4	SLC22A11	11q13.1	K, P	AP	NM_018484
OCT1	SLC22A1	6q26	K, L, B, I	BL	NM_003057
OCT2	SLC22A2	6q26	K, B, I	BL	NM_003058
OCT3	SLC22A3	6q27	P, K, B, I	nd	NM_021977
OCTN1	SLC22A4	5q31.1	K, L	AP	NM_003059
OCTN2	SLC22A5	5q31	K, L, B, I	AP	NM_003060
PepT1	SLC15A1	13q33-q34	I, K	AP	NM_005073
PepT2	SLC15A2	3q21.1	K, B	AP	NM_021082
MDR1	ABCB1	7q21.1	K, L, B, I, P	AP	NM_000927
MDR3	ABCB4	7q21.1	L	AP	NM_000443
BSEP	ABCB11	2q24	L	AP	NM_003742
MRP1	ABCC1	16p13.1	Ub	BL	NM_004996
MRP2	ABCC2	10q24	K, L, B, I	AP	NM_000392
MRP3	ABCC3	17q22	L, A, P, K, I	BL	NM_003786
MRP4	ABCC4	13q32	Ub	AP/BL	NM_005845
MRP5	ABCC5	3q27	Ub	BL	NM_005688
MRP6	ABCC6	16p13.1	L, K	BL	NM_001171
MRP7	ABCC10	6p21.1	K, B	nd	NM_033450
MRP8	ABCC11	16q12.1	L, K, Lu	nd	NM_032583
MRP9	ABCC12	16q12.1	L, K, Lu	nd	NM_033226
BCRP	ABCG2	4q22	L, I, P	AP	NM_004827

Abbreviations: A, adrenal gland; AP, apical; B, brain; BL, basolateral; I, intestine; K, kidney; L, liver; Lu, lung; nd, not determined; P, placenta; T, testis; Ub, ubiquitous

pump (BSEP), and the breast cancer resistance protein (BCRP) (Table 1). Transporters, whether they mediate uptake or efflux, tend to be localized to certain key organs such as the intestine (3), liver (4), kidney (5), brain (6), and are therefore critical modulators of drug absorption, tissue distribution, and elimination (Fig. 1A–C).

Genetic polymorphisms in drug metabolizing enzymes have long been recognized to be responsible for interindividual variability in drug response and adverse reactions (7). There is now increasing evidence to suggest that genetic heterogeneity in drug transporters may also contribute to interindividual and population variability in drug disposition.

The goal of this review is to summarize the current state of knowledge in the area of pharmacogenetics of human transporters known to be involved in drug disposition. The molecular, biochemical, and physiological role of each transporter will be systematically outlined (Table 1) as well as the functional relevance of genetic polymorphisms in such transporters (Tables 2 and 3, www.vanderbilt.edu/kimlab).

2. ORGANIC ANION TRANSPORTING POLYPEPTIDE FAMILY

The organic anion transporting polypeptides (OATPs) represent a family of important proteins involved in the membrane transport of a large number of endogenous and therapeutic compounds. Members of the OATPs are expressed in multiple tissues such as the liver, kidney, brain, and intestine (8). In the liver, all known OATPs are localized to the basolateral membrane and facilitate the hepatocellular accumulation of drug substrates prior to metabolism and efflux transporter-mediated excretion into bile. Endogenous substrates include bilirubin, prostaglandins, bile acids, steroids, and endogenous peptides like cholecystokinin, indicating that these transporters participate in a variety of important physiological processes. Xenobiotic substrates include organic anions, cations, neutral, or zwitterionic compounds as well as certain peptidomimetic agents, suggesting these transporters are likely to play an important role in drug disposition. Currently 11 human OATPs have been identified. Due to the growing number of newly identified OATP transporters and the potential confusion relating to their nomenclature, a new classification based on evolutionary conservation and degree of amino acid sequence identities has recently been accepted by the HUGO Gene Nomenclature Committee (updated nomenclature found at www.kpt.unizh.ch/oatp/).

Given the broad substrate specificity and the localization in various organs, it seems likely that for some drugs functional genetic variations may be a basis for interindividual differences in drug disposition and response. Specific details on substrates specificities and distribution as well as information regarding the identities and functional relevance of known genetic polymorphisms will be provided in the following sections.

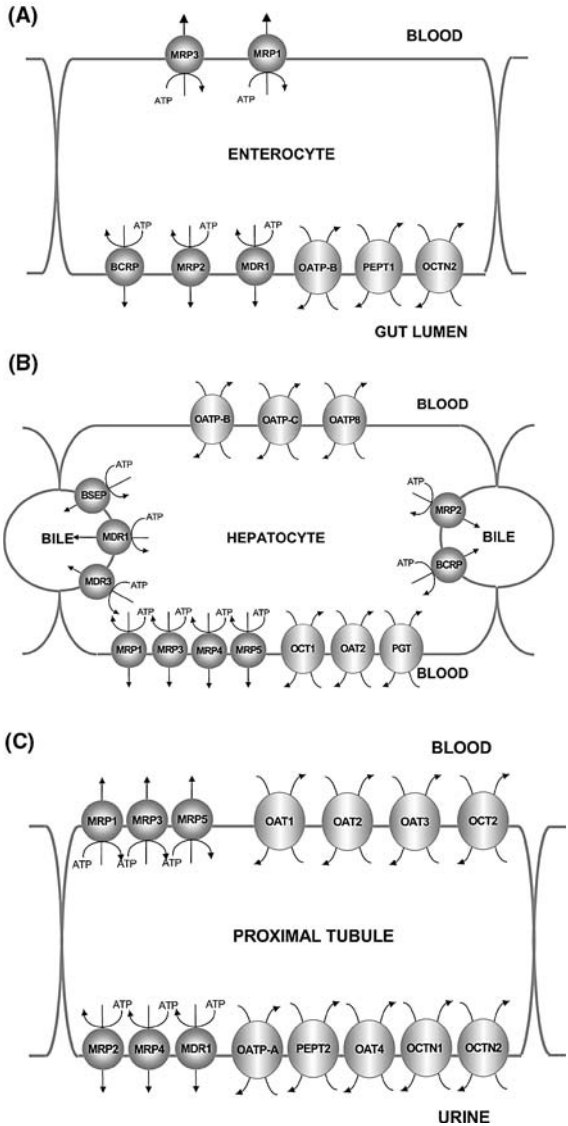


Figure 1 Schematic diagram depicting drug transporters and their localization in the human small intestinal enterocyte (**A**), hepatocyte (**B**), and renal tubular cell (**C**).

2.1. OATP-A (OATP1A2)

Rat Oatp1, the first identified member of this family, was cloned from a liver cDNA library (9). In the rat kidney, Oatp1 protein is localized to the apical proximal tubules (10), suggesting that this transporter facilitates solute

Table 2 Consequences of MDR1 Polymorphisms on Pharmacokinetics of Drug Substrate and on Protein Expression/Function

				Pharmacokinetics	
SNP	Study	Population	Drug	Parameter	Effect
C3435T exon 26	Hoffmeyer et al. (114)	21 Caucasians	Digoxin (po, ss)	Drug level	Higher for TT
	Johne et al. (120)	24 Caucasians	Digoxin (po, ss)	AUC (0–4 hr), Cmax	Higher for TT[a]
	Verstuyft et al. (121)	32 Caucasians +A	Digoxin (po, sd)	AUC	Higher for TT
	Gerloff et al. (122)	50 Caucasians	Digoxin (po, sd)	AUC (0–4 hr), Cmax	No difference
	Sakaeda et al. (123)	114 Japanese	Digoxin (po, sd)	AUC	Lower for TT
	Horinouchi et al. (124)	117 Japanese	Digoxin (po, sd)	AUC, tmax	Lower for TT
	Kurata et al. (125)	15 Japanese	Digoxin (po/iv, sd)	Bioavailability	Higher for TT[b]
	Drescher et al. (126)	20 Caucasians	Fexofenadine	AUC	No difference
	Kim et al. (116)	60 Caucasians + AA	Fexofenadine	AUC	Lower for TT
	Fellay et al. (127)	123 Caucasians	Nelfinavir	Drug level	Lower for TT
	von Ahsen et al. (128)	124 RT Caucasians	Cyclosporine	Trough level	No difference
	Min and Ellingrad (129)	14 Caucasians + AA	Cyclosporine	AUC, Cmax, tmax	No difference
	Yates et al. (130)	10 Caucasians + AA RT patients	Cyclosporine	AUC, Cmax	lower for CT/TT
	Mai et al. (131)	98 Caucasians RT patients	Cyclosporine	AUC (0–12), trough level	No difference[b]
	Haufroid et al. (132)	50 Caucasians RT patients	Cyclosporine	Trough level	No difference[c]
	Chowbay et al. (133)	14 Asian HT patients	Cyclosporine	AUC (0–4 hr; 0–12), Cmax	Higher for TT[c]
	Balram et al. (134)	10 Chinese HT patients	Cyclosporine	AUC (0–4 hr)	Higher for TT (trend)

(*Continued*)

Table 2 Consequences of MDR1 Polymorphisms on Pharmacokinetics of Drug Substrate and on Protein Expression/Function (*Continued*)

				Pharmacokinetics	
SNP	Study	Population	Drug	Parameter	Effect
	Kuzuya et al. (135)	97 Japanese RT patients	Cyclosporine	AUC (0–2 hr)	No difference
	Macphee et al. (136)	180 RT patients	Tacrolimus	Drug level	Higher for TT
	Zheng et al. (137)	69 HT pediatric patients	Tacrolimus	Drug level	Higher for CT/TT[d]
	Anglicheau et al. (138)	81 Caucasians + A RT patients	Tacrolimus	Drug level	No difference
	Haufroid et al. (132)	50 Caucasians RT patients	Tacrolimus	Trough level	No difference[c]
	Goto et al. (139)	69 Japanese LT patients	Tacrolimus	Drug level, dose ratio	No difference
	Yamauchi et al. (140)	17 Japanese LT patients	Tacrolimus	Tacr. induced neurotoxicity	No difference
	Roberts et al. (141)	78 Caucasians	Nortriptyline	Drug level	No difference
	Siegmund et al. (142)	67 Caucasians	Talinolol	AUC	No difference
	Pauli-Magnus et al. (143)	16 Caucasians	Loperamide	Cmax, AUC, CNS effects	No difference
	Skarke et al. (144)	21 volunteers	Loperamide	Cmax, AUC, pupil size	Higher for TT[a,h]
	Putnam et al. (145)	17 volunteers	Dicloxacillin	Drug level	No difference
	Goh et al. (146)	32 Asian patients	Docetaxel	Clearance	No difference
	Sai et al. (147)	49 Japanese patients	irinotecan	Clearance	Lower for TT[c]
	Kerb et al. (148)	96 Turkish	Phenytoin	Drug level	Higher for TT
G2677T/A	Verstuyft et al. (121)	32 Caucasians + African	Digoxin (po, sd)	AUC	Higher for TT (trend)
exon 21	Gerloff et al. (122)	50 Caucasians	Digoxin (po, sd)	AUC (0–4 hr), Cmax	No difference

Horinouchi et al. (124)	117 Japanese	Digoxin (po, sd)	AUC, tmax	Lower for TT
Kurata et al. (125)	15 Japanese	Digoxin (po/iv, sd)	Bioavailability	Higher for TT[b]
Kim et al. (116)	60 Caucasians + AA	Fexofenadine	AUC	Lower for TT
Haufroid et al. (132)	50 Caucasians RT patients	Cyclosporine	Trough level	No difference[c]
Chowbay et al. (133)	14 Asian HT patients	Cyclosporine	AUC (0–4 hr; 0–12), Cmax	Higher for TT[c]
Kuzuya et al. (135)	97 Japanese RT patients	Cyclosporine	AUC (0–2 hr)	No difference
Haufroid et al. (132)	50 Caucasians RT patients	Tacrolimus	Trough level	No difference[c]
Zheng et al. (137)	69 HT pediatric patients	Tacrolimus	Drug level	Higher for GT/TT[d]
Anglicheau et al. (138)	81 Caucasians + A RT patients	Tacrolimus	Drug level	Higher for TT
Goto et al. (139)	69 Japanese LT patients	Tacrolimus	Drug level, dose ratio	No difference
Yamauchi et al. (140)	17 Japanese LT patients	Tacrolimus	Tacr, induced neurotoxicity	Higher for TT (trend)
Siegmund et al. (142)	67 Caucasians	Talinolol	AUC	Slightly higher for TA/TT
Sai et al. (147)	49 Japanese patients	Irinotecan	Clearance	Lower for TT[c]

[a]Haplotype 12: 2677GT + 3435TT is associated with higher AUC, haplotype 11: 2677GG + 3435CC is associated with AUC.
[b]Haplotype analysis performed for the polymorphisms in exons 21 and 26.
[c]Haplotype analysis performed for the polymorphisms in exons 12, 21, and 26.
[d]No correlation of 3 months post transplantation.
[e]The genotype affect CYP3A4 mRNA level (TT < CT < CC).
[f]Verapamil, digoxin, vinblastine, cyclosporine.
[g]Paclitaxel, verapamil, daunorubicin, vinblastine, prazosin, bisantrene, forskolin, calcein.
[h]Most pronounced increase of the miotic effects of loperamide with quinidine co-administration.
Abbreviations: A, African; AA, African American; AML, acute myeloid leukemia; AUC, area under the curve; Cmax, maximum concentration; HT, heart transplant; iv, intravenous administration; LT, liver transplant; po, oral administration; RCC, renal cell carcinoma; RT, renal transplant; sd, single-dose; ss, steady-state; tmax, time to peak plasma drug concentration.

Table 3 Consequences of MDR1 C3435T Polymorphism on Disease Outcome

			Clinical investigations	
Study	Disease	Population	Outcome marker	Effect
Siegsmund et al. (152)	Renal epithelial tumor	212 Caucasian patients	Susceptibility to renal tumor	Higher in TT[a]
Atasanova et al. (153)	Endemic nephropathy	96 patients	Susceptibility to nephropathy	Lower in TT[a,b]
Illmer et al. (154)	Acute myeloid leukemia	405 patients	Overall survival	Higher in TT[c]
Efferth et al. (155)	Acute lymphoblastic leukemia	53 patients	Overall survival	No statistical difference
Kafka et al. (156)	Breast cancer	68 patients	Response to preop. chemotherapy	Higher in TT
Schwab et al. (157)	Ulcerative colitis	149 Caucasians	Susceptibility to ulcerative colitis	Higher in TT[a]
Schwab et al. (157)	Crohn's disease	126 Caucasian patients	Susceptibility to Crohn's disease	No statistical difference[a]
Zheng et al. (158)	Heart transplantation (HT)	69 pediatric patients	Steroid weaning 1 year post-HT	Higher in TT[d]
Asano et al. (159)	Kidney transplantation (KT)	136 patients	Steroid-induced osteonecrosis	Lower in TT
Furuno et al. (160)	Parkinson's disease (PD)	95 Caucasian patients	Susceptibility to PD	No statistical differences[a,e]
Drozdzik et al. (161)	Parkinson's disease (PD)	107 Caucasian patients	Susceptibility to PD	No statistical differences[a]
Drozdzik et al. (161)	Parkinson's disease (PD)	59 pesticides exposed patients	PD post exposition to pesticides	Higher in CT[f]
Siddiqui et al. (162)	Epilepsy	315 patients	Response to antiepileptic	Higher in TT[a,g]

Fellay et al. (127)	HIV infection	80 Caucasian patients	CD4 response to treatment	Higher in TT
Nasi et al. (163)	HIV infection	149 Caucasian patients	CD4 response to treatment	No statistical difference
Ifergan et al. (164)	HIV infection	137 Caucasians exposed to HIV	Host susceptibility to HIV infection	No statistical difference[h]
Bleiber et al. (165)	HIV infection	411 patients	HIV progression before treatment	No statistical differences[e]
Brumme et al. (166)	HIV infection	461 patients	Length of virological suppression	Higher in TT (trend p: 0.07)
Haas et al. (167)	HIV infection	31 patients	Phase 1 viral decay	No statistical difference
Roberts et al. (141)	Depression	78 Caucasian patients	Nortriptyline-induced hypotension	Higher in TT
De Luca et al. (168)	Depression	55 patients	Antidepressant-induced mania	No statistical difference

[a]Compared to healthy controls.
[b]Decreased risk for carries of the haplotype 12 (2677G/3435T).
[c]Overall survival was also higher when considering MDR1 2677TT (exon 21) and 1236TT (exon 12) genotypes.
[d]Steroid weaning was also faster when considering MDR1 2677TT (exon 21) genotype.
[e]No statistical difference also when considering exon 21 polymorphism.
[f]Compared to PD patients not exposed to pesticides.
[g]Compared to patients with drug-responsive epilepsy.
[h]Comparison between infected and non infected Caucasians highly exposed to HIV.

reabsorption from urine. Rat Oatp1 is also expressed on the apical surface of the choroids plexus (11) and acts to remove organic solutes from the cerebrospinal fluid. The cloning of rat Oatp1 soon led to the isolation the first human OATP, OATP-A, from human liver (12). Although OATP-A was originally identified in human liver, its level of expression in this organ is low and substantially higher mRNA levels were found in kidney and in the brain (13). The range of OATP-A substrates is wide since it is capable of transporting structurally diverse compounds including bromosulfophthalein (12), bile acids (12), steroid sulfates (14), bulky organic cations (15), thyroid hormones (16), opioid peptides (17), and fexofenadine (18). Single-nucleotide polymorphisms (SNPs) have been described for OATP-A in the 5′-flanking region (19) and only recently in exonic regions (20) (www.vanderbilt.edu/kimlab). Interestingly, it was reported that the presence of a non-synonymous SNP at codon 172 significantly reduced the transport activity of estrone sulfate and fexofenadine in cells expressing the Q172D variant. The in vivo relevance of sequence variations in OATP-A-mediated CNS drug entry will require further clinical investigations.

2.2. OATP-B (OATP2B1)

Tamai et al. (8) cloned human OATP-B from a brain cDNA library and transporter mRNA was also noted to be expressed in various tissues including the liver, kidney, lung, placenta, heart, and small intestine (8). In liver, OATP-B is localized to the basolateral membrane (16) indicating a blood to liver uptake role for this transporter. OATP-B exhibits restricted substrate specificity, thus its importance in hepatic drug elimination is currently unclear. Interestingly, recent studies indicated that the uptake of estrone-3-sulfate and pravastatin by OATP-B was higher at acidic pH (pH 5.5) (21). Since the physiological pH close to the surface of intestinal epithelial cells is acidic, the role of OATP-B in the small intestine might be different from that in other tissue (22). Nozawa et al. (23) described the presence of a non-synonymous SNP in OATP-B at codon 486 that reduces the uptake rate of estrone sulfate in cells expressing the variant transporter. The functional importance of this variant to the in vivo disposition of substrate drugs remains to be defined.

2.3. OATP-C (OATP1B1)

The OATP-C, also referred to as liver specific transporter 1 (LST-1) or OATP2, was cloned by number of groups (8,24,25) and was shown to be exclusively expressed in the basolateral membrane of the liver (25). As with OATP-A, OATP-C has broad substrate specificity and transports bile acids (13), sulfate and glucuronide conjugates (25), thyroid hormones (13), and peptides (16), and drugs such as pravastatin (24), methotrexate (26), and rifampin (27). Given its liver-specific expression and broad substrate

specificity, it is likely that OATP-C polymorphisms might influence drug disposition. Several groups (23,26,28,29) screened the coding regions the OATP-C gene and various synonymous and non-synonymous SNPs were identified whose allelic frequencies were dependent on race (www.vanderbilt.edu/kimlab). In vitro experiments with cultured cells expressing the wild type and mutated OATP-C revealed that several variants exhibited markedly reduced uptake of estrone's sulfate and 17β-D-glucuronide conjugates (26). Alterations in transport were specifically associated with SNPs that introduced amino acid changes within the transmembrane-spanning domains and extracellular loop 5. The common amino acid variants with altered transporter function were V174A (OATP-C*5) in European-Americans and G488A (OATP-C*9) in African-Americans. Interestingly, the variant A1964G reduced the transport activity of estrone sulfate but not that of 17β-D-glucuronide, suggesting possible substrate-dependent polymorphisms. Cell surface trafficking defects proved to be responsible for altered transport function of many of these OATP-C variants (26). Recently, Nishizato et al. (29) have investigated the in vivo importance of genetic OATP-C variants on pravastatin pharmacokinetics and demonstrated that the subjects carrying the OATP-C*15 allele [a haplotype consisting of OATP-C*1B (Asp130) and *5 (Ala174)] had increased pravastatin plasma levels as compared to individuals carrying only the OATP-C*1B allele (Asp130Val174) (29).

These findings suggest much of the loss of function defect associated with the OATP-C*15 haplotype is related to the V174A mutation. Interestingly, Tirona et al. (27) demonstrated that various SNPs in OATP-C markedly reduced the hepatocellular uptake of rifampin. This finding suggests that carriers of these functionally deficient OATP-C variants may exhibit reduced capacity for rifampin-mediated induction of hepatic drug metabolizing enzymes and transporters (27).

2.4. OATP8 (OATP1B3)

OATP8 is similar to OATP-C in terms of amino acid identity (80%), liver-specific tissue distribution and substrate specificity (25). Although OATP-C and OATP8 exhibit broad overlapping substrate specificities, OATP8 is unique in transporting digoxin and has a high transport activity for anionic peptides and cholecystokinin-8 (16). Bile salts, substrates for OATP-C, are not transported by OATP8 (30). While a number of SNPs in OATP8 have been identified, their functional significance has not been reported (19).

2.5. OATP-D, OATP-E, OATP-F, OATP-H, OATP-I, OATP-J, and PGT

The OATP-D was cloned from human brain, but noted to be ubiquitously expressed (31). It has been shown to transport prostaglandin E1/E2 and $F_{2\alpha}$ (31). OATP-E was cloned from human kidney and is also ubiquitously

expressed (8). Some substrates transported by OATP-E include estrone sulfate (8), prostaglandin E_2 (8), taurocholate (32), and thyroid hormones (32). OATP-F has been cloned recently from human brain and displays a noticeable high affinity for thyroxin (T_3, T_4). This transporter could be important for thyroid hormone disposition in brain (33). OATP-H was recently identified as a digoxin transporter at the basolateral membrane of renal proximal tubular cells (34). The OATP-J sequence has been reported in the database and its functional properties are currently under investigation. OATP-I has been isolated from human testis but so far no functional characterization has been published (35). Finally, the prostaglandin carrier PGT was cloned from human kidney and has been detected in several tissues (36). Further studies are required to determine the impact of these transporters on the drug distribution and elimination as well as the consequences of genetic polymorphisms.

3. ORGANIC ANION TRANSPORTER FAMILY

Members of the organic anion transporters (OATs) are structurally distinct from OATP transporters and appear to have important functions, especially in the kidney. Whereas, larger and more lipophilic organic anions are extracted in the liver by OATPs, small hydrophilic organic anions are efficiently transported by OATs, which are highly expressed on the basolateral and apical membranes of renal proximal tubules (Fig. 1C). The OAT substrates include a wide variety of clinically important anionic drugs such as β-lactam antibiotics, diuretics, non-steroidal anti-inflammatory drugs (NSAIDs), nucleoside/nucleotide antiviral drugs, and anticancer agents. Currently four human OATs have been identified.

3.1. OAT1 (*SLC22A6*)

The prototypical organic anion transporter in the kidney was first cloned from the rat (Oat1) (37), followed by the subsequent identification of a human ortholog, which was cloned as two isoforms as a result of alternative splicing (38–40). The two variants, OAT1-1 (563 amino acids) and OAT1-2 (550 amino acids) appeared to be functionally similar (38). Recently, two new splice variants were found expressed in kidney (OAT1-3 and OAT1-4), however, they do not appear functional (41). OAT1 is expressed at the basolateral membrane of the proximal tubule cells where it mediates the uptake of small and hydrophilic organic anions such as *p*-aminohippurate (PAH) from blood using an exchange mechanism for intracellular dicarboxylates (e.g., α-KG^-) (39,40). OAT1 has been reported to transport PAH and antiviral agents such as adefovir, cidofovir, zidovudine, acyclovir, and ganciclovir (42,43). To date, several synonymous and non-synonymous SNPs have been reported (www.vanderbilt.edu/kimlab). Preliminary studies

of variants (R50H), (P104L), and (A190Y) showed no alteration of OAT1 transport function in vitro experiments (44,45). Although (R50H) did not affect PAH handling by OAT1, this amino acid change resulted in increased affinity for cidofovir and adefovir compared to the wild-type transporter. However, more detailed studies are required with a variety of OAT1 drug substrates to determine the clinical relevance of OAT1 SNPs to renal drug elimination and drug-associated nephrotoxicity.

3.2. OAT2 (*SLC22A7*)

Although, OAT transporters are mainly expressed in the kidney, rat Oat2 was initially identified in the liver and named NLT (novel liver-specific transport protein) (46). Subsequently, Sekine et al. found that NLT transports organic anions and proposed renaming NLT as rat Oat2 (47). Human OAT2 was isolated from liver but was also found to a lesser extent in the basolateral side of the proximal tubules (48). OAT2 transports relatively small and hydrophilic organic anions such as indomethacin and salicylate (49). To date, little is known about the existence of genetic polymorphisms in OAT2 and their possible functional consequences.

3.3. OAT3 (*SLC22A8*)

The OAT3 was cloned from human kidney where it is found at high levels. The OAT3 protein is expressed at the basolateral membrane of renal proximal tubules (50) where it mediates uptake of estrone sulfate, estradiol, estradiol 17β-D-glucuronide, dehydroepiandrosterone sulfate, prostaglandins, PAH, methotrexate, cimetidine, NSAIDs, a number of antiviral drugs, taurocholate, bile acids, tetraethylammonium (TEA) (50), and pravastatin (51). The Oat3 knockout mice have significantly reduced renal uptake of organic anions (52). In addition, Oat3 null mice with absent transporter expression at the apical membrane of the choroid plexus exhibit decreased CSF uptake (52,53). Recently, Nishizato et al. (29) found several polymorphisms in OAT3 and assessed their effect on pravastatin pharmacokinetics. They demonstrated that variants in OAT3 were not associated with changes in renal and tubular secretory clearance of pravastatin, on the contrary and as discussed previously, polymorphisms in OATP-C appeared to be a critical determinant. However, given the broad substrate specificity of OAT3 and its key role in the regulation of renal and CSF anionic drug clearance, further studies are needed in order to assess possible functional consequences of polymorphisms in this transporter.

3.4. OAT4 (*SLC22A11*)

The OAT4 was cloned from human kidney but is also expressed at appreciable levels in the placenta (54). In kidney, OAT4 is localized to the apical

side of renal proximal tubules (55,56). Steroid sulfates and ochratoxin A are efficiently transported by OAT4 whereas PAH is only poorly transported (54). Zidovudine and methotrexate are also substrates of OAT4 (57,58). OAT4 transport appears to be driven by dicarboxylate exchange (56). The role of OAT4 in regulating renal drug elimination and placental permeability as well as the existence of genetic polymorphisms remains to be clarified.

4. ORGANIC CATION TRANSPORTER FAMILY

The organic cation transporters (OCTs) are capable of transporting organic cations in a number of epithelial tissues including the intestine, liver, and kidney. These transporters mediate the uptake of a variety of structurally divergent compounds such as endogenous cations (i.e., guanidine, monoamine neurotransmitters, and *N*-methylnicotinamide), cationic toxins (i.e., 1-methyl-4-phenylpyridinium, MPP^+), and cationic drugs (TEA, cimetidine, procainamide, quinidine, and cardiac glycosides). Currently three human OCTs have been identified.

4.1. OCT1 (*SLC22A1*)

The first human member of the OCT family, OCT1 was cloned from human liver (59). OCT1 mRNA was detected predominantly in liver and to a lesser extent in kidney and intestine (59,60). The OCT1 mediates the uptake of several substrates drugs including TEA (60), quinidine (61), and certain antiviral drugs (57). Interestingly, different substrate selectivities have been demonstrated for human OCT1 as compared to rat Oct1 (62) suggesting that OCT1 may be partly responsible for interspecies differences in the disposition of organic cations. Genetic variations in OCT1 have been identified and functionally characterized in, in vitro systems (63–65). Four polymorphisms, namely R61C, C88R, P341L, and G401S, resulted in a reduction of transport activity. Interestingly, C88R and G401S variants could mediate significant uptake of TEA and serotonin yet were unable to transport a prototypical OCT1 substrate, MPP^+. The P283L and R287G variants exhibited a complete loss of uptake for both TEA and MPP^+ (65). Although OCT1 is a major hepatic uptake transporter for cationic compounds, the clinical relevance of these polymorphisms remains to be delineated since there are other transporters with overlapping substrate specificity.

4.2. OCT2 (*SLC22A2*)

Okuda et al. (66) isolated rat Oct2 from kidney, which was subsequently followed by the identification of human OCT2 (60). The OCT2 mRNA has been detected in kidney and brain (60). The OCT2 has been localized

to the basolateral membrane of the renal proximal tubule (67) suggesting that OCT2 participates in the secretion of cations by proximal tubules. The OCT2 was also shown to be expressed on the apical side of the distal tubule (60). In the brain, OCT2 is known to mediate the sodium-independent transport of monoamine neurotransmitters and the antiparkinsonian drug amantadine (68). Recently, two variants, M165V and R400C, were found to be associated with altered transport activity in vitro (69). Similar to OCT1, the clinical relevance of these polymorphisms awaits further studies.

4.3. OCT3 (*SLC22A3*)

The OCT3 was cloned from a human kidney carcinoma cell line (70). The OCT3 mRNA can be detected in multiple tissues including intestine, kidney, and brain with the highest expression noted in placenta (71). This transporter mediates the uptake of neurotransmitters such as epinephrine, norepinephrine, and MPP^+ (70). To date, several synonymous variants have been reported for OCT3 (72,73) (www.vanderbilt.edu/kimlab).

5. NOVEL ORGANIC CATION TRANSPORTER FAMILY

The novel organic cation transporters (OCTNs) represent a subfamily of OCTs with only 30% amino acid sequence similarity to the OCTs. Two members have been identified.

5.1. OCTN1 (*SLC22A4*)

The OCTN1 was initially cloned from human fetal liver and its mRNA has been detected in various tissues including skeletal muscle and kidney but not in adult liver (74). The OCTN1 transports cations such as TEA, quinidine, verapamil, and the zwitterion carnitine (75). Several non-synonymous SNPs have been reported for OCTN1, however, their functional consequences are currently unknown (64). Recently, an SNP in intron 1 of OCTN1 was found to be associated with rheumatoid arthritis (76). This study established that OCTN1 was highly expressed in hematopoietic tissues, leukocytes, and inflammatory joints and that alteration of the transcription factor RUNX1 activity at this polymorphic site influences OCTN1 transcription and susceptiblity to rheumatoid arthritis (76).

5.2. OCTN2 (*SLC22A5*)

Human OCTN2 was cloned from a human placental trophoblast cell line and human kidney (77,78). OCTN2 is expressed in skeletal muscle, kidney, lung, pancreas, small intestine, adrenal gland, thyroid, and brain (77,78). OCTN2 mediates Na^+-independent transport of TEA whereas

carnitine uptake is mediated in a sodium-dependent fashion (77,78). Expression of rat Octn2 in the apical membrane of renal proximal tubule cells (79) supports the notion that this transporter may be essential for the reabsorption of filtered carnitine. Mutations in the OCTN2 gene have been causally linked to primary systemic carnitine deficiency, an autosomal recessive disease characterized by hypoglycemia, progressive cardiomyopathy, and skeletal myopathy due to impaired fatty acid oxidation (80–91) (www.vanderbilt.edu/kimlab). Functional assessment of several OCTN2 variants has demonstrated a loss of carnitine transport activity. Seth et al. (91) demonstrated that two mutations, namely P478L and L352R, resulted in a complete loss of carnitine transport function. Interestingly, P478L variant affected only carnitine transport, but not other organic cations. These findings suggest that the bindings sites for carnitine and organic cations overlap but are not identical.

6. PEPTIDE TRANSPORTER (PepT) FAMILY

Members of the PepT family are located on the apical membranes of intestinal and renal epithelial cells and mediate the uptake of several peptidomimetic drugs like β-lactam antibiotics, angiotensin-converting enzyme inhibitors, and nucleoside analogs. Currently, two human PepTs have been identified.

6.1. PepT1 (*SLC15A1*)

Human PepT1 was initially cloned from intestine (92) and was found to be localized to the brush border of intestinal epithelial cells (93) and in S1 segment of apical proximal tubules (94). PepT1 transports β-lactam antibiotics (95), antiviral drugs such as valacyclovir and valganciclovir (96), and the angiotensin converting enzyme inhibitor captopril (97). Polymorphisms have been reported, however little is known regarding their functional consequences (64).

6.2. PepT2 (*SLC15A2*)

The PepT2 was identified by Liu et al. (98) and was found to be expressed in the kidney but not in the intestine. Rat PepT2 is localized to the apical proximal tubule membrane and participates in the renal reabsorption of filtered substrates (94). Studies with knockout mice confirmed the importance of PepT2 in the renal tubular reabsorption of dipeptides (99). Moreover, apically expressed PepT2 at the choroid plexus regulates the efflux of dipeptides and peptidomimetic drugs from the CSF to the blood (100–102). These findings indicate an important role for PepT2 in renal clearance and brain penetration of drugs. Substrates for PepT2 are similar to

PepT1 although differences in affinity exist. Again, the clinical relevance of SNPs in PepT2 has not been established.

7. MULTIDRUG RESISTANCE FAMILY

The multidrug resistance P-glycoproteins (P-gps) (MDRs) represent a family of ATP-dependent transporters that are involved in the cellular efflux of numerous endogenous and exogenous compounds. This family has three members: MDR1 (P-gp), MDR3, and BSEP [a liver specific homolog of MDR1-P-gp also called sister-P-glycoprotein (SPGP)].

7.1. MDR1 (*ABCB1*)

P-gp, the gene product of MDR1, has been the most thoroughly investigated efflux-transporter. Human P-gp is a phosphorylated and glycosylated transmembrane protein that is 1280 amino acids long and is composed of two homologous and symmetrical sequences, each of which contains six transmembrane domains and an ATP-binding motif. P-gp functions as a transmembrane efflux pump thereby moving drugs from the intracellular to extracellular domain; it can also act on drug molecules entrapped within the cell membrane lipid bilayer. ATP hydrolysis provides the energy for active drug transport, enabling the transporter to function against steep concentration gradients.

P-gp was first isolated from colchicine-resistant Chinese hamster ovary cells (103). Subsequently, the gene coding for P-gp (MDR1) was identified because of its overexpression in tumor cells associated with an acquired cross-resistance to multiple cytotoxic anticancer agents (104). This drug transporter was also recognized to be expressed in many normal tissues, suggestive of a physiological function (105). P-gp is found on the canalicular surface of hepatocytes, the apical surface of proximal tubular cells in the kidney, the brush border surface of enterocytes, the epithelium of the brain choroid plexus as well as the luminal surface of blood capillaries in the brain, the placenta, the ovaries, the testes, and certain lymphocytes (105–107). The function and the localization of P-gp suggest that this transporter acts as a protective barrier to keep toxins out of the body by excreting these compounds into bile, urine, and the intestinal lumen and limiting their accumulation in critical organs such as the brain, fetus, gonads, and bone marrow. Interestingly, P-gp is co-localized with the drug-metabolizing enzyme CYP3A4 in the small intestine and liver suggesting that this transporter also plays a significant role in the absorption and elimination of drugs. Evidence supporting such a role was obtained from animal models. Both a mouse strain (CF-1 mice) and a genetic knock-out deficient in *mdr1a*, demonstrated marked sensitivity to the neurotoxic effects of ivermectin, an antiparasiticide substrate of P-gp (108,109). The absence of P-gp in the blood–brain barrier

of such mice resulted in a more than 80-fold higher brain accumulation of ivermectin, resulting in neurotoxicity (108). In addition to limiting the CNS entry of drugs, P-gp also reduced oral absorption of drugs by extruding them from enterocytes back into the intestinal lumen (110) or enhanced their biliary and renal excretion (111).

P-gp mediates the transport of a broad range of amphipathic hydrophobic substrates which includes a variety of pharmacologically distinct agents used in cancer chemotherapy, hypertension, allergy, infection, immunosuppression, neurology, and inflammation (see review by Marzolini et al. (112)).

Given the broad substrate specificity and the broad localization in various organs, it seems likely that genetic variability of this transporter may influence drug disposition as well as treatment efficacy. Accordingly, many studies related to the discovery of MDR1 polymorphisms and their effects on drug disposition or disease outcome have now been reported.

Genetic polymorphisms in *MDR1* were first identified by Kioka et al. (113) from in vitro studies with cancer cells. Subsequently other groups, including Hoffmeyer and coworkers screened the entire *MDR1* coding region (114–119). A synonymous SNP in exon 26 (C3435T) was the first variant to be associated with altered protein expression although the SNP does not change the encoded amino acid (Ile) (114). In the same study P-gp expression in duodenal biopsy samples among healthy Caucasians with the homozygous T allele (variant) was noted to be decreased when compared to those with the C allele (wild-type). Subjects with the variant allele were also shown to have increased digoxin plasma concentrations after oral admisnitration, suggesting greater drug absorption in individuals with low intestinal P-gp levels. As summarized in Table 2, the observation that the 3435T allele in exon 26 is associated with higher digoxin levels was further confirmed by studies by Johne et al. (120), Verstuyft et al. (121), and Kurata et al. (125) although these authors did not measure intestinal P-gp expression. However, this observation has recently been contradicted by the study of Gerloff et al. (122), which showed no differences in digoxin levels among healthy Caucasians carrying the 3435T allele or the wild-type 3435C allele. Furthermore, Sakaeda et al. (123) and Horinouchi et al. (124) demonstrated that the AUC of digoxin was in fact lower in Japanese healthy subjects carrying the 3435T allele. While quantifying MDR1 mRNA in the duodenum, Nakamura et al. (149) showed greater levels in healthy Japanese subjects carrying the 3435T allele in exon 26 as compared to those with the 3435C allele. This is opposite to the findings of Hoffmeyer et al. (114) who found lower levels of the protein in subject carrying the 3435T allele. Discordant clinical data have been noted for other P-gp drug substrates including fexofenadine (116,126), cyclosporine (128–135), and tacrolimus (136–140) (Table 2). Similarly conflicting findings have also been noted for the exon 21 (G2677T/A) polymorphism (Table 2). Several in vitro studies have also

tried to assess the functional consequences of *MDR1* polymorphisms (150,151). Such studies have so far failed to demonstrate a consistent association between the frequently observed SNPs in *MDR1* and alterations in P-gp transporter function.

The SNPs in *MDR1* may also alter the physiological protective role of P-gp and therefore influence disease risk. Recently, a number of studies describing to such an association with P-gp have been published (Table 3). For example, Siegsmund et al. (152) showed that in patients with renal cell carcinoma, the frequency of exon 26 3435T allele was significantly higher in comparison to healthy controls. Furthermore, the expressed level of P-gp in renal tissue was lower in patients with the 3435T genotype, suggesting that reduced renal P-gp expression could potentially expose the organism to higher concentrations of toxic P-gp substrates, thereby increasing organ exposure to the damaging effects of such agents. The *MDR1* exon 26 3435T allele has also been suggested to increase the susceptibility to Parkinson disease and especially among the patients with a history of exposure to pesticides (161). This assumes that pesticides, in general, are substrates of the P-gp and that CNS exposure to such agents may cause neurodegenerative diseases. Another interesting example of increased risk for disease associated with the exon 26 T allele relates to inflammatory bowel diseases such as ulcerative colitis and Crohn's disease (157). Interestingly, among *mdr1a* knockout mice, one age-related phenotype observed is colitis. The inflammation is thought to result from toxins produced by intestinal bacteria, which are normally kept out of the intestinal wall by *mdr1a* P-gp activity (169). The efficacy of HIV has also been studied in relation to *MDR1* polymorphisms. It is known that all the currently available HIV protease inhibitors are transported by P-gp. Recently, Fellay et al. (127) found a relationship between expressions of P-gp in peripheral blood mononuclear cells (PBMC) of HIV infected patients and CD4 lymphocyte response to treatment. Patients with the 3435T allele in exon 26 had a significantly greater rise in CD4 cell count 6 months after starting antiretroviral therapy. It was hypothesized that this benefit associated with the T allele could result from an enhanced HIV protease inhibitor penetration into CD4 cells. However, this observation has been contradicted by two studies in which no correlation was observed between polymorphisms in *MDR1* and changes of lymphocyte subset over time (164,167).

It should be noted that most of the *MDR1* genotype-based studies have focused on the exon 26 C3435T synonymous polymorphism. Several hypotheses have been put forward to account for its somewhat unexpected effect. One suggestion is that the Ile encoded by the synonymous SNP has reduced translation efficiency. Others have suggested allele specific differences in RNA folding may influence downstream mRNA splicing, processing, or translational control and regulation (170). It is also possible that certain SNPs increase mRNA stability and consequently results in increased

protein expression (171). An additional consideration is that of linkage disequilibrium. Multiple studies have demonstrated a linkage disquilibrium between the synonymous polymorphism in exon 26 and other SNPs in MDR1 including the exon 21 (G2677T/A) and exon 12 (C1236T), suggesting that the functional effects may be haplotype-dependent (116,120,124,138,139,142,154,160). Since the number of relevant haplotypes in *MDR1* has been shown to significantly differ in various populations based on racial ancestry, determination of haplotype may prove to be important when assessing the effects of *MDR1* SNPs to in vivo functional consequences. Moreover, dietary differences, which alter extent of *MDR1* expression (172) and probe drugs used to study *MDR1* function in vivo must also be considered as potential confounding factors, which may contribute to discordant or variable genotype–phenotype associations.

7.2. MDR3 (*ABCB4*)

Despite its close amino acid sequence identity with MDR1 (75%), MDR3 does not appear to encode a multidrug transporter. Instead, MDR3 is a phospholipid translocator that is highly expressed on the canalicular membrane of the hepatocyte (173). The function of MDR3 has been elucidated by disruption of its homolog in the mouse, the *mdr2* gene. Mdr2 (−/−) knockout mice do not secrete phospholipids and cholesterol into bile and develop severe liver disease characterized by inflammation of the portal tracts, proliferation of the bile ducts, and fibrosis (174). Mutations in MDR3 have been clearly linked to an array of cholestatic syndromes, including progressive familial intrahepatic cholestasis type 3 (PFIC3) (175–183), intrahepatic cholestasis of pregnancy (180–183), and a form of cholesterol gallstone disease (179) (www.vanderbilt.edu/kimlab). These phenotypes are the result of genetic lesions including nucleotide deletions, insertions, and missense mutations. Genetic variations in MDR3 (or BSEP, see below) do not appear to be linked to primary biliary cirrhosis or primary sclerosis cholangitis (184).

7.3. BSEP (*ABCB11*)

BSEP also known as sister-P-glycoprotein (SPGP) was originally cloned from pig liver (185). BSEP is localized on the canalicular membrane of hepatocytes and is responsible for the secretion of bile salts across the canalicular membrane into bile. BSEP appears to be the predominant bile salt efflux system for hepatocytes, and is a critical component in the enterohepatic circulation of bile acids. A number of mutations in the transporter were found to the basis for progressive familial intrahepatic cholestasis type 2 (PFIC2) (186–188). Mutations found in PFIC2 patients include frameshifts, missense mutations, and premature termination codons. Most PFIC2 patients lack immunohistochemically detectable BSEP in their liver. Recently, seven

human PFIC2 mutations in highly conserved regions were assessed using a full-lengh cDNA encoding rat Bsep (189). Five mutations (G238V, E297G, G982R, R1153C, and R1268Q) prevented the protein from trafficking to the apical membrane and E297G, D482G, G982R, R1153C, and R1268Q also abolished taurocholate transport activity. The D482G mutation did not affect the apical expression of BSEP but showed decreased transport activity (www.vanderbilt.edu/kimlab).

8. MULTIDRUG RESISTANCE-RELATED PROTEIN (MRP) FAMILY

Initially, overexpression of MDR1 was thought to be the major mechanism by which multidrug resistance developed to cytotoxic cancer chemotherapeutic agents. However, this proved not to be the case when MRP1 was identified as another major contributor. Currently, nine members have been identified.

8.1. MRP1 (*ABCC1*)

MRP1 was originally identified from a human multidrug-resistant cancer cell line (190). Expression of MRP1 was shown to be associated with cellular resistance to chemotherapeutic agents including vinca alkaloids, taxol, methotrexate, and etoposide (191). MRP1 is expressed to the basolateral side multiple epithelial tissues in the body with highest levels being found in the testes, skeletal muscle, heart, kidney, and lung (192). MRP1 acts as a cellular efflux pump for hydrophobic compounds, which are conjugated to the anionic tripeptide glutathione (GSH), glucuronic acid, or sulfate (193). Knockout mice lacking Mrp1 are viable and fertile, but they show deficiencies in leukotriene C4 (LTC4) mediated inflammatory reactions, suggesting that secretion of LTC4 is an important physiological function of MRP1 (194). Polymorphisms in MRP1 have been associated with a connective tissue disease, pseudoxanthoma elasticum (PXE). However polymorphisms in MRP1, without the inclusion of variations in MRP6, do not appear to play a role in PXE (195–201) (www.vanderbilt.edu/kimlab). Screening of healthy Japanese subjects revealed several SNPs in MRP1 although their functional relevance was not assessed (195). MRP1 R723Q has been studied regarding its influence on duodenal mRNA levels and was not found to be associated with altered expression (202). A G671V variant, whose mutation is located within the nucleotide binding domain 1 (NBD1), was able to transport the prototypical substrates LTC4, estradiol 17β-D-glucuronide, and estrone sulfate with similarly efficiency as compared to cells wild-type for MRP1 (200). However, the R433S polymorphism resulted in a two-fold reduction in transport function and a two-fold increase in resistance to doxorubicin (199). Additional stu-

dies are required to examine the impact of MRP1 polymorphisms on clinical drug disposition as well as the reponse to chemotherapeutic agents.

8.2. MRP2 (*ABCC2*)

Mrp2 was first cloned from rat and originally designated as a canalicular multispecific organic anion carrier (cMOAT) due to its predominant localization in canalicular membranes of hepatocytes (203). MRP2 was also found in the apical membranes of enterocytes (204) and epithelial cells of proximal tubules in the kidney (205). Rats lacking functional Mrp2 expression (i.e., Wistar TR^- and Eisai rat strains) are hyperbilirubinemic as a result of their inability to excrete bilirubin conjugates into bile (203), thus, suggesting that bilirubin glucuronide conjugates are important substrates of MRP2 (206). Importantly, the absence of MRP2 in humans results in Dubin–Johnson syndrome (DJS) characterized by conjugated hyperbilirubinemia (207). In addition to glucuronide and glutathione conjugates, nonconjugated compounds such as PAH (208), vinblastine (209), HIV protease inhibitors (210), and telmisartan (211) are also MRP2 substrates. A number of genetic polymorphisms in MRP2 have been associated with DJS (212–221) (www.vanderbilt.edu/kimlab). Known genetic variations in DJS result in loss of transporter function and aberrant RNA splicing with the vast majority resulting in absence of immunochemically detectable MRP2. Some of these mutations have been extensively characterized in vitro. Protein maturation and sorting was not altered in the R1150H and the Q1382R mutations although both showed impaired transport function in vitro (214–215). In contrast I1173F and R768W mutations cause deficient maturation and impaired sorting leading to inactive transport proteins (214, 218). Although the altered disposition of many compounds has been well characterized in TR^- and Eisai rats, little is known regarding drug disposition among DJS patients.

8.3. MRP3 (*ABCC3*)

The MRP3 has been cloned by several groups (222,223) and expressed in tissues such as the liver, kidney, and intestine (222). MRP3 is localized to the basolateral membrane of hepatocytes and mediates the efflux of glucuronide and GSH conjugates, bile acid conjugates as well as drugs such as MTX back into the circulation (224,225). The upregulation of MRP3 in the liver under some cholestatic conditions and the ability of MRP3 to transport bile salts have led to the speculations that MRP3 might play a role in the enterohepatic recycling of bile salts and in the removal of toxic organic anions from the hepatocyte under cholestatic conditions (226,227). The MRP3 expression is associated with poor outcome in childhood acute lymphoblastic leukemia (228). Several SNPs have been identified however

their functional has not yet been determined (www.vanderbilt.edu/kimlab) (201,229,230).

8.4. MRP4 (*ABCC4*)

The MRP4 was discovered as a result of screening human expressed-cDNA sequence tags in an effort to discover and characterize additional mammalian ABC transporters. The MRP4 is ubiquitously expressed with highest levels in the prostate where it can be localized to the basolateral membrane of tubuloacinar cells (231). In the kidney, MRP4 is localized to the apical side of renal proximal tubules (232). The MRP4 appears to be involved in the efflux of monophosphate metabolites of nucleoside analogs such as HIV reverse transcriptase inhibitors (233) and PMEA (234). Overexpression of MRP4 is associated with resistance to AZT, 3TC, and d4T (233) as well as methotrexate (231) and the purine analogs 6-mercaptopurine and 6-thioguanine (235). MRP4 was found to transport the cylic nucleotides cAMP and cGMP (235,236), steroid sulfates (237), and prostaglandins (238). MRP4 may also play a biological role in hepatic bile acid homeostasis since the loss of the main bile acid efflux transporter, BSEP, leads to a strong compensatory upregulation in MRP4 expression (239) and that Mrp4 is upregulated in liver in experimental obstructive cholestasis (240,241). Moreover, MRP4 has been shown to transport bile acids via a GSH-cotransport mechanism (242). Recently, MRP4 (and Sult2a1) was shown to be regulated by the transcription factor, constitutive androstane receptor, as part of a coordinated mechanism to protect hepatocytes from bile acid toxicity (243). Polymorphisms in MRP4 are known to exist, however, their functional relevance remains to be established (201,244).

8.5. MRP5 (*ABCC5*)

Similar to the discovery of MRP4, MRP5 was identified after screening databases of human expressed cDNA sequence tags. MRP5 is widely expressed in multiple tissues and appears to be localized to the basolateral membrane (245). MRP5 functions as an efflux transporter of cyclic nucleotides. Similar to MRP4, MRP5 overexpression confers drug resistance towards PMEA, 6-mercaptopurine, and thioguanine (234,245). MRP5 is capable of transporting the natural cyclic nucleotides cGMP and cAMP (236,246). Furthermore, expression of MRP5 in cardiomyocytes and coronary endothelium suggests an important role in cardiovascular function (241). Several SNPs have been identified in Japanese and Caucasian populations, however their functional relevance has not been determined (201,241).

8.6. MRP6 (*ABCC6*)

The MRP6 is expressed in the liver and kidney (247) and is thought to be an amphipathic anion transporter, based on its ability to transport BQ-123, an anionic cyclopentapeptide (248,249), and glutathione conjugates such as LTC_4 (249). Mutations in MRP6 have been associated with PXE, a disorder of the connective tissue characterized by ophthalmological, dermatological, and cardiovascular abnormalities (250–264). A variety of missense, splice, insertion, and deletion mutations have been associated with PXE (www.vanderbilt.edu/kimlab).

8.7. MRP8 (*ABCC11*) and MRP9 (*ABCC12*)

Using a cluster of expressed sequence tags to screen cDNA libraries, Bera et al. (265,266) identified MRP8 and MRP9. Both transporters are expressed in various tissues including liver, lung, and kidney (267). Little is known about the existence of genetic polymorphisms in MRP8 and MRP9 or their possible functional consequences.

9. BREAST CANCER RESISTANCE PROTEIN (BCRP)

BCRP, also known as MXR or ABCP, was first cloned from mitoxantrone and anthracycline-resistant breast and colon cancer cells (268). Since BCRP has only one ABC and six putative transmembrane domains it was suggested that BCRP is a half-transporter. However, it has recently been determined that BCRP may function as a homotetramer (269). The BCRP mRNA expression analyses of normal tissues indicate highest expression in the placenta, heart, ovary, and kidney and lower levels in the liver, colon, small intestine, prostate, and brain (268). Furthermore, hematopoietic, muscular, neural, and testicular stem cells abundantly express Bcrp, thus conferring Hoechst 33342 dye transport and mitoxanthrone resistance (270,271). Interestingly, Bcrp1(−/−) mice develop protoporphyria and diet-dependent phototoxicity (272). It appears that Bcrp prevents heme or porphyrin accumulation in cells, enhancing hypoxic cell survival (273). Several polymorphisms have been discovered in BCRP (274–279) (www.vanderbilt.edu/kimlab). Two polymorphisms, C421A and C376T, have been demonstrated in a cohort of Japanese subjects (274). These polymorphisms result in reduced protein expression and might thus increase the risk of hypersensitivity to BCRP substrates in certain individuals. Non-synonymous polymorphisms were found that cause V12M, Q141K, and D620N transversions in diverse populations (276,277). However, the relative importance of polymorphisms in BCRP to disease remains to be clarified. BCRP expression appears to be a predictor of survival in platinum-based chemotherapy for advanced non-small cell lung cancer (280).

10. CONCLUSION

Drug transporters expressed in tissues such as the intestine, liver, kidney, and brain play an important role in the absorption, distribution, and excretion of many drugs in clinical use. It is now becoming increasingly clear that genetic heterogeneity in such transport processes not only contribute to the observed interindividual variation in drug disposition, but also response to drugs. Indeed, targeted and polarized expression of transporters in organs such as the brain can often limit the extent of CNS drug entry. Accordingly, genetic variations in such transporters could manifest as unexpected CNS drug effects or toxicity. Moreover, since many transporters have significant tissue expression and drug substrate overlap with a number of important drug metabolizing enzymes, such as members of the cytochrome P450 (CYP) superfamily, genetically determined differences in transporter function could indirectly affect the metabolism of shared substrate drugs by altering the extent of achieved intracellular drug concentrations in organs such as the liver and intestine. Therefore, inclusion of drug transporter pharmacogenetics to what is already known regarding metabolizing enzyme genetic polymorphisms has the potential to better predict drug efficacy and toxicity, and taken together, will likely be an important component of the goal of individualized drug selection and dosing regimens.

ACKNOWLEDGMENTS

This work was supported by U.S. Public Health Service Grants, GM31304, GM54724, and the NIH/NIGMS Pharmacogenetics Research Network and Database (U01GM61374, http://pharmgkb.org/) under Grant U01 HL65962. Marzolini was supported by a grant from the Swiss National Research Foundation (No. 81LA-69443).

REFERENCES

1. Burckhardt G, Wolff NA. Structure of renal organic anion and cation transporters. Am J Physiol Renal Physiol 2000; 278:F853–F866.
2. Higgins CF, Gottesman MM. Is the multidrug transporter a flippase? Trends Biochem Sci 1992; 17(1):18–21.
3. Suzuki H, Sugiyama Y. Role of metabolic enzymes and efflux transporters in the absorption of drugs from the small intestine. Eur J Pharm Sci 2000; 12:3–12.
4. Suzuki H, Sugiyama Y. Transport of drugs across the hepatic sinusoidal membrane: sinusoidal drug influx and efflux in the liver. Semin Liver Dis 2000; 20:251–263.
5. Inui KI, Masuda S, Saito H. Cellular and molecular aspects of drug transport in the kidney. Kidney Int 2000; 58:944–958.

6. Tamai I, Tsuji A. Transporter-mediated permeation of drugs across the blood–brain barrier. J Pharm Sci 2000; 89:1371–1388.
7. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999; 286:487–491.
8. Tamai I, Nezu JI, Uchino H, Sai Y, Oku A, Shimane M, Tsuji A. Molecular identification and characterization of novel members of the human organic anion transporter (OATP) family. Biochem Biophys Res Commun 2000; 273:251–260.
9. Jacquemin E, Hagenbuch B, Stieger B, Wolkoff AW, Meier PJ. Expression cloning of a rat liver Na(+)-independent organic anion transporter. PNAS 1994; 91:133–137.
10. Bergwerk AJ, Shi X, Ford AC, Kanai N, Jacquemin E, Burk RD, Bai S, Novikoff PM, Stieger B, Meier PJ, Schuster VL, Wolkoff AW. Immunologic distribution of an organic anion transport protein in rat liver and kidney. Am J Physiol 1996; 271:G231–G238.
11. Angeletti RH, Novikoff PM, Juvvadi SR, Fritschy JM, Meier PJ, Wolkoff AW. The choroids plexus epithelium is the site of the organic anion transport protein in the brain. PNAS 1994; 94:283–286.
12. Kullak-Ublick GA, Hagenbuch B, Stieger B, Schteingar CD, Hofmann AF, Wolkoff AW, Meier PJ. Molecular and functional characterization of an organic anion transporting polypeptide cloned from human liver. Gastroenterology 1995; 109:1274–1282.
13. Abe T, Kakyo M, Tokui T, Nakagomi R, Nishio T, Nakai D, Nomura H, Unno M, Suzuki M, Naitoh T, Matsuno S, Yawo H. Identification of a novel gene family encoding human liver-specific organic anion transporter LST-1. J Biol Chem 1999; 274:17159–17163.
14. Kullak-Ublick GA, Fisch T, Oswald M, Hagenbuch B, Meier PJ, Beuers U, Paumgartner G. Dehydroepiandrosterone sulphate (DHEAS): identification of a carrier protein in human liver and brain. FEBS Lett 1998; 424:173–176.
15. Bossuyt X, Muller M, Meier PJ. Multispecific amphipathic substrate transport by an organic anion transporter of human liver. J Hepatol 1996; 25: 733–738.
16. Kullak-Ublick GA, Ismair MG, Stieger B, Landmann L, Huber R, Pizzagalli F, Fattinger K, Meier PJ, Hagenbuch B. Organic anion-transporting polypeptide-B (OATP-B) and its functional comparison with three other OATPs of human liver. Gastroenterology 2001; 120:525–533.
17. Gao B, Hagenbuch B, Kullak-Ublick GA, Benke D, Aguzzi A, Meier PJ. Organic anion-transporting polypeptides mediate transport of opioid peptides across blood–brain barrier. J Pharmacol Exp Ther 2000; 294:73–79.
18. Cvetkovic M, Leake B, Fromm MF, Wilkinson GR, Kim RB. OATP and P-glycoprotein transporters mediate the cellular uptake and excretion of fexofenadine. Drug Metab Dispos 1999; 27:866–871.
19. Iida A, Saito S, Sekine A, Mishima C, Kondo K, Kitamura Y, Harigae S, Osawa S, Nakamura Y. Catalog of 258 single-nucleotide polymorphisms (SNPs) in genes encoding three organic anion transporters, three organic anion-transporting polypeptides, and three NADH: ubiquinone oxidoreductase flavoproteins. J Hum Genet 2001; 46:668–683.

20. Lee W, Glaeser H, Smith LH, Roberts RL, Moeckel GW, Gervasini G, Leake BF, Kim RB. Polymorphisms in human organic anion-transporting polypeptide 1A2 (OATP1A2): implications for altered drug disposition and central nervous system drug entry. J Biol Chem 2005; 280:9610–9617.
21. Kobayashi D, Nozawa D, Imai K, Nezu J, Tsuji A, Tamai I. Involvement of human organic anion transporting polypeptide OATP-B (SLC21A9) in pH-dependent transport across intestinal apical membrane. J Pharmacol Exp Ther 2003; 306:703–708.
22. Nozawa T, Imai K, Nezu J, Tsuji A, Tamai I. Functional characterization of pH-sensitive organic anion transporting polypeptide OATP-B in human. J Pharmacol Exp Ther 2004; 308:438–445.
23. Nozawa T, Nakajima M, Tamai I, Noda K, Nezu JI, Sai Y, Tsuji A, Yokoi T. Genetic polymorphisms of human organic anion transporters OATP-C (SLC21A6) and OATP-B (SLC21A9): allele frequencies in the Japanese population and functional analysis. J Pharmacol Exp Ther 2002; 302:804–813.
24. Hsiang B, Zhu Y, Wang Z, Wu Y, Sasseville V, Yang WP, Kirchgessner A. A novel human hepatic organic anion transporting polypeptide (OATP2). Identification of a liver-specific human organic anion transporting polypeptide and identification of rat and human hydroxymethylglutaryl-CoA reductase inhibitors transpoters. J Biol Chem 1999; 274:37161–37168.
25. Konig J, Cui Y, Nies AT, Keppler D. A novel human organic anion transporting polypeptide localized to the basolateral hepatocyte membrane. Am J Physiol Gastrointest Liver Physiol 2000; 278:G156–G164.
26. Tirona RG, Leake BF, Merino G, Kim RB. Polymorphisms in OATP-C. J Biol Chem 2001; 276:35669–35675.
27. Tirona RG, Leake BF, Wolkoff AW, Kim RB. Human organic anion transporting polypeptide-C (SLC21A6) is a major determinant of rifampin-mediated pregnane X receptor activation. J Pharmacol Exp Ther 2003; 304: 223–228.
28. Michalski C, Cui Y, Nies AT, Nuessler AK, Neuhaus P, Zanger UM, Klein K, Eichelbaum M, Keppler D, Konig J. A naturally occurring mutation in the SLC21A6 gene causing impaired membrane localization of the hepatocyte uptake transporter. J Biol Chem 2002; 277:43058–43063.
29. Nishizato Y, Ieiri I, Suzuki H, Kimura M, Kawabata K, Hirota T, Takane H, Irie S, Kusuhara H, Urasaki Y, Urae A, Higuchi S, Otsubo K, Sugiyama Y. Polymorphisms of OATP-C (SLC21A6) and OAT3 (SLC22A8) genes: consequences for pravastatin pharmacokinetics. Clin Pharm Ther 2003; 73:554–565.
30. Konig J, Cui Y, Nies AT, Keppler D. Localization and genomic organization of a new hepatocellular organic anion transporting polypeptide. J Biol Chem 2000; 275:23161–23168.
31. Adachi H, Suzuki T, Abe M, Asano N, Mizutamari H, Tanemoto M, Nishio T, Onogawa T, Toyohara T, Kasai S, Satoh E, Suzuki M, Tokui T, Unno M, Shimasagawa T, Matsuno S, Ito S, Abe T. Molecular characterization of human and rat organic anion transporter OATP-D. Am J Physiol Renal Physiol 2003; 286:F1188–F1197.
32. Fujiwara K, Adachi H, Nishio T, Unno M, Tokui T, Okabe M, Onogawa T, Suzuki T, Asano N, Tanemoto M, Seki M, Shiiba K, Suzuki M, Kondo Y,

Nunoki K, Shimosegawa T, Linuma K, Ito S, Matsuno S, Abe T. Identification of thyroid hormone transporters in humans: different molecules are involved in a tissue-specific manner. Endocrinology 2001; 142:2005–2012.
33. Pizzagalli F, Hagenbuch B, Stieger B, Klenk U, Folkers G, Meier PJ. Identification of a new human organic anion transporting polypeptide as a high affinity thyroxine transporter. Mol Endocrinol 2002; 16:2283–2296.
34. Mikkaichi T, Suzuki T, Onogawa T, Tanemoto M, Mizutamari H, Okada M, Chaki T, Masuda S, Tokui T, Eto N, Abe M, Satoh F, Unno M, Hishinuma T, Inui K, Ito S, Goto J, Abe T. Isolation and characterization of a digoxin transporter and its rat homologue expressed in the kidney. PNAS 2004; 101:3569–3574.
35. Suzuki T, Onogawa T, Asano N, Mizutamari H, Mikkaichi T, Tanemoto M, Abe M, Satoh F, Unno M, Nunoki K, Suzuki M, Hishinumo T, Goto J, Shimosegawa T, Matsuno S, Ito S, Abe T. Identification and characterization of novel rat and human gonad specific organic anion transporters. Mol Endocrinol 2003; 17:1203–1215.
36. Lu R, Kanai N, Bao Y, Schuster VL. Cloning, in vitro expression, and tissue distribution of a human prostaglandin transporter cDNA (hPGT). J Clin Invest 1996; 98:1142–1149.
37. Sekine T, Watanabe N, Hosoyamada M, Kanai Y, Endou H. Expression cloning and characterization of a novel multispecific organic anion transporter. J Biol Chem 1997; 272:18526–18529.
38. Hosoyamada M, Sekine T, Kanai Y, Endou H. Molecular cloning and functional expression of a multispecific organic anion transporter from human kidney. Am J Physiol 1999; 276:F122–F128.
39. Lu R, Chan BS, Schuster VL. Cloning of the human kidney PAH transporter: narrow substrate specificity and regulation by protein kinase C. Am J Physiol 1999; 276:F295–F303.
40. Race JE, Grassi SM, Williams WJ, Holtzman EJ. Molecular cloning and characterization of two novel human renal organic anion transporters (hOAT1 and hOAT3). Biochem Biophys Res Commun 1999; 255:508–514.
41. Bahn A, Ebbinghaus C, Ebbinghaus D, Ponimaskin EG, Fuzesï L, Burckhardt G, Hagos Y. Expression studies and functional characterization of renal human organic anion transporter 1 isoforms. Drug Metab Dispos 2004; 32:424–430.
42. Cihlar T, Lin DC, Pritchard JB, Fuller MD, Mendel DB, Sweet DH. The antiviral nucleotide analogs cidofovir and adefovir are novel substrates for human and rat renal organic anion transporter 1. Mol Pharmacol 1999; 56: 570–580.
43. Ho ES, Lin DC, Mendel DB, Cihlar T. Citotoxicity of antiviral nucleotides adefovir and cidofovir is induced by the expression of human renal organic anion transporter 1. J Am Soc Nephrol 2000; 11:383–393.
44. Bleasky K, Mohrenweiser HW, Pritchard JB. Identification of single nucleotide polymorphisms in the human organic anion transporter hOAT1. Drug Metab Rev 2002; 34(suppl 1):165.
45. Lee W, Ingram AD, Leake BF, Kim RB. Identification of nonsynonymous polymorphisms of human organic anion transporter 1 (OAT1) among African,

European- and Chinese-American subjects [abstr]. Drug Metab Rev 2003; 35(149):75.
46. Simonson GD, Vincent AC, Roberg KJ, Huang Y, Iwanij V. Molecular cloning and characterization of a novel liver-specific transport protein. J cell Sci 1994; 107:1065–1072.
47. Sekine T, Cha SH, Tsuda M, Apiwattanakul N, Nakajima N, Kanai Y, Endou H. Identification of multispecific organic anion transporter 2 expressed predominantly in the liver. FEBS Lett 1998; 429:179–182.
48. Enomoto A, Takeda M, Shimoda M, Narikawa S, Kobayashi Y, Yamamoto T, Sekine T, Cha SH, Niwa T, Endou H. Interaction of human organic anion transporters 2 and 4 with organic anion transport inhibitors. J Pharmacol Exp Ther 2002; 301:797–802.
49. Morita N, Kusuhara H, Sekine T, Endou H, Sugiyama Y. Functional characterization of rat organic anion tansporter 2 in LLC-PK1cells. J Pharmacol Exp Ther 2001; 298:1179–1184.
50. Cha SH, Sekine T, Fukushima JI, Kanai Y, Kobayashi Y, Goya T, Endou H. Identification and characterization of human organic anion transporter 3 expressing predominantly in the kidney. Mol Pharmacol 2001; 59:1277–1286.
51. Hasegawa M, Kusuhara H, Sugiyama D, Ito K, Ueda S, Endou H, Sugiyama Y. Functional involvement of rat organic transporter 3 (rOat3, Slc22a8) in the renal uptake of organic anions. J Pharmacol Exp Ther 2002; 300:746–753.
52. Sweet DH, Miller DS, Pritchard JB, Fujiwara Y, Beier DR, Nigam SK. Impaired organic anion transport in kidney and choroid plexus of organic anion transporter 3 (Oat3 (Slc22a8)) knockout mice. J Biol Chem 2002; 277:26934–26943.
53. Sykes D, Sweet DH, Lowes S, Nigam SK, Pritchard JB, Miller DS. Organic anion transport in choroid plexus from wild-type and organic anion transporter 3 (Slc22a8)-null mice. Am J Physiol 2004; 286:F972–F978.
54. Cha SH, Sekine T, Kusuhara H, Yu E, Kim JY, Kim DK, Sugiyama Y, Kanai Y, Endou H. Molecular cloning and characterization of multispecific organic anion transporter 4 expressed in the placenta. J Biol Chem 2000; 275:4507–4512.
55. Babu E, Takeda M, Narikawa S, Kobayashi Y, Enomoto A, Tojo A, Cha SH, Sekine T, Sakthisekaran D, Endou H. Role of human organic anion transporter 4 in the transport of ochratoxin A. Biochim Biophys Acta 2002; 1590: 64–75.
56. Ekaratanawong S, Anzai N, Jutabha P, Miyazaki H, Noshiro R, Takeda M, Kanai Y, Sophasan S, Endou H. Human organic anion transporter 4 is a renal apical organic anion/dicarboxylate exchanger in the proximal tubules. J Pharmacol Sci 2004; 94:297–304.
57. Takeda M, Khamdang S, Narikawa S, Kimura H, Kobayashi Y, Yamamoto Y, Cha SH, Sekine T, Endou H. Human organic anion transporters and human organic cation transporters mediate renal antiviral transport. J Pharmacol Exp Ther 2002; 300:918–924.
58. Takeda M, Khamdang S, Narikawa S, Kimura H, Hosoyamoda M, Cha SH, Sekine T, Endou H. Characterization of methotrexate transport and its drug interactions with human organic anion transporters. J Pharmacol Exp Ther 2002; 302:666–671.

59. Zhang L, Dresser MJ, Gray AT, Yost SC, Terashita S, Giacomini KM. Cloning and functional expression of a human liver organic cation transporter. Mol Pharmacol 1997; 51:913–921.
60. Gorboulev V, Ulzheimer JC, Akhoundova A, Ulzheimer-Teuber I, Karbach U, Quester S, Baumann C, Lang F, Busch AE, Koepsell H. Cloning and characterization of two human polyspecific organic cation transporters. DNA Cell Biol 1997; 16:871–881.
61. Van Montfoort JE, Muller M, Groothuis GM, Meijer DKF, Koepsell H, Meier PJ. Comparison of type I and type II organic cation transport by organic cation transporters and organic anion-transporting polypeptides. J Pharmacol Exp Ther 2001; 298:110–115.
62. Dresser MJ, Gray AT, Giacomini KM. Kinetic and selectivity differences between rodent, rabbit, and human organic cation transporters (OCT1). J Pharmacol Exp Ther 2000; 292:1146–1152.
63. Kerb R, Brinkmann U, Chatskaia N, Gorbunov D, Gorboulev V, Mornhinweg E, Keil A, Eichelbaum M, Koepsell H. Identification of genetic variations of the human organic cation trasporter hOCT1 and their functional consequences. Pharmacogenetics 2002; 12:591–595.
64. Saito S, Iida A, Sekine A, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y. Catalog of 238 variations among six human genes encoding solute carriers (hSLCs) in the Japanese population. J Hum Genet 2002; 47:576–584.
65. Sakata T, Anzai N, Shin HJ, Noshiro R, Hirata T, Yokoyama H, Kanai Y, Endou H. Novel single nucleotide polymorphisms of organic cation transporter 1 (SLC22A1) affecting transport functions. Biochem Biophys Res Commun 2004; 313:789–793.
66. Okuda M, Saito H, Urakami Y, Takano M, Inui K. cDNA cloning and functional expression of novel rat kidney organic cation transporter, OCT2. Biochem Biophys Res Commun 1996; 224:500–507.
67. Motohashi H, Sakurai Y, Saito H, Masuda S, Urakami Y, Goto M, Fukatsu A, Ogawa O, Inui K. Gene expression levels and immunolocalization of organic ion transporters in the human kidney. J Am Soc Nephrol 2002; 13:866–874.
68. Busch AE, Karbach U, Miska D, Gorboulev V, Akhoundova A, Volk C, Arndt P, Ulzheimer JC, Sonders MS, Baumann C, Waldegger S, Lang F, Koepsell H. Human neurons express the polyspecific cation transporter hOCT2, which translocates monoamine neurotransmitters, amantadine, and memantine. Mol Pharmacol 1998; 54:342–352.
69. Leabman MK, Huang CC, Kawamoto M, Johns SJ, Stryke D, Ferrin TE, De Young J, Taylor T, Clark AG, Herskowitz I, Giacomini KM. Pharmacogenetics of membrane transporters investigators. Polymorphisms in a human kidney xenobiotic transporter, OCT2, exhibit altered function. Pharmacogenetics 2002; 12:395–405.
70. Grundemann D, Schechinger B, Rappold GA, Schomig E. Molecular identification of the corticosterone-sensitive extraneuronal catecholamine transporter. Nat Neurosci 1998; 1:349–351.
71. Wu X, Huang W, Ganapathy ME, Wang H, Kekuda R, Conway SJ, Leibach FH, Ganapathy V. Structure, function, and regional distribution of

the organic cation transporter OCT3 in the kidney. Am J Physiol Renal Physiol 2000; 279:F449–F458.
72. Wieland A, Hayer-Zillgen M, Bonisch H, Bruss M. Analysis of the gene structure of the human (SLC22A3) and murine (Slc22a3) extraneuronal monoamine transporter. J Neural Transm 2000; 107:1149–1157.
73. Lazar A, Grundemann D, Berkels R, Taubert D, Zimmermann T, Schomig E. Genetic variations of the extraneuronal monoamine transporter EMT (SLC22A3). J Hum Genet 2003; 48:226–230.
74. Tamai I, Yabuuchi H, Nezu J, Sai Y, Oku A, Shimane M, Tsuji A. Cloning and characterization of a novel human pH-dependent organic cation transporter, OCTN1. FEBS Lett 1997; 419:107–111.
75. Yabuuchi H, Tamai I, Nezu J, Sakamoto K, Oku A, Shimane M, Sai Y, Tsuji A. Novel membrane transporter OCTN1 mediates multispecific, bidirectional, and pH-dependent transport of organic cations. J Pharmacol Exp Ther 1999; 289: 768–773.
76. Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T, Suzuki M, Nagasaki M, Ohtsuki M, Ono M, Furukawa H, Nagashima M, Yoshino S, Mabuchi A, Sekine A, Saito S, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet 2003; 35:341–348.
77. Tamai I, Ohashi R, Nezu J, Yabuuchi H, Oku A, Shimane M, Sai Y, Tsuji A. Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2. J Biol Chem 1998; 273:20378–20382.
78. Wu X, Prasad PD, Leibach FH, Ganapathy V. cDNA sequence transport function genomic organization of human OCTN2 a new member of the organic cation transporter family. Biochem Biophys Res Commun 1998; 246: 589–595.
79. Tamai I, China K, Sai Y, Kobayashi D, Nezu J, Kawahara E, Tsuji A. Na^+-coupled transport of L-carnitine via high-affinity carnitine transporter OCTN2 and its subcellular localization in kidney. Biochim Biophys Acta 2001; 1512:273–284.
80. Wang Y, Korman SH, Ye J, Gargus JJ, Gutman A, Taroni F, Garavaglia B, Longo N. Phenotype and genotype variation in primary carnitine deficiency. Genet Med 2001; 3:387–392.
81. Nezu JI, Tamai I, Oku A, Ohashi R, Yabuuchi H, Hashimoto N, Nikaido H, Sai Y, Koizumi A, Shoji Y, Takada G, Matsuishi T, Yoshino M, Kato H, Ohura T, Tsujimoto G, Hayakawa JI, Shimane M, Tsuji A. Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter. Nat Genet 1999; 21:91–94.
82. Wang Y, Kelly MA, Cowan TM, Longo N. A missense mutation in the OCTN2 gene associated with residual carnitine transport activity. Hum Mutat 2000; 15:238–245.
83. Burwinkel B, Kreuder J, Schweitzer S, Vorgerd M, Gempel K, Gerbitz KD, Kilimann MW. Carnitine transporter OCTN2 mutations in systemic primary carnitine deficiency: a novel Arg169Gln mutation and a recurrent Arg282ter

mutation associated with an unconventional splicing abnormality. Biochem Biophys Res Commun 1999; 261:484–487.
84. Tang NLS, Ganapathy V, Wu X, Hui J, Seth P, Yuen PMP, Fok TF, Hjelm NM. Mutations of OCTN2, an organic cation/carnitine transporter, lead to deficient cellular carnitine uptake in primary carnitine deficiency. Hum Mol Genet 1999; 8: 655–660.
85. Vaz FM, Scholte HR, Ruiter J, Hussaarts-Odijk LM, Rodrigues Pereira R, Schweitzer S, de Klerk JB, Waterham HR, Wanders RJA. Identification of two novel mutations in OCTN2 of three patients with systemic carnitine deficiency. Hum Genet 1999; 105:157–161.
86. Wang Y, Taroni F, Garavaglia B, Longo N. Functional analysis of mutations in the OCTN2 transporter causing primary carnitine deficiency: lack of genotype–phenotype correlation. Hum Mutat 2000; 16:401–407.
87. Koizumi A, Nozaki J, Ohura T, Kayo T, Wada Y, Nezu J, Ohashi R, Tamai I, Shoji Y, Takada G, Kibira S, Matsuishi T, Tsuji A. Genetic epidemiology of the carnitine transporter OCTN2 gene in Japanese population and phenotypic characterization in Japanese pedigrees with primary systemic carnitine deficiency. Hum Mol Genet 1998; 8:2247–2254.
88. Cederbaum SD, Koo-McCoy S, Tein I, Hsu BYL, Ganguly A, Vilain E, Dipple K, Cvitanovic-Sojat L, Stanley C. Carnitine membrane transporter deficiency: a long-term follow up and OCTN2 mutation in the first documented case of primary carnitine deficiency. Mol Genet Metab 2002; 77: 195–201.
89. Wang Y, Ye J, Ganapathy V, Longo N. Mutations in the organic cation/carnitine transporter OCTN2 in primary carnitine deficiency. PNAS 1999; 96: 2356–2360.
90. Mayapetek E, Nezu J, Tamai I, Ooku A, Katsura M, Shimane M, Tsuji A. Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a patient with primary systemic carnitine deficiency. Hum Mutat 2000; 15:118.
91. Seth P, Wu X, Huang W, Leibach FH, Ganapathy V. Mutations in novel organic cation transporter (OCTN2), an organic cation/carnitine transporter, with differential effects on the organization transport function and the carnitine transport function. J Biol Chem 1999; 274:33388–33392.
92. Liang R, Fei YJ, Prasad PD, Ramamoorthy S, Han H, Yang-Geng TL, Hediger MA, Ganapathy V, Leibach FH. Human intestinal H^+/peptide cotransporter. Cloning, functional expression, and chromosomal localization. J Biol Chem 1995; 270:6456–6463.
93. Ogihara H, Saito H, Shin BC, Terado T, Takenoshita S, Nagamachi Y, Inui K, Takata K. Immuno-localization of H^+/peptide cotransporter in rat digestive tract. Biochem Biophys Res Commun 1996; 220:848–852.
94. Shen H, Smith DE, Yang T, Huang YG, Schnermann JB, Brosius FC. Localization of PEPT1 and PEPT2 proton-coupled oligopeptide transporter mRNA and protein in rat kidney. Am J Physiol 1999; 276:F658–F665.
95. Ganapathy ME, Brandsch M, Prasad PD, Ganapathy V, Leibach FH. Differential recognition of beta-lactam antibiotics by intestinal and renal peptide transporters, PEPT1 and PEPT2. J Biol Chem 1995; 270:25672–25677.

96. Sugawara M, Huang W, Fei YJ, Leibach FH, Ganapathy V, Ganapathy ME. Transport of valganciclovir, a ganciclovir prodrug, via peptide transporters PEPT1 and PEPT2. J Pharm Sci 2000; 89:781–789.
97. Zhu T, Chen XZ, Steel A, Hediger MA, Smith DE. Differential recognition of ACE inhibitors in Xenopus laevis oocytes rat PEPT1 and PEPT2. Pharm Res 2000; 17:526–532.
98. Liu W, Liang R, Ramamoorthy S, Fei YJ, Ganapathy ME, Hediger MA, Ganapathy V, Leibach FH. Molecular cloning of PEPT2, a new member of the H^+/peptide cotransporter family, from human kidney. Biochim Biophys Acta 1995; 1235:461–466.
99. Rubio-Aliaga I, Frey I, Boll M, Groneberg DA, Eichinger HM, Balling R, Daniel H. Targeted disruption of the peptide transporter Pept2 gene in mice defines its physiological role in the kidney. Mol Cell Biol 2003; 23:3247–3252.
100. Shu C, Shen H, Teuscher NS, Lorenzi PJ, Keep RF, Smith DE. Role of PEPT2 in peptide/mimetic trafficking at the blood–cerebrospinal fluid barrier: studies in rat choroid plexus epithelial cells in primary culture. J Pharmacol Exp Ther 2002; 301:820–829.
101. Shen H, Smith DE, Keep RF, Xiang J, Brosius FC. Targeted disruption of the PEPT2 gene markedly reduces dipeptide uptake in choroid plexus. J Biol Chem 2003; 278:4786–4791.
102. Ocheltree SM, Shen H, Hu Y, Xiang J, Keep RF, Smith DE. Mechanisms of cefadroxil uptake in the choroid plexus: studies in wild-type and PEPT2 knockout mice. J Pharmacol Exp Ther 2004; 308:462–467.
103. Juliano RL, Ling V. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta 1976; 455: 152–162.
104. Ueda K, Cardarelli C, Gottesman MM, Pastan I. Expression of a full-length cDNA for the human *MDR1* gene confers resistance to colchicines, doxorubicin, and vinblastine. PNAS 1987; 84:3004–3008.
105. Thiebaut F, Tsuruo T, Hamada H, Gottesman MM, Pastan I, Willingham MC. Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. PNAS 1987; 84:7735–7738.
106. Cordon-Cardo C, O'Brien JP, Casals D, Rittman-Grauer L, Biedler JL, Melamed MR, Bertino JR. Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood–brain barrier sites. PNAS 1989; 86: 695–698.
107. Klimecki WT, Futscher BW, Grogan TM, Dalton WS. P-glycoprotein expression and function in circulating blood cells from normal volunteers. Blood 1994; 83:2451–2458.
108. Schinkel A, Smit J, van Tellingen O, Beijnen J, Wagenaar E, van Deemter L, Mol CA, van der Valk MA, Robanus-Maandag EC, Riele PJ, Berns AJM, Borst P. Disruption of the mouse mdr1a P-glycoprotein gene leads to a deficiency in the blood–brain barrier and to increased sensitivity to drugs. Cell 1994; 77:491–502.
109. Kwei GY, Alvaro RF, Chen Q, Jenkins HJ, Hop CEAC, Keohane CA, Ly VT, Strauss JR, Wang RW, Pippert TR, Umbenhauer DR. Disposition of

ivermectin and cyclosporin A in CF-1 mice deficient in mdr1a P-glycoprotein. Drug Metab Dispos 1999; 27:581–587.
110. Meerum Terwgot J, Malingre M, Beijnen J, ten Bokkel Huinink W, Rosing H, Koopman F, van Telligen O, Swart M, Schellens JH. Co-administration of cyclosporin A enables oral therapy with paclitaxel. Clin Cancer Res 1999; 5:3379–3384.
111. Kawahara M, Sakata A, Miyashita T, Tamai I, Tsuji A. Physiologically based pharmacokinetics of digoxin in mdr1a knockout mice. J Pharm Sci 1999; 88:1281–1287.
112. Marzolini C, Paus E, Buclin T, Kim RB. Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance. Clin Pharm Ther 2004; 75:13–33.
113. Kioka N, Tsubota J, Kakehi Y, Komano T, Gottesman MM, Pastan I, Ueda K. P-glycoprotein gene (MDR1) cDNA from human adrenal: normal P-glycoprotein carries Gly^{185} with an altered pattern of multidrug resistance. Biochem Biophys Res Commun 1989; 162:224–231.
114. Hoffmeyer S, Burk O, von Richter O, Arnold H, Brockmoller J, Johne A, Cascorbi I, Gerloff T, Roots I, Eichelbaum M, Brinkmann U. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. Proc Natl Acad Sci 2000; 97:3473–3478.
115. Sai K, Kaniwa N, Itoda M, Saito Y, Hasegawa R, Komamura K, Ueno K, Kamakura S, Kitaaze M, Shirao K, Minami H, Ohtsu A, Yoshida T, Saijo N, Kitamura Y, Kamatani N, Ozawa S, Sawada JI. Haplotype analysis of ABCB1/MDR1 blocks in a Japanese population reveals genotype-dependent renal clearance of irinotecan. Pharmacogenetics 2003; 13:741–757.
116. Kim RB, Leake B, Choo E, Dresser G, Kubba S, Schwarz U, Taylor A, Xie HG, McKinsey J, Zhou S, Lan LB, Schuetz JD, Schuetz EG, Wilkinson GR. Identification of functionally variant MDR1 alleles among European-Americans and African-Americans. Clin Pharm Ther 2001; 70:189–199.
117. Tanabe M, Ieiri I, Nagata N, Inoue K, Ito S, Kanamori Y, Takahashi M, Kurata Y, Kigawa J, Higuchi S, Terakawa N, Otsubo K. Expression of P-glycoprotein in human placenta: relation to genetic polymorphism of the multidrug resistance (MDR)-1 gene. J Pharmacol Exp Ther 2001; 297: 1137–1143.
118. Mickley LA, Lee JS, Weng Z, Zhan Z, Alvarez M, Wilson W, Bates SE, Fojo T. Genetic polymorphism in MDR-1: a tool for examining allelic expression in normal cells, unselected and drug-selected cell lines, and human tumors. Blood 1998; 91(5):1749–1756.
119. Cascorbi I, Gerloff T, Johne A, Meisel C, Hoffmeyer S, Schwab M, Schaeffeler E, Eichelbaum M, Brinkmann U, Roots I. Frequency of single nucleotide polymorphisms in the P-glycoprotein drug transporter MDR1 gene in white subjects. Clin Pharm Ther 2001; 69:169–174.
120. Johne A, Kopke K, Gerloff T, Mai I, Rietbrock S, Meisel C, Hoffmeyer S, Kerb R, Fromm MF, Brinkmann U, Eichelbaum M, Brockmoller J, Cascorbi I. Modulation of steady-state kinetics of digoxin by haplotypes of the P-glycoprotein MDR1 gene. Clin Pharmacol Ther 2002; 72:584–594.

121. Verstuyft C, Schwab M, Schaeffeler E, Kerb R, Brinkmann U, Jaillon P, Funck-Brentano C. Digoxin pharmacokinetics and MDR1 genetic polymorphisms. Eur J Clin Pharmacol 2003; 58:809–812.
122. Gerloff T, Schaefer M, Johne A, Oselin K, Meisel C, Cascorbi I, Roots I. MDR1 genotypes do not influence the absorption of a single oral dose of 1 mg digoxin in healthy white males. Br J Clin Pharmacol 2002; 54:610–616.
123. Sakaeda T, Nakamura T, Horinouchi M, Kakumoto M, Ohmoto N, Sakai T, Morita Y, Tamura T, Aoyama N, Hirai M, Kasuga M, Okumura K. MDR1 genotype-related pharmacokinetics of digoxin after single oral administration in healthy Japanese subjects. Pharm Res 2001; 18:1400–1404.
124. Horinouchi M, Sakaeda T, Nakamura T, Morita Y, Tamura T, Aoyama N, Kasuga M, Okumura K. Significant genetic linkage of MDR1 polymorphisms at positions 3435 and 2677: functional relevance to pharmacokinetics of digoxin. Pharm Res 2002; 19:1581–1585.
125. Kurata Y, Ieiri I, Kimura M, Morita T, Irie S, Urae A, Ohdo S, Ohtani H, Sawada Y, Higuchi S, Otsubo K. Role of human MDR1 gene polymorphism in bioavailability and interaction of digoxin, a substrate of P-glycoprotein. Clin Pharmacol Ther 2002; 72:209–219.
126. Drescher S, Schaeffeler E, Hitzl M, Hofmann U, Schwab M, Brinkmann U, Eichelbaum M, Fromm MF. MDR1 gene polymorphisms and disposition of the P-glycoprotein substrate fexofenadine. Br J Clin Pharmacol 2002; 53:526–534.
127. Fellay J, Marzolini C, Meaden E, Back D, Buclin T, Chave J, Decosterd LA, Furrer H, Opravil M, Pantaleo G, Retelska D, Ruiz L, Schinkel AH, Vernazza P, Eap CB, Telenti A. Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. Lancet 2002; 359:30–36.
128. Von Ahsen N, Richter M, Grupp C, Ringe B, Oellerich M, Armstrong VW. No influence of the MDR-1 C3435T polymorphism or a CYP3A4 promoter polymorphism (CYP3A4-V allele) on dose-adjusted cyclosporin A trough concentrations or rejection incidence in stable renal transplant recipients. Clin Chem 2001; 47:1048–1052.
129. Min DI, Ellingrod VL. C3435T mutation in exon 26 of the human MDR1 gene and cyclosporine pharmacokinetics in healthy subjects. Ther Drug Monit 2002; 24:400–404.
130. Yates C, Zhang W, Song P, Li S, Gaber A, Kotb M, Honaker MR, Alloway RR, Meibohm B. The effect of CYP3A5 and MDR1 polymorphic expression on cyclosporine oral disposition in renal transplant patients. J Clin Pharmacol 2003; 43:555–564.
131. Mai I, Stormer E, Goldhammer M, Johne A, Kruger H, Budde K, Roots I. MDR1 haplotypes do not affect the steady-state pharmacokinetics of cyclosporine in renal transplant patients. J Clin Pharm 2003; 43:1101–1107.
132. Haufroid V, Mourad M, Van Kerckhove V, Wawrzyniak J, De Meyer M, Chaib Eddour D, Malaise J, Lison D, Squifflet JP, Wallemacq P. The effect of CYP3A5 and MDR1 (ABCB1) polymorphisms on cyclosporine and tacrolimus dose requirements and through blood levels in stable renal transplant patients. Pharmacogenetics 2004; 14:147–154.

133. Chowbay B, Cumaraswamy S, Cheung YB, Zhou Q, Lee EJ. Genetic polymorphisms in MDR1 and CYP3A4 genes in Asians and the influence of MDR1 haplotypes on cyclosporin disposition in heart transplant recipients. Pharmacogenetics 2003; 13:89–95.
134. Balram C, Sharma A, Sivathasan C, Lee EJD. Frequency of C3435T single nucleotide MDR1 genetic polymorphism in an Asian population: phenotypic–genotypic correlates. Br J Clin Pharmacol 2003; 56:78–83.
135. Kuzuya T, Kobayashi T, Moriyama N, Nagasaka T, Yokoyama I, Uchida K, Nakao A, Nabeshima T. Amlodipine, but not MDR1 polymorphisms, alters the pharmacokinetics of cyclosporine A in Japanese kidney transplant recipients. Transplantation 2003; 76:865–868.
136. Macphee I, Fredericks S, Tai T, Syrris P, Carter N, Johnston A, Goldberg L, Holt DW. Tacrolimus pharmacogenetics: polymorphisms associated with expression of cytochrome P4503A5 and P-glycoprotein correlate with dose requirement. Transplantation 2002; 74:1486–1489.
137. Zheng H, Webber S, Zeevi A, Schuetz E, Zhang J, Bowman P, Boyle G, Law Y, Miller S, Lamba J, Burckart GJ. Tacrolimus dosing in pediatric heart transplant patients is related to CYP3A5 and MDR1 gene polymorphisms. Am J Transplant 2003; 3:477–483.
138. Anglicheau D, Verstuyft C, Laurent-Puig P, Becquemont L, Schlageter MH, Cassinat B, Beaune P, Legendre P, Thervet E. Association of the multidrug resistance-1 gene single-nucleotide polymorphisms with the tacrolimus dose requirements in renal transplant recipients. J Am Soc Nephrol 2003; 14:1889–1896.
139. Goto M, Masuda S, Saito H, Uemoto S, Kiuchi T, Tanaka K, Inui K. C3435T polymorphism in the MDR1 gene affects the enterocyte expression level of CYP3A4 rather than Pgp in recipients of living-donor liver transplantation. Pharmacogenetics 2002; 12:451–457.
140. Yamauchi A, Ieiri I, Kataoka Y, Tanabe M, Nishizaki T, Oishi R, Higuchi S, Otsubo K, Sugimachi K. Neurotoxicity induced by tacrolimus after liver transplantation: relation to genetic polymorphisms of the ABCB1 (MDR1) gene. Transplantation 2002; 74:571–578.
141. Roberts RL, Joyce PR, Mulder RT, Begg EJ, Kennedy MA. A common P-glycoprotein polymorphism is associated with nortriptyline-induced postural hypotension in patients treated for major depression. Pharmacogenomics J 2002; 2:191–196.
142. Siegmund W, Ludwig K, Giessmann T, Dazert P, Schroeder E, Sperker B, Warzok R, Kroemer HK, Cascorbi I. The effects of the human MDR1 genotype on the expression of duodenal P-glycoprotein and disposition of the probe drug talinolol. Clin Pharm Ther 2002; 72:572–583.
143. Pauli-Magnus C, Feiner J, Brett C, Lin E, Kroetz DL. No effect of MDR1 C3435T variant on loperamide disposition and central nervous system effects. Clin Pharm Ther 2003; 74:487–498.
144. Skarke C, Jarrar M, Schmidt H, Kauert G, Langer M, Geisslinger G, Lotsch J. Effects of ABCB1 (multidrug resistance transporter) gene mutations on disposition and central nervous effects of loperamide in healthy volunteers. Pharmacogenetics 2003; 13:651–660.

145. Putnam W, Desai DG, Hung Y, Woo JM, Benet LZ. The effect of induction conditions and MDR1 genotypes on dicloxacillin pharmacokinetics [abstr]. Clin Pharm Ther 2003; 73:57.
146. Goh BC, Lee SC, Wang LZ, Fan L, Guo JY, Lamba J, Schuetz E, Lim R, Lim HL, Ong AB, Lee HS. Explaining interindividual variability of docetaxel pharmacokinetics and pharmacodynamics in Asians through phenotyping and genotyping strategies. J Clin Oncol 2002; 20:3683–3690.
147. Sai K, Kaniwa N, Itoda M, Saito Y, Hosegawa M, Kamamura K, Ueno K, Kamakura S, Kitakaze M, Shirao K, Minami H, Ohtsu A, Yoshida T, Saijo N, Kitamura Y, Komatoni N, Ozawa S, Sawada JI. Haplotype analysis of ABCB1/MDR1 blocks in a Japanese population reveals genotype-dependent renal clearance of irinotecan. Pharmacogenetics 2003; 13:741–757.
148. Kerb R, Aynacioglu AS, Brockmoller J, Schlagenhaufer R, Bauer S, Szekeres T, Hamwi A, Fritzer-Szekeres M, Baumgartner C, Ongen HZ, Guzelbey P, Roots I, Brinkmann U. The predictive value of MDR1, CYP2C9, and CYP2C19 polymorphisms for phenytoin plasma levels. Pharmacogenomics J 2001; 1:204–210.
149. Nakamura T, Sakaeda T, Horinouchi M, Tamura T, Aoyama N, Shirakawa T, Matsuo M, Kasuga M, Okumura K. Effect of the mutation (C3435T) at exon 26 of the MDR1 gene on expression level of MDR1 messenger ribonucleic acid in duodenal enterocytes of healthy Japanese subjects. Clin Pharmacol Ther 2002; 71(4):297–303.
150. Morita N, Yasumori T, Nakayama K. Human polymorphism: G2677T/A and C3435T have no effect on MDR1 transport activities. Biochem Pharmacol 2003; 65:1843–1852.
151. Kimchi-Sarfaty C, Gribar JJ, Gottesman MM. Functional characterization of coding polymorphisms in the human MDR1 gene using a vaccinia virus expression system. Mol Pharmacol 2002; 62:1–6.
152. Siegsmund M, Brinkmann U, Schaffeler E, Weirich G, Schwab M, Eichelbaum M, Fritz P, Burk O, Decker J, Alken P, Rothenpieler U, Kerb R, Hoffmeyer S, Brauch H. Association of the P-glycoprotein transporter MDR1 (C3435T) polymorphism with the susceptibility to renal epithelial tumors. J Am Soc Nephrol 2002; 13:1847–1854.
153. Atanasova S, von Ahsen N, Dimitrov T, Armstrong V, Oellerich M, Toncheva D. MDR1 haplotypes modify BEN disease risk: a study in Bulgarian patients with balkan endemic nephropathy compared to healthy controls. Nephron Exp Nephrol 2004; 96:e7–e13.
154. Illmer T, Schuler U, Thiede C, Schwarz U, Kim R, Gotthard S, Freund D, Schaekel U, Ehninger G, Schaich M. MDR1 gene polymorphisms affect therapy outcome in acute myeloid leukemia patients. Cancer Res 2002; 62:4955–4962.
155. Efferth T, Sauerbrey A, Steinbach D, Gebhart E, Drexler HG, Chitambar CR, Becker CM, Zintl F, Humeny A. Analysis of single nucleotide polymorphism C3435T of the multidrug resistance gene MDR1 in acute lymphoblastic leukemia. Int J Oncol 2003; 23:509–517.
156. Kafka A, Sauer G, Jaeger C, Grundmann R, Kreienberg R, Zeillinger R, Deissler H. Polymorphism C3435T of the MDR-1 gene predicts response to preoperative chemotherapy in locally advanced breast cancer. Int J Oncol 2003; 22:1117–1121.

157. Schwab M, Schaeffeler E, Marx C, Fromm MF, Kaskas B, Metzler J, Stange E, Herfarth H, Schoelmerich J, Gregor M, Walker S, Cascorbi I, Roots I, Brinkmann U, Zanger UM, Eichelbaum M. Association between the C3435T MDR1 gene polymorphism and susceptibility for ulcerative colitis. Gastroenterology 2003; 124:26–33.
158. Zheng H, Webber S, Zeevi A, Schuetz E, Zhang J, Lamba J, Bowman P, Burckart GJ. The MDR1 polymorphisms at exons 21 and 26 predict steroid weaning in pediatric heart transplant patients. Hum Immunol 2002; 63: 765–770.
159. Asano T, Takahashi KA, Fujioka M, Inoue S, Okamoto M, Sugioka N, Nishino H, Tanaka T, Hirota Y, Kubo T. ABCB1 C3435T and G2677T/A polymorphism decreased the risk for steroid-induced osteonecrosis of the femoral head after kidney transplantation. Pharmacogenetics 2003; 13: 675–682.
160. Furuno T, Landi MT, Ceroni M, Caporaso N, Bernucci I, Nappi G, Martignoni E, Schaeffeler E, Eichelbaum M, Schwab M, Zanger UM. Expression polymorphism of the blood–brain barrier component P-glycoprotein (MDR1) in relation to Parkinson's disease. Pharmacogenetics 2002; 12: 529–534.
161. Drozdzik M, Bialecka M, Mysliwiec K, Honczarenko K, Stankiewicz J, Sych Z. Polymorphism in the P-glycoprotein drug transporter MDR1 gene: a possible link between environmental and genetic factors in Parkinson's disease. Pharmacogenetics 2003; 13:259–263.
162. Siddiqui A, Kerb R, Weale ME, Brinkmann U, Smith A, Goldstein DB, Wood NW, Sisodiya SM. Association of multidrug resistance in epilepsy with a polymorphism in the drug-transporter gene ABCB1. N Engl J Med 2003; 348:1442–1448.
163. Nasi M, Borghi V, Pinti M, Bellodi C, Lugli E, Maffei S, Troiano L, Richeldi L, Mussini C, Esposito R, Cossarizza A. MDR1 C3435T genetic polymorphism does not influence the response to antiretroviral therapy in drug-naive HIV-positive patients. AIDS 2003; 17:1696–1698.
164. Ifergan I, Bernard NF, Bruneau J, Alary M, Tsoukas CM, Roger M. Allele frequency of three functionally active polymorphisms of the MDR-1 gene in high-risk HIV-negative and HIV-positive Caucasians. AIDS 2002; 16: 2340–2342.
165. Bleiber G, May M, Suarez C, Martinez R, Marzolini C, Egger M, Telenti A. MDR1 genetic polymorphism does not modify cell permissiveness to HIV-1 or disease progression before treatment. J Infect Dis 2004; 189: 583–586.
166. Brumme Z, Dong WWY, Chan KJ, Hogg RS, Montaner JSG, O'Shaughnessy MV, Harrigan PR. Influence of polymorphisms within the CX3CR1 and MDR-1 genes on initial antiretroviral therapy response. AIDS 2003; 17:201–208.
167. Haas DW, Wu H, Li H, Bosch RJ, Lederman MM, Kuritzkes D, Landay A, Connick E, Benson C, Wilkinson GR, Kessler H, Kim RB. MDR1 gene polymorphisms and phase 1 viral decay during HIV-1 infection. J AIDS 2003; 34:295–298.

168. De Luca V, Mundo E, Trakalo J, Wong GWH, Kennedy JL. Investigation of polymorphism in the MDR1 gene and antidepressant-induced mania. Pharmacogenomics J 2003; 3:297–299.
169. Panwala CM, Jones JC, Viney JL. A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, mdr1a, spontaneously develop colitis. J Immunol 1998; 161:5733–5744.
170. Shen LX, Basilion JP, Stanton VPJR. Single-nucleotide polymorphisms can cause different structural folds of mRNA. PNAS 1999; 96:7871–7876.
171. Frittitta L, Ercolino T, Bozzali M, Argiolas A, Graci S, Santagati MG, Spampinato D, Di Paola R, Cisternino C, Tassi V, Vigneri R, Pizzuli A, Trischitta V. A cluster of three single nucleotide polymorphisms in the 3′-untranslated region of human glycoprotein PC-1 gene stabilizes PC-1 mRNA and is associated with increased PC-1 protein content and insulin resistance-related abnormalities. Diabetes 2001; 50:1952–1955.
172. Schuetz EG, Umbenhauer DR, Yasuda K, Brimer C, Nguyen L, Relling MV, Schuetz JD, Schinkel AH. Altered expression of hepatic cytochromes P-450 in mice deficient in one or more mdr1 genes. Mol Pharmacol 2000; 57:188–197.
173. Van Helvoort A, Smith AJ, Sprong H, Fritzsche I, Schinkel AH, Borst P, van Meer G. MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine. Cell 1996; 87:507–517.
174. Smit JJ, Schinkel AH, Oude Elferink RPJ, Groen AK, Wagenaar E, van Deemter L, Mol CA, Ottenhoff R, van der Lugt NMT, van Roon MA, van der Valk MA, Offerhaus GJA, Berns AJM, Borst P. Homozygous disrupting of the murine mdr2 P-glycoprotein gene leads to a complete absence of phospholipids from bile and to liver disease. Cell 1993; 75:451–462.
175. De Vree JM, Jacquemin E, Sturm E, Cresteil D, Bosma PJ, Aten J, Deleuze JF, Desrochers M, Burdelski M, Bernard O, Oude Elferink RPJ, Hadchouel M. Mutations in the MDR3 gene cause progressive familial intrahepatic cholestasis. PNAS 1998; 95:282–287.
176. Chen HL, Chang PS, Hsu HC, Lee JH, Ni YH, Hsu HY, Tseng YM, Chang MH. Progressive familial intrahepatic cholestasis with high gamma-glutamyltranspeptidase levels in Taiwanese infants: role of MDR3 gene defect? Pediatr Res 2001; 50:50–55.
177. Jacquemin E, de Vree JM, Cresteil D, Sokal EM, Sturm E, Dumont M, Scheffer GL, Paul M, Burdelski M, Bosma PJ, Bernard O, Hadchouel M, Oude Elferink RPJ. The wide spectrum of multidrug resistance 3 deficiency: from neonatal cholestasis to cirrhosis of adulthood. Gastroenterology 2001; 120:1448–1458.
178. Rosmorduc O, Hermelin B, Boelle PY, Parc R, Taboury J, Poupon R. ABCB4 gene mutation-associated cholelithiasis in adults. Gastroenterology 2003; 125:452–459.
179. Rosmorduc O, Hermelin B, Poupon R. MDR3 gene defect in adults with symptomatic intrahepatic and gallbladder cholesterol cholelithiasis. Gastroenterology 2001; 120:1459–1467.
180. Pauli-Magnus C, Lang T, Meier Y, Zodan-Marin T, Jung D, Breymann C, Zimmermann R, Kenngott S, Beuers U, Reichel C, Kerb R, Penger A,

Meier PJ, Kullak-Ublick G. Sequence analysis of bile salt export pump (ABCB11) and multidrug resistance P-glycoprotein 3 (ABCB4, MDR3) in patients with intrahepatic cholestasis of pregnancy. Pharmacogenetics 2004; 14:91–102.
181. Lucena JP, Herrero I, Quiroga J, Sangro B, Garcia-Foncillas J, Zabalegui N, Sola J, Herraiz M, Medina JF, Prieto J. A multidrug resistance 3 gene mutation causing cholelithiasis, cholestasis of pregnancy, and adulthood biliary cirrhosis. Gastroenterology 2003; 124:1037–1042.
182. Dixon PH, Weerasekera N, Linton KJ, Donaldson O, Chambers J, Egginton E, Weaver J, Nelson-Piercy C, de Swiet M, Warnes G, Elias E, Higgins CF, Johnston DG, McCarthy MI, Williamson C. Heterozygous MDR3 missense mutation associated with intrahepatic cholestasis of pregnancy: evidence for a defect in protein trafficking. Hum Mol Genet 2000; 9:1209–1217.
183. Jacquemin E, Cresteil D, Manouvrier A, Boute O, Hadchoel M. Heterozygous non-sense mutation of the MDR3 gene in familial intrahepatic cholestasis of pregnancy. Lancet 1999; 353:210–211.
184. Pauli-Magnus C, Kerb R, Fattinger K, Lang T, Anwald B, Kullak-Ublick GA, Beuers U, Meier PJ. BSEP and MDR3 haplotype structure in healthy Caucasians, primary biliary cirrhosis and primary sclerosing cholangitis. Hepatology 2004; 39:779–791.
185. Childs S, Yeh RL, Georges E, Ling V. Identification of a sister gene to P-glycoprotein. Cancer Res 1995; 55:2029–2034.
186. Jansen PLM, Strautnieks SS, Jacquemin E, Hadchouel M, Sokal EM, Hooiveld GJEJ, Koning JH, de Jager-Krikken A, Kuipers F, Stellaard F, Bijleveld CMA, Gouw A, van Goor H, Thompson RJ, Muller M. Hepatocanalicular bile salt export pump deficiency in patients with progressive familial intrahepatic cholestasis. Gastroenterology 1999; 117:1370–1379.
187. Strautnieks SS, Bull LN, Knisely AS, Kocoshis SA, Dahl N, Arnell H, Sokal E, Dahan K, Childs S, Ling V, Tanner MS, Agalwalla AF, Nemeth A, Pawlowska J, Baker A, Mieli-Vergani G, Freimer NB, Gardiner RM, Thompson RJ. A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nat Genet 1998; 20:233–238.
188. Goto K, Sugiyama K, Sugiura T, Ando T, Mizutani F, Terabe K, Ban K, Togari H. Bile salt export pump gene mutations in two Japanese patients with progressive familial intrahepatic cholestasis. J Pediatr Gastroenterol Nutr 2003; 36:647–650.
189. Wang L, Soroka CJ, Boyer Jl. The role of bile salt export pump mutations in progressive familial intrahepatic cholestasis type II. J Clin Invest 2002; 110:965–972.
190. Cole SPC, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, Almquist KC, Stewart AJ, Kurz EU, Duncan AM, Deeley RG. Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. Science 1992; 258:1650–1654.
191. Allen JD, Brinkhuis RF, van Deemter L, Wijnholds J, Schinkel AH. Extensive contribution of the multidrug transporters P-glycoprotein and Mrp1 to basal drug resistance. Cancer Res 2000; 60:5761–5766.
192. Flens MJ, Zaman GJ, van der Valk P, Izquierdo MA, Schroeijers AB, Scheffer AB, van der Groep GL, de Haas M, Meijer CJ, Scheper RJ. Tissue

distribution of the multidrug resistance protein. Am J Pathol 1996; 148: 1237–1247.
193. Leier I, Jedlitschky G, Buchholz U, Cole SPG, Deeley RG, Keppler D. The MRP gene encodes an ATP-dependent export pump for leukotriene C4 and structurally related conjugates. J Biol Chem 1994; 269:27807–27810.
194. Wijnholds J, Evers R, van Leusden MR, Mol CAAM, Zaman GJR, Mayer U, Beijnen J, van der Valk M, Krimpenfort I, Borst P. Increased sensitivity to anticancer drugs and decreased response in mice lacking the multidrug resistance-associated protein. Nat Med 1997; 11:1275–1279.
195. Ito S, Ieiri I, Tanabe M, Suzuki A, Higuchi S, Otsubo K. Polymorphism of the ABC transporter genes, MDR1, MRP1 and MRP2/cMOAT, in healthy Japanese subjects. Pharmacogenetics 2001; 11:175–184.
196. Perdu J, Germain DP. Identification of novel polymorphisms in the pM5 and MRP1 (ABCC1) genes at locus 16p13.1 and exclusion of both genes as responsible for pseudoxanthoma elasticum. Hum Mutat 2001; 17:74–75.
197. Oselin K, Mrozikiewicz PM, Gaikovitch E, Pahkla P, Roots I. Frequency of MRP1 genetic polymorphisms and their functional significance in Caucasians: detection of a novel mutation G816A in the human MRP1 gene. Eur J Clin Pharmacol 2003; 59:347–350.
198. Le Saux O, Urban Z, Tschuch C, Csiszar K, Bacchelli B, Quaglino D, Pasquali-Ronchetti I, Pope FM, Richards A, Terry S, Bercovitch L, de Paepe A, Boyd CD. Mutations in a gene encoding an ABC transporter cause pseudoxanthoma elasticum. Nat Genet 2000; 25:223–227.
199. Conrad S, Kauffmann HM, Ito K, Leslie EM, Deeley RG, Schrenk D, Cole SPC. A naturally occurring mutation in MRP1 results in a selective decrease in organic anion transport and in increased doxorubicin resistance. Pharmacogenetics 2002; 12:321–330.
200. Conrad S, Kauffmann HM, Ito K, Deeley RG, Cole SP, Schrenk D. Identification of human multidrug resistance protein 1 (MRP1) mutations and characterization of a G671V substitution. J Hum Genet 2001; 46:656–663.
201. Saito S, Iida A, Sekine A, Miura Y, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y. Identification of 779 genetic variations in eight genes encoding members of the ATP-binding cassette, subfamily C (ABCC/MRP/CFTR). J Hum Genet 2002; 47:147–171.
202. Moriya Y, Nakamura T, Horinouchi M, Sakaeda T, Tamura T, Aoyama N, Shirakawa T, Gotoh A, Fujimoto S, Matsuo M, Kasuga M, Okumura K. Effects of polymorphisms of MDR1, MRP1, and MRP2 genes on their mRNA expression levels in duodenal enterocytes of healthy Japanese. Biol Pharm Bull 2002; 25:1356–1359.
203. Paulusma CC, Bosma PJ, Zaman GJR, Baker CTM, Otter M, Scheffer GL, Scheper RJ, Borst P, Oude Elferink RPJ. Congenital jaundice in rats with a mutation in a multidrug resistance-associated protein gene. Science 1996; 271:1126–1128.
204. Van Aubel RA, Hartog A, Bindels RJ, van Os CH, Russel FG. Expression and immunolocalization of multidrug resistance protein 2 in rabbit small intestine. Eur J Pharmacol 2000; 400:195–198.
205. Schaub TP, Kartenbeck J, Konig J, Spring H, Dorsan J, Staehler G, Starkel S, Than WF, Keppler D. Expression of the conjugate export pump encoded by

the mrp2 gene in the apical membrane of kidney proximal tubules. J Am Soc Nephrol 1997; 8:1213–1221.
206. Kamisako T, Leier I, Cui Y, Konig J, Bucholz U, Hummel-Eisenbeiss J, Keppler D. Transport of monoglucuronosyl and bisglucuronosyl bilirubin by recombinant human and rat multidrug resistance protein 2. Hepatology 1999; 30:485–490.
207. Kartenbeck J, Leuschner U, Mayer R, Keppler D. Absence of the canalicular isoform of the MRP gene-encoded conjugate export pump from the hepatocytes in Dubin–Johnson syndrome. Hepatology 1996; 23:1061–1066.
208. Leier I, Hummel-Eisenbeiss J, Cui Y, Keppler D. ATP-dependent para-aminohippurate transport by apical multidrug resistance protein MRP2. Kidney Int 2000; 57:1636–1642.
209. Evers R, Kool M, van Deemter L, Janssen H, Calafat J, Oomen LC, Paulusma CC, Oude Elferink RP, Baas F, Schinkel AH, Borst P. Drug export activity of the human canalicular multispecific organic anion transporter in polarized kidney MDCK cells expressing cMOAT (MRP2) cDNA. J Clin Invest 1998; 101:1310–1319.
210. Huisman MT, Smit JW, Crommentuyn KM, Zelcer N, Wiltshire HR, Beijnen JH, Schinkel AH. Multidrug resistance protein 2 (MRP2) transports HIV protease inhibitors, and transport can be enhanced by other drugs. AIDS 2002; 16:2295–2301.
211. Nishino A, Kato Y, Igarashi T, Sugiyama Y. Both cMOAT/MRP2 and another unknown transporter(s) are responsible for the biliary excretion of glucuronide conjugate of the nonpeptide angiotensin II antagonist, telmisartan. Drug Metab Dispos 2000; 28:1146–1148.
212. Itoda M, Saito Y, Soyama A, Saeki M, Murayama N, Ishida S, Sai K, Suzuki M, Sugiyama Y, Ozawa S, Sawada JI. Polymorphisms in the ABCC2 (cMOAT /MRP2) gene found in 72 established cell lines derived from Japanese individuals: an association between single nucleotide polymorphisms in the 5′-untranslated region and exon 28. Drug Metab Dispos 2002; 30:363–364.
213. Shoda J, Suzuki H, Suzuki H, Sugiyama Y, Hirouchi M, Utsunomiya H, Oda K, Kawamoto T, Matsuzaki Y, Tanaka N. Novel mutations identified in the human multidrug resistance-associated protein 2 (MRP2/ABCC2) gene in a Japanese patient with Dubin–Johnson syndrome. Hepatology Res 2003; 27:322–325.
214. Mor-Cohen R, Zivelin A, Rosenberg N, Shani M, Muallem S, Seligsohn U. Identification and functional analysis of two novel mutations in the multidrug resistance protein 2 gene in Israeli patients with Dubin–Johnson syndrome. J Biol Chem 2001; 276:36923–36930.
215. Toh S, Wada M, Uchiumi T, Inokuchi A, Makino Y, Horie Y, Adachi Y, Sakisaka S, Kuwano M. Genomic structure of the canalicular multispecific organic anion-transporter gene (MRP2/cMOAT) and mutations in the ATP-binding-cassette region in Dubin–Johnson syndrome. Am J Hum Genet 1999; 64:739–746.
216. Kajihara S, Hisatomi A, Mizuta T, Hara T, Ozaki I, Wada I, Yamamoto K. A splice mutation in the human canalicular multispecific organic anion transporter gene causes Dubin–Johnson syndrome. Biochem Biophys Res Commun 1998; 253:454–457.

217. Wakusawa A, Machida I, Suzuki S, Hayashi H, Yano M, Yoshioka K. Identification of a novel 2026G-C mutation of the MRP2 gene in a Japanese patient with Dubin–Johnson syndrome. J Hum Genet 2003; 48:425–429.
218. Wada M, Toh S, Taniguchi K, Nakamura T, Uchiumi T, Kohno K, Yoshida I, Kimura A, Sakisaka S, Adachi Y, Kuwano M. Mutations in the canalicular multispecific organic anion transporter (cMOAT) gene, a novel ABC transporter in patients with hyperbilirubinemia II/Dubin–Johnson syndrome. Hum Mol Genet 1998; 7:203–207.
219. Paulusma CC, Kool M, Bosma PJ, Scheffer GL, Borg FT, Scheper RJ, Tytgat GNJ, Borst P, Baas F, Oude Elferink RPJ. A mutation in the human canalicular multispecific organic anion transporter gene causes the Dubin–Johnson syndrome. Hepatology 1997; 25:1539–1542.
220. Tate G, Li M, Suzuki T, Mitsuya T. A new mutation of the ATP-binding cassette, sub-family C, member 2 (ABCC2) gene in a Japanese patient with Dubin–Johnson syndrome. Genes Genet Syst 2002; 77:117–121.
221. Tsujii H, Konig J, Rost D, Stockel B, Leuschner U, Keppler D. Exon–intron organization of the human multidrug-resistance protein 2 (MRP2) gene mutated in Dubin–Johnson syndrome. Gastroenterology 1999; 117:653–660.
222. Kiuchi Y, Suzuki H, Hirohashi T, Tyson CA, Sugiyama Y. cDNA cloning and inducible expression of human multidrug resistance associated protein 3 (MRP3). FEBS Lett 1998; 433:149–152.
223. Fromm MF, Leake B, Roden DM, Wilkinson GR, Kim RB. Human MRP3 transporter: identification of the 5′-flanking region, genomic organization and alternative splice variants. Biochim Biophys Acta 1999; 1415:369–374.
224. Konig J, Rost D, Cui Y, Keppler D. Characterization of the human multidrug resistance protein isoform MRP3 localized to the basolateral hepatocyte membrane. Hepatology 1999; 29:1156–1163.
225. Kool M, van der Linden M, de Haas M, Scheffer GL, de Vree JM, Smith AJ, Jansen G, Peters GJ, Ponne N, Scheper RJ, Elferink RP, Baas F, Borst P. MRP3, an organic anion tranporter able to transport anti-cancer drugs. PNAS 1999; 96:6914–6919.
226. Soroka CJ, Lee JM, Azzaroli F, Boyer JL. Cellular localization and up-regulation of multidrug resistance-associated protein 3 in hepatocytes and cholangiocytes during obstructive cholestasis in rat liver. Hepatology 2001; 33:783–791.
227. Wagner M, Fickert P, Zollner G, Fuchsbichler A, Silbert D, Tsybrovskyy O, Zatloukal K, Guo GL, Schuetz JD, Gonzalez FJ, Marschall HU, Denk H, Trauner M. Role of farnesoid X receptor in determining hepatic ABC transporter expression and liver injury in bile duct-ligated mice. Gastroenterology 2003; 125:825–838.
228. Steinbach D, Wittig S, Cario G, Viehmann S, Mueller A, Gruhn B, Haefer R, Zintl F, Sauerbrey A. The multidrug resistance-associated protein 3 (MRP3) is associated with a poor outcome in childhood ALL and may account for the worse prognosis in male patients and T-cell immunophenotype. Blood 2003; 102:4493–4498.
229. Lang T, Hitzl M, Burk O, Mornhinweg E, Keil A, Kerb R, Klein K, Zanger UM, Eichelbaum M, Fromm MF. Genetic polymorphisms in the multidrug resistance-

associated protein 3 (ABCC3, MRP3) gene and relationship to its mRNA and protein expression in human liver. Pharmacogenetics 2004; 14:155–164.
230. Lee YMA, Cui Y, Konig J, Risch A, Jager B, Drings P, Bartsch H, Keppler D, Nies AT. Identification and functional characterization of the natural variant MRP3-Arg^{1297} His of human multidrug resistance protein 3 (MRP3/ABCC3). Pharmacogenetics 2004; 14:213–223.
231. Lee K, Klein-Szanto AJ, Kruh GD. Analysis of the MRP4 drug resistance profile in transfected NIH3T3 cells. J Natl Cancer Inst 2000; 92:1934–1940.
232. Van Aubel RA, Smeets PH, Peters JG, Bindels RJ, Russel FG. The MRP4/ABCC4 gene encodes a novel apical organic anion transporter in human kidney proximal tubules: putative efflux pump for urinary cAMP and cGMP. J Am Soc Nephrol 2002; 13:595–603.
233. Schuetz JD, Connelly MC, Sun D, Paibir SG, Flynn PM, Srinivas RV, Kumar A, Fridland A. MRP4 a previously unidentified factor in resistance to nucleoside-based antiviral agents. Nat Med 1999; 5:1048–1051.
234. Reid G, Wielinga P, Zelcer N, De Haas M, Van Deemter L, Wijnholds J, Balzarini J, Borst P. Characterization of the transport of nucleoside analog drugs by the human multidrug resistance proteins MRP4 and MRP5. Mol Pharm 2003; 63:1094–1103.
235. Chen ZS, Lee K, Kruh GD. Transport of cyclic nucleotides and estradiol 17-beta-D-glucuronide by multidrug resistance protein 4: resistance to 6-mercaptopurine and 6-thioguanine. J Biol Chem 2001; 276:33747–33754.
236. Wielinga PR, van der Heijden I, Reid G, Beijnen JH, Wijnholds J, Borst P. Characterization of the MRP4- and MRP5-mediated transport of cyclic nucleotides from intact cells. J Biol Chem 2003; 278:17664–17671.
237. Zelcer N, Reid G, Wielinga P, Kuil A, van der Heijden I, Schuetz JD, Borst P. Steroid and bile acid conjugates are substrates of human multidrug-resistance protein (MRP) 4 (ATP-binding cassette C4). Biochem J 2003; 371:361–367.
238. Reid G, Wielinga P, Zelcer N, van der Heijden I, Kuil A, de Haas M, Wijnholds J, Borst P. The human multidrug resistance protein MRP4 functions as a prostaglandin efflux transporter and is inhibited by nonsteroidal antiinflammatory drugs. PNAS 2003; 100:9244–9249.
239. Sampath J, Adachi M, Hatse S, Naesens L, Balzarini J, Flatley RM, Matherly LH, Schuetz JD. Role of MRP4 and MRP5 in biology and chemotherapy. AAPS PharmSci 2002; 4:E14.
240. Denk GU, Soroka CJ, Takeyama Y, Chen WS, Schuetz JD, Boyer JL. Multidrug resistance-associated protein 4 is up-regulated in liver but down-rregulated in kidney in obstructive cholestasis in the rat. J Hepatol 2004; 40: 585–591.
241. Dazert P, Meissner K, Vogelgesang S, Heydrich B, Eckel L, Bohm M, Warzok R, Kerb R, Brinkmann U, Schaeffeler E, Schwab M, Cascorbi I, Jedlitschky G, Kroemer HK. Expression and localization of the multidrug resistance protein 5 (MRP5/ABCC5), a cellular export pump for cyclic nucleotides, in human heart. Am J Pathol 2003; 163:1567–1577.
242. Rius M, Nies AT, Hummel-Eisenbeiss J, Jedlitschky G, Keppler D. Cotransport of reduced glutathione with bile salts by MRP4 (ABCC4) localized to the basolateral hepatocyte membrane. Hepatology 2003; 38:374–384.

243. Assem M, Schuetz EG, Leggas M, Sun D, Yasuda K, Reid G, Zelcer N, Adachi M, Strom S, Evans RM, Moore DD, Borst P, Schuetz JD. Interactions between hepatic Mrp4 and Sult2A as revealed by the constitutive androstane receptor and Mrp4 knockout mice. J Biol Chem 2004; 279:22250–22257.
244. Marzolini C, Leake BF, Kim RB. Identification of single nucleotide polymorphisms in the multidrug resistance-associated protein 4 (MRP4) [abstr]. 105th Annual Meeting of the American Society for Clinical Pharmacology and Therapeutics, Miami, March 24-27, 2004.
245. Wijnholds J, Mol CA, van Deemter L, de Haas M, Scheffer GL, Baas F, Beijnen JH, Scheper RJ, Hatse S, De Clercq E, Balzarini J, Borst P. Multidrug-resistance protein 5 is a multispecific organic anion transporter able to transport nucleotide analogs. PNAS 2000; 97:7476–7481.
246. Jedlitschy G, Burchell B, Keppler D. The multidrug resistance protein 5 functions as an ATP-dependent export pump for cyclic nucleotides. J Biol Chem 2000; 275:30069–30074.
247. Kool M, van der Linden M, de Haas M, Baas F, Borst P. Expression of human MRP6, a homologue of the multidrug resistance protein gene MRP1, in tissues and cancer cells. Cancer Res 1999; 59:175–182.
248. Madon J, Hagenbuch B, Landmann L, Meier PJ, Stieger B. Transport function and hepatocellular localization of mrp6 in rat liver. Mol Pharmacol 2000; 57:634–641.
249. Belinsky MG, Chen ZS, Shchaveleva I, Zeng H, Kruh GD. Characterization of the drug resistance and transport properties of multidrug resistance protein 6 (MRP6, ABCC6). Cancer Res 2002; 62:6172–6177.
250. Longhurst TJ, O'Neill GM, Harvie RM, Davey RA. The anthracycline resistance-associated (ara) gene, a novel gene associated with multidrug resistance in a human leukaemia cell line. Br J Cancer 1996; 74:1331–1335.
251. Le Saux O, Beck K, Sachsinger C, Silvestri C, Treiber C, Goring HH, Johnson EW, de Paepe A, Pope FM, Pasquali-Ronchetti I, Bercovitch L, Marais AS, Viljoen DL, Terry SF, Boyd CD. A spectrum of ABCC6/MRP6 mutations is responsible for pseudoxanthoma elasticum. Am J Hum Genet 2001; 69:749–764.
252. Pulkkinen L, Nakano A, Ringpfeil F, Uitto J. Identification of ABCC6/MRP6 pseudogenes on human chromosome 16q: implications for mutation detection in psedoxanthoma elasticum. Hum Genet 2001; 109:356–565.
253. Le Saux O, Beck K, Achsinger C, Treiber C, Goring HHH, Curry K, Johnson EW, Bercovitch L, Marais AS, Terry SF, Viljoen DL, Boyd CD. Evidence for a founder effect for pseudoxanthma elasticum in the Afrikaner population of South Africa. Hum Genet 2002; 111:331–338.
254. Germain DP, Remones V, Perdu J, Jeunemaitre X. Identification of two polymorphisms (c189G > C; c190T > C) in exon 2 of the human MRP6 gene (ABCC6) by screening of pseudoxanthoma elasticum patients: possible sequence correction?. Hum Mutat 2000; 16:449.
255. Le Saux O, Urban Z, Tschuch C, Csiszar K, Bacchelli B, Quaglino D, Pasquali-Ronchetti I, Pope FM, Richards A, Terry S, Bercovitch L, de Paepe A, Boyd CD. Mutations in a gene encoding an ABC transporter cause pseudoxanthoma elasticum. Nat Genet 2000; 25:223–227.

256. Meloni I, Rubegni P, De Aloe G, Bruttini M, Pianigiani E, Cusano R, Seri M, Mondillo S, Frederico A, Bardelli AM, Andreassi L, Fimiani M, Renieri A. Pseudoxanthoma elasticum: point mutations in the ABCC6 gene and a large deletion including also ABCC1 and MYH11. Hum Mut 2001; 18:85.
257. Cai L, Lumsden A, Guenther UP, Ramesar SA, Hohl D, Gallen DF, Neldner KH, Lindpaintner K, Richards RI, Struk B. A novel Q378X mutation exists in the transmembrane transporter protein ABCC6 and its pseudogene: implications for mutation analysis in pseudoxanthoma elasticum. J Mol Med 2001; 79:536–546.
258. Ringpfeil F, Pulkkinen L, Uitto L. Molecular genetics of pseudoxanthoma elasticum. Exp Dermatol 2001; 10:221–228.
259. Struk B, Cai L, Zach S, Chung WJJ, Lumsden A, Stumm M, Huber M, Schaen L, Kim CA, Goldsmith LA, Viljoen D, Figuera LE, Fuchs W, Munier F, Ramesar R, Hohl D, Richards R, Neldner KH, Lindpaintner K. Mutations of the gene encoding the transmembrane transporter protein ABC-C6 cause pseudoxanthoma elasticum. J Mol Med 2000; 78:282–286.
260. Bergen AA, Plomp AS, Schuurman EJ, Terry S, Breuning M, Dauwerse H, Swart J, Kool M, van Soest S, Baas F, ten Brink JB, de Jong PTVM. Mutations in ABCC6 cause pseudoxanthoma elasticum. Nat Genet 2000; 25:228–231.
261. Xu H, Plomp A, Wijnholds J, ten Brink J, van Soest S, van der Born LI, Leys A, Peek R, de Jong PTVM, Bergen AAB. ABCC6/MRP6 mutations: further insight into the molecular pathology of pseudoxanthoma elasticum. Eur J Hum Genet 2003; 11:215–224.
262. Ringpfeil F, Lebwohl MG, Christiano AM, Uitto J. Psedoxanthoma elasticum: mutations in the MRP6 gene encoding a transmembrane ATP-binding cassette (ABC) transporter. PNAS 2000; 97:6001–6006.
263. Wang J, Near S, Young K, Connelly PW, Hegele RA. ABCC6 gene polymorphisms associated with variation in plasma lipoproteins. J Hum Genet 2001; 46:699–705.
264. Ilias A, Urban Z, Seidl TL, Le Saux O, Sinko A, Boyd CD, Sarkadi B, Varadi A. Loss of ATP-dependent transport activity in pseudoxanthoma elasticum-associated mutants of human ABCC6 (MRP6). J Biol Chem 2002; 277:16860–16867.
265. Bera TK, Lee S, Salvatore G, Lee B, Pastan I. MRP8, a new member of ABC transporter superfamily, identified by EST database mining and gene prediction program, is higly expressed in breast cancer. Mol Med 2001; 7:509–516.
266. Bera TK, Iavarone C, Kumar V, Lee S, Lee B, Pastan I. MRP9, an unusual truncated member of the ABC transporter supefamily is highly expressed in breast cancer. PNAS 2002; 99:6997–7002.
267. Yabuuchi H, Shimizu H, Takayanagi SI, Isikawa T. Multiple splicing variants of two new human ATP-binding cassette transporters, ABCC11 and ABC12. Biochem Biophys Res Commun 2001; 288:933–939.
268. Doyle LA, Yang W, Abruzzo LV, Krogmann T, Gao Y, Rishi AK, Ross DD. A multidrug resistance transporter from human MCF-7 breast cancer cells. PNAS 1998; 95:15665–15670.
269. Xu J, Liu Y, Yang Y, Bates S, Zhang JT. Characterization of oligomeric human half ABC transporter ATP-binding cassette G2. J Biol Chem 2004; 279:19781–19789.

270. Zhou S, Schuetz JD, Bunting KD, Colapietro AM, Sampath J, Morris JJ, Lagutina I, Grosveld GC, Osawa M, Nakauchi H, Sorrentino BP. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nat Med 2001; 7:1028–1034.
271. Zhou S, Morris JJ, Barnes Y, Lan L, Schuetz JD, Sorrentino BP. Bcrp1 gene expression is required for normal numbers of side population stem cells in mice, and confers relative protection to mitoxantrone in hematopoietic cells in vivo. PNAS 2001; 99:12339–12344.
272. Jonker JW, Buitelaar M, Wagenaar E, van der Valk MA, Scheffer GL, Scheper RJ, Plosch T, Kuipers F, Oude Elferink RP, Rosing H, Beijnen JH, Schinkel AH. The breast cancer resistance protein protects against a major chlorophyll-derived dietary phototoxin and protoporphyria. PNAS 2002; 99: 15649–15654.
273. Krishnamurthy P, Ross DD, Nakanishi T, Bailey-Dell K, Zhou S, Mercer KE, Sorrentino BP, Schuetz JD. The stem cell marker Bcrp/ABCG2 enhances hypoxic cell survival through interactions with heme. J Biol Chem 2004; 279: 24218–24225.
274. Imai Y, Nkane M, Kage K, Tsukahara S, Ishikawa E, Tsuruo T, Miki Y, Sugimoto Y. C421A polymorphism in the human breast cancer resistance protein gene is associated wih low expression of Q141K protein and low-level drug resistance. Mol Cancer Ther 2002; 1:611–616.
275. Zamber CP, Lamba JK, Yasuda K, Farnum J, Thummel K, Schuetz JD, Schuetz EG. Natural allelic variants of breast cancer resistance protein (BCRP) and their relationship to BCRP expression in human intestine. Pharmacogenetics 2003; 13:19–28.
276. Honjo Y, Morisaki K, Huff LM, Robey RW, Hung J, Dean M, Bates SE. Single-nucleotide polymorphism (SNP) analysis in the ABC half-transporter ABCG2 (MXR/BCRP/ABCP1). Cancer Biol Ther 2002; 1:696–702.
277. Backstrom G, Taipalensuu J, Melhus H, Brandstrom H, Svensson AC, Artursson P, Kindmark A. Genetic variation in the ATP-binding cassette transporter gene ABCG2 (BCRP) in a Swedish population. Eur J Pharm Sci 2003; 18:359–364.
278. Honjo Y, Hrycyna CA, Yan QW, Medina-Perez WY, Robey RW, van den Laar A, Litman T, Dean M, Bates SE. Acquired mutations in the MXR/BCRP/ABCP gene alter substrate specificity in MXR/BCRP/ABCP-overexpressing cells. Cancer Res 2001; 61:6635–6639.
279. Mizuarai S, Aozasa N, Kotani H. Single nucleotide polymorphisms result in impaired membrane localization and reduced ATPase activity in multidrug transporter ABCG2. Int J Cancer 2004; 109:238–246.
280. Yoh K, Ishii G, Yokose T, Minegishi Y, Tsuta K, Goto K, Nishiwaki Y, Kodama T, Suga M, Ochiai A. Breast cancer resistance protein impacts clinical outcome in platinum-based chemotherapy for advanced non-small cell lung cancer. Clin Cancer Res 2004; 10:1691–1697.

6

Variability in Induction of Human Drug Metabolizing Enzymes

Allan B. Okey

Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada

1. INTRODUCTION: ENZYME INDUCTION IN CONTEXT

It has been evident for several decades that individuals differ markedly, one from the other, in their ability to metabolize drugs and other foreign chemicals (collectively termed "xenobiotics"). As described elsewhere in this volume, phenotypic variation in drug metabolism and drug response can have profound clinical implications (1–3). Hence it is important to decipher the mechanisms responsible for interindividual variation.

The total activity of a particular enzyme in vivo is determined primarily by three factors: (1) the inherent catalytic competence of the enzyme (turnover number: moles of substrate transformed per mole of enzyme per unit time); (2) the level of that enzyme which is expressed in relevant tissues; (3) the possible presence of agents that inhibit enzyme activity by competitive or non-competitive actions on the enzyme protein.

Until recently, when genetic explanations have been sought for phenotypic variations in drug metabolism, most emphasis has been placed on polymorphisms in the enzymes themselves. Numerous polymorphisms have been cataloged in coding regions of DMEs that alter the enzymes' catalytic functions or alter stability of the proteins. Additionally, some individuals are "poor metabolizers" for particular drugs because the relevant gene has been deleted whereas other individuals exhibit an "ultra-rapid" metabolic phenotype due to duplication or amplification of a DME gene

such as with CYP2D6 (4); for review see Reference 5 or other chapters in this monograph.

The expression level of DMEs is not a fixed immutable property of the individual. Many DMEs are highly responsive to environmental influences—for example, to modulation by endogenous regulatory signals or to induction by xenobiotic chemicals. The five P450 enzymes that typically are most abundant in human liver (CYPs 3A4, 2C9, 1A2, 2E1, and 2C19) (6,7) all are subject, to various degrees, to induction by xenobiotic chemicals as are important Phase II enzymes in the GST or UGT families.

Allan Conney has provided splendid personal narratives regarding the history of induction of drug-metabolizing enzymes (8–10); these constitute important sources for understanding how induction was discovered and why induction matters.

The focus in this chapter is on variability of induction in humans. Clearly, however, a great debt is owed to investigations in non-human species where many of the fundamental mechanisms and their variations first were discovered.

2. INDUCTION MECHANISMS

2.1. Levels at Which Regulation Can Occur

Induction occurs predominantly at the level of transcription (Fig. 1), often being regulated by specific receptors that interact with the chemical inducer (Fig. 2; Table 1). During the last decade it has become feasible to study the impact, on DME expression, of polymorphisms in regulatory factors in addition to genetic variation in coding regions of enzymes themselves. In a few instances, elevation of enzyme level and enzyme activity is

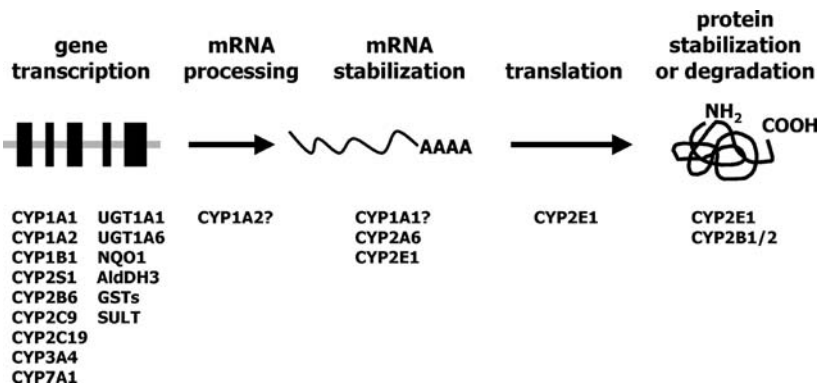


Figure 1 Overview of levels at which induction of drug-metabolizing enzymes can be regulated. *Source*: Modified from Ref. 223.

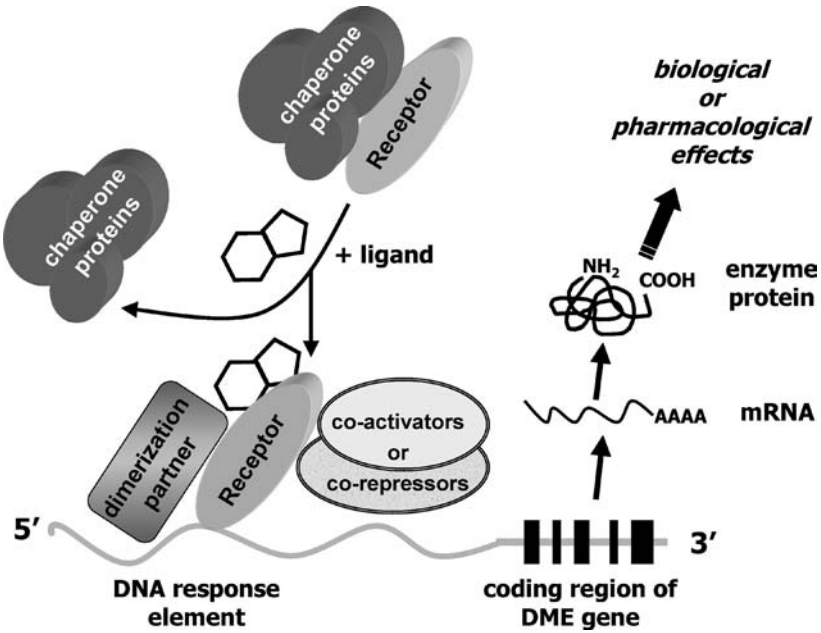


Figure 2 Schematic cell model of transcriptional regulation by nuclear receptors.

predominantly at post-transcriptional stages; the most notable example is the increase in CYP2E1 in response to ethanol and other small inducers, which stabilize the CYP2E1 protein (and also produce some up-regulation of transcription) (11–13).

2.2. A Brief Overview of Nuclear Receptors

Nuclear receptors are by no means the only cellular components responsible for regulating DME induction but essentially they provide the axis on which the main induction mechanism turns. Figure 2 provides a simplified generic scheme of how nuclear receptors go about their business of altering transcription rates. Nuclear receptors are soluble proteins that reside either in cytoplasm or in the nucleus prior to ligand binding. Typically, the nascent receptor protein is bound to chaperone proteins such as hsp90 that prevent receptor degradation by proteasomes. Ligand binding serves to expel the chaperone proteins, following which the receptor dimerizes either as a homodimer or with other common dimerization partners such as RXR (retinoid X receptor). Dimerization is essential for the receptor to be recognized by specific nucleotide sequences (response elements) located in the 5′-flanking regions of genes that are regulated by the particular receptor. AGGTCA is the "classical" core sequence for response elements for steroid

Table 1 Nuclear Receptors That Play a Major Role in Regulating Human Drug-Metabolizing Enzymes

Receptor or regulator	Prototype agonists	P450s Regulated	Phase II enzymes and other pharmacologically relevant genes-regulated	Selected references and reviews
CAR	Phenobarbital; 5β-pregnane 3,2,dione	CYP2A6; CYP2B6; CYP2C9[a]; CYP3A4; CYP2C19	UGT1A1; SULT	16, 132, 160, 190, and 191
FXR	Bile acids	CYP7A1 (down-regulated)		192
GR	Dexamethasone	CYP3A4? CYP2B6? CYP2C8/9? CYP2C19?		17 and 132
LXR	Oxysterols	CYP7A1; CYP3A4; CYP2B6		192, 193, and 194
PPARα	Clofibrate; statins	CYP4A11?		195
PPARγ	Thiazoladinediones		GSTA2 (Rat)	
PXR	PCN; rifampicin; hyperforin	CYP2A6; CYP3A4; CYP3A7; CYP2B6; CYP2C9; CYP2C19; CYP1A1; CYP1A2	MDR1; MRP2; UGT1A1; UGT1A6	16, 17, 19, 129, 132, 148, 160, 190, and 196–199
RXR		CYP26?		

VDR	$1\alpha,25(OH)_2D_3$	CYP3A4; CYP2B6; CYP2C9		29 and 200
AHR	2,3,7,8-TCDD 3-methylcholanthrene β-naphthoflavone	CYP1A1; CYP1A2; CYP1B1; CYP2S1	UGT1A1; UGT1A6; AldDH3; NQO1; GSTA1-1	14, 159, 160, 201, and 202
Nrf2	Phenolic antioxidants		GSTs; NQO1; HO-1	14 and 203

Note: This table focuses on regulation of enzyme induction in human systems; the literature cited is predominantly from studies in human cells or in "humanized mice" that carry a human transgene, which expresses the human form of that particular receptor.

[a]CAR appears to regulate basal expression of human CYP2C9 but is not responsible for 2C9 induction by phenobarbital.

Agents that act as inducers for FXR, CAR and Nrf2 cause these proteins to accumulate in the nucleus and bind to specific regulatory nucleotide sequences in the genes that these receptors regulate. However, direct binding of ligands to these "receptors" has been difficult to demonstrate, possibly because interactions of agonists with FXR, CAR, and Nrf2 are of relatively low affinity and technically difficult to detect. For Nrf2, the function of the chemical inducer may be to stimulate phosphorylation of Nrf2 by protein kinase C, releasing Nfr2 from the cytoplasmic Keap1 protein and permitting Nrf2 to translocate to the nucleus where it binds to the ARE.

There can be differences among various animal species regarding gene regulation by some of these receptors. For example, CYP4A enzymes are highly induced by PPARα agonists in rodents; however the effect in human cells is marginal (25). CYP3A4 induction via the PXR occurs in both human and mouse. However PCN is a potent PXR ligand in mouse cells but not in human whereas rifampicin is a potent 3A4 inducer in human cells but not in mouse cells, a difference that is correlated with their differential ability to activate reporter constructs bearing either mouse or human promoter sequences (198). Xie et al. (186) found, by creating "humanized mice" which express human PXR, that the species variation in chemical selectivity is due to the species origin of the receptor rather than to the promoter structure of the *CYP3A4* genes. In contrast, Zhang et al. (204), using a human *CYP3A4* reporter vector injected in vivo in intact mice, reported that the chemical selectivity is significantly influenced by the host-cellular-environment.

In most instances these receptors *increase* expression of the genes listed. However, certain ligands may *decrease* gene expression. For example, it has been proposed (194) that bile acids activate LXR which then competes with CXR, PXR, and CAR for binding to response elements, thereby inhibiting induction of CYP7A1, CYP3A4, and CYP2B6 by xenobiotics. Agents, that up-regulate expression of specific DMEs by acting as agonists for these receptors also can down-regulate expression of many other forms of P450 by mechanisms that are not yet well understood.

Source: From Refs. 131, 14, 20, 21.

receptors; however the ARE core is TGAC (14) and for the AHR is GCGTG (15).

While bound to its cognate response element, the liganded/dimerized receptor recruits co-activator proteins that link with additional transcription factors, often leading to acetylation of histones, which "opens up" the nucleosome to admit RNA polymerase II to the transcription start site. As would be expected, given that there is a sizeable superfamily of nuclear receptors and numerous interacting proteins, this simplified central theme is subject to many variations and complexities that allow subtle fine-tuning of regulatory responses.

Steroid receptors contain multiple functional domains (Fig. 3) that are modular and whose sequences can be extensively conserved across the superfamily. Thus it is not surprising that there is considerable "cross talk" between and among members of the superfamily and their various shared partner proteins. As will be evident from Table 3, there is considerable overlap in the spectrum of enzymes regulated by nuclear receptors. Many examples exist wherein an individual nuclear receptor regulates multiple enzymes (16).

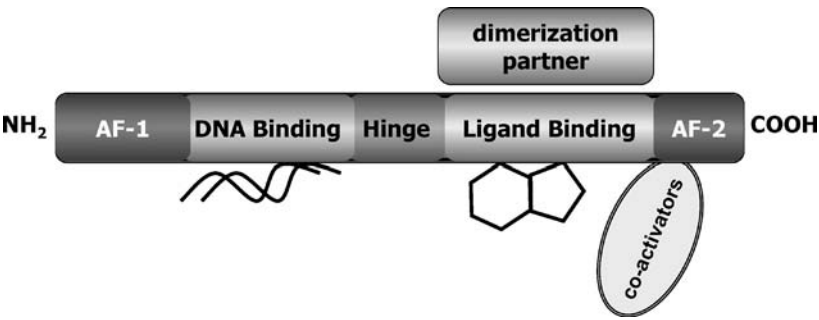


Figure 3 The generic diagram of functional domains in members of the steroid-receptor superfamily. The ligand-binding domain binds not only ligand but also serves as the recognition site for heat shock proteins, nuclear dimerization partners, and co-activator proteins. The DNA-binding domain contains two zinc-fingers that are characteristic motifs in members of the steroid receptor superfamily. The flexible hinge region permits the DNA-binding domain to rotate into contact with response element nucleotide sequences in the 5′-flanking region of genes being regulated. Transactivation function is widely dispersed, including key sequences in the activation function-1 domain (AF-1) near the amino terminus and the AF-2 domain at the carboxy terminus, which is essential to recruitment of co-activator proteins (adapted from Refs. 224 and 24). The AH receptor is not a member of the steroid receptor superfamily and does not contain zinc-fingers in the DNA binding domain. Rather, the AHR is a member of the bHLH (basic helix-loop-helix) PAS family of transcriptional regulators. Despite belonging to a different gene family, the general pattern of regulation by the AHR is similar to that of receptors for steroid hormones.

Conversely, as illustrated in Table 3, the expression of a particular enzyme can be regulated or influenced by several different receptors (17). A few examples: CYP2B6 expression in human primary hepatocytes or in human HepG2 hepatoma cells can be transactivated both by CAR and by PXR and their respective ligands; maximal induction involves both the phenobarbital-responsive enhancer module (PBREM) and a more distal responsive element that can bind either CAR or PXR (18); CYP2B6 is strongly induced in human primary hepatocyte cultures by phenobarbital, rifampicin, clotrimazole, and ritonavir—therapeutic agents that also cause induction of CYP3A4. The PXR appears to be the primary mediator of induction of both CYP2B6 and CYP3A4 by these compounds (19).

Although this chapter emphasizes up-regulation (induction) of DMEs by xenobiotic chemicals there are abundant examples of down-regulation (suppression) of DME levels by xenobiotic chemicals, perhaps even involving some of the same regulatory pathways responsible for induction (reviewed in Refs. 20 and 21).

One of the fascinating developments in gene regulation during the past decade has been the removal of "orphan" status from a throng of nuclear receptors. Orphan receptors were so named because they were discovered by methods of molecular biology ["reverse endocrinology" (22)] prior to anyone knowing their ligands or regulatory functions. Several previous orphans—CAR, FXR, LXR, and PXR (Table 1)—turn out to be key regulators of cytochromes P450 and other DMEs. For detailed current accounts of the role of nuclear receptors in regulation of drug-metabolizing enzymes, please see the excellent reviews by Waxman (23), Handschin and Meyer (24) and Tamasi et al. (25).

3. MAJOR FACTORS THAT INFLUENCE ENZYME INDUCTION: A SYNOPSIS

Section 3 is a brief outline and overview of the various factors or events, which might lead to variation in induction of human DMEs. Section 4 summarizes what currently is known about the factors that appear to be most important in governing induction of specific drug-metabolizing enzymes.

To what extent does variation in the activity of an inducible enzyme in vivo reflect the capacity (genetic potential) to *respond* to an inducer and to what extent does it reflect a difference in *exposure* to that inducer? This is the classical "nature/nurture" problem wherein both factors frequently are unknown. The capability for induction in humans usually has been assessed by studying human cells in culture (26–29) rather than by exposing human subjects to inducing agents. However, some in vivo human studies exist and they confirm that intact humans can, indeed, respond to inducers. There have not been many circumstances where induction per se was tested in humans. Most in vivo studies compare enzyme activities in populations

exposed to inducers with activities in unexposed control populations. For example, hepatic CYP1A2 is elevated in populations of smokers compared with non-smokers (30,31) and placental CYP1A1 is highly elevated in smoking populations compared with non-smokers (32,33).

In a few instances induction has been tested directly by measuring enzyme activity in an individual in a "basal" state, then administering a presumed inducting agent and re-measuring activity. By this procedure Michnovicz et al. (34–36) showed that indole-3-carbinol stimulates estrogen metabolism in women, probably by elevating levels of CYP1 enzymes. MacLeod et al. (37) found that consumption of pan-fried meat temporarily increased CYP1A2 activity (as measured by urinary caffeine metabolite ratios) and Murray et al. (38) observed a similar induction when subjects were fed a diet rich in cruciferous vegetables; caffeine metabolism reverted to pre-treatment levels when the diet was returned to normal.

Table 2 shows that there are very wide ranges of activity in humans for those DMEs which are amenable to induction. However, in most instances, we do not know what portion of this variation is due to variation in inducibility vs. variation in "basal" activity of the enzymes. Nor do we usually know what portion of the variation is due to differential exposure to chemicals that induce or that inhibit the inducible enzyme.

Some DMEs are simultaneously polymorphic and inducible; CYP2C9 and CYP2C19 are notable examples (39,40). Coding region polymorphisms that change enzyme structure can dramatically alter or even abolish the enzyme's catalytic function. Therefore, the range of enzyme activities observed in populations in vivo or in tissue samples can be very large—essentially infinite in the case where some individuals possess an enzyme whose catalytic function is near zero. Thus, it can be difficult to know what portion of variation in a population is attributable to induction vs. how much is accounted for by variation in catalytic function per se. This discrimination is greatly aided by identifying the relevant polymorphisms that affect enzyme structure as well as the polymorphisms that alter regulation of enzyme expression. Additionally, activity of some enzymes is highly susceptible to inhibition (competitive or otherwise) due to exposure to numerous xenobiotic chemicals.

3.1. Exposure to Xenobiotic Chemicals

Many inducers are substrates for the enzymes that they induce. From a teleological perspective, the "purpose" of enzyme induction is to foster clearance of xenobiotic chemicals from the body (the dioxin, TCDD, is a notable exception to this generalization). TCDD is a potent inducer of many DMEs that are regulated by the AHR (Table 3); however, TCDD is highly resistant to metabolism by any of the enzymes that it induces, particularly in

Table 2 Inducible Human Drug Metabolizing Enzymes: Examples of the Range of Variation of Expression Observed in Human Populations In Vivo or In Vitro

Enzyme	Range of variation (fold)	In vivo phenotyping probe	Or In vitro (tissue/endpoint)	References
CYP1A1	20		Liver/mRNA level	205 and 206
	26		Liver/mRNA level	101
	>40		Lung/catalytic activity	207
	67		Lung/protein level	86
	50		Placenta/catalytic activity	32
	400		Placenta/catalytic activity	
CYP1A2	6	Caffeine		30
	8	Caffeine		95
	13	Caffeine		93
	38–70	Caffeine		31
	70	Caffeine		208
	40		Liver/mRNA level	205
	500		Liver/mRNA level	206
CYP1B1	16		Liver/mRNA level	206
	42		Lung/protein level	207
	4000		Leukocytes/mRNA level	43
CYP2B6	9	Bupropion		209
	20		Liver/protein level	117
	>25		Liver/protein level	210
	41		Liver/protein level	211
	278		Liver/mRNA level	115
	45		Liver/catalytic activity	116

(*Continued*)

Table 2 Inducible Human Drug Metabolizing Enzymes: Examples of the Range of Variation of Expression Observed in Human Populations In Vivo or In Vitro (*Continued*)

Enzyme	Range of variation (fold)	In vivo phenotyping probe	Or In vitro (tissue/endpoint)	References
	288		Liver/protein level	116
CYP2S1	5		Skin/mRNA level	77
CYP2C9	>7	Warfarin		212
	>11		Intestine/mRNA level	213
	>7		Intestine/protein level	213
	>5		Intestine/catalytic activity	213
	>13		Liver/mRNA level	213
	>12		Liver/protein level	213
	>4		Liver/catalytic activity	213
CYP2C19	19	Mephenytoin		39 and 214
	>20	Mephenytoin		213
	>5		Intestine/mRNA level	213
	>10		Intestine/protein level	213
	>20		Intestine/catalytic activity	213
	>13		Liver/mRNA level	213
	>4		Liver/protein level	213
	>68		Liver/catalytic activity	

CYP2E1	5	Chlorzoxazone		215
	4	Chlorzoxazone		138
	6		Liver/protein level	135
	7		Liver/catalytic activity	135
	>50		Liver/protein level	136
CYP3A4	10		Liver/mRNA level	216
	20		Liver/catalytic activity	217
	31		Liver/catalytic activity	145
	118		Liver/mRNA level	218
UGT1A1	30		Liver/catalytic activity	219
UGT2B15	[a]21		Liver/catalytic activity	220
GSTA1-1	15		Liver/protein level	221
NQO1	1.5		Liver or lung/catalytic activity	172
	3		Breast tumor/catalytic activity	173
	12		Lung tumor/catalytic activity	173

[a]The authors were able to attribute about half of this variation to a coding-region polymorphism (D85Y) in the *UGT2B15* gene or to gender variation; they postulate that the remaining variation is due to regulatory polymorphisms or to exposure to xenobiotic inducers.

Table 3 Matrix Illustrating the Overlap in Enzymes That Are Regulated or Strongly Influenced by Various Nuclear Receptors in Human Cells or Tissues

	Cytochrome P450 Species													
	-1A1	1A2	1B1	2S1	2A6	2B6	2C9	2C19	3A4	4A	7A1	UGT1A1	GSTA1-1	NQO1
CAR					•	•	•	•				•		
FXR											•			
GR						•	•	•	•					
LXR						•			•	•[a]	•			
PXR	•	•			•	•	•		•			•		
VDR						•	•		•					
AHR	•	•	•	•								•	•	•
Nrf2													•	•

[a]The CYP4A11 is up-regulated by Wy-14,643 in human HepG2 cells stably transfected to overexpress PPARα (195). CYP4A1 was highly up-regulated by the potent PPARα agonist, Wy-14,643 in rat FaO hepatoma cells but not in human HepG2 hepatoma cells (222).

humans where the estimated biological half-life for TCDD is over 7 years (41).

Variation in the extent of exposure certainly makes a major contribution to interindividual variation in enzyme activities. From a research perspective it would be ideal to be able to identify human populations where exposure levels are known and where exposure is uniform so that the investigator could partition the magnitude of the induction response into genetic components vs. environmental components. Opportunities to study induction under circumstances of uniform exposure in humans are rare except, perhaps, under maximal inducing conditions such as heavy smoking or during prolonged high-dose exposure to therapeutic agents such as anticonvulsant drugs.

Some inducers are ingested as contaminants in food or air, some are dietary constituents originating in plant or animal tissue and some are therapeutic agents. A few examples of inducing chemicals that are relevant to human responses are summarized below.

3.1.1. Environmental or Dietary Chemicals

All humans inevitably are exposed to trace amounts of persistent environmental chemicals, even individuals who are not at special risk of exposure due to occupation or accident. The DMEs can be induced in humans by many persistent organic pollutants—for example: DDT, polychlorinated biphenyls (PCBs), dioxins, and polycyclic aromatic hydrocarbons such as benzo[a]pyrene (42). Cigarette smoke probably constitutes the most prevalent human exposure to a potent inducing stimulus, given that about one-third of adults in developed nations are regular smokers. As described above, cigarette smoke is a well-established inducer of CYP1A1 and CYP1A2 in humans via the AHR. Another CYP1 enzyme, CYP1B1, also is induced by AHR ligands, including dioxins. A study in Japanese subjects exposed to effluent from waste incinerators showed that CYP1B1 expression in peripheral blood lymphocytes was highly correlated with plasma dioxin levels (43).

3.1.2. Therapeutic Agents and Herbal Preparations

Anti-convulsant drugs such as barbiturates or phenytoin have long been known to stimulate their own metabolism and that of other drugs (24,42). Induction sometimes interferes with therapeutic responses by causing excessively rapid clearance of drugs such as anti-coagulants. These undesirable drug–drug interactions now are known to be due to induction of multiple DMEs, including CYP2B6, CYP2C9, CYP2C19, and CYP3A4 (Table 1).

Exposure to inducing agents is common but not uniform across the human population. The ultimate degree of enzyme induction depends both upon the extent of exposure and upon the receptivity of the host's cells to the inducing stimulus. Receptivity probably is strongly influenced by genetics.

However, as summarized below, there are multiple and complex stages at which receptivity might be influenced. Much remains to be learned about specific genetic and non-genetic factors that govern individual sensitivity to enzyme induction by xenobiotics.

3.2. Polymorphisms

3.2.1. Polymorphisms in Receptors that Regulate Induction

By convention, polymorphisms are genetic variations where the least common allele or least common phenotype is present at a frequency of at least 1%. Human populations also display numerous genetic variations (termed rare genetic variants) that are present at a frequency lower than 1%. The major receptors that regulate induction of DMEs in humans (Table 1) exhibit few genetic variations that qualify as true polymorphisms (24).

Before discussing potential sites of genetic variation, a reminder regarding the significance of genetic variation in nuclear receptors: for the receptors listed in Table 1, any polymorphisms that alter a receptor's function probably will affect numerous cellular pathways in addition to their effect on induction of drug-metabolizing enzymes. Although the focus of this chapter is on variation of DME induction, it is clear that most of the nuclear receptors involved in induction have duties other than just regulation of drug-metabolizing enzymes. For example, in mice, CAR regulates not only induction of specific DMEs but also affects constitutive expression of some of these enzymes as revealed by studies in CAR-null mice (44). Moreover, gene-array expression studies in CAR-null mice reveal that a wide spectrum of genes other than those encoding DMEs is altered by deletion of CAR and/or by treatment with CAR ligands. The AHR mediates induction of multiple Phases I and II DMEs (Table 3) but also plays important roles in fundamental cellular processes such as cell cycle regulation, cell proliferation, and apoptosis (45). The GR and VDR obviously play critical roles in regulation of metabolic pathways and mineral homeostasis in addition to their modulating effects on expression of several DMEs.

As shown in Figure 2, nuclear receptors have a modular construction. Specific functions reside primarily within specialized domains that have defined boundaries. Hypothetically, polymorphisms could exist in any domain. A polymorphism within a particular domain may selectively alter a single function of the receptor without necessarily disturbing other receptor functions. For example, a receptor could carry a polymorphism that interferes with ligand binding but might have normal transactivation domains that are fully capable of enhancing gene expression; however, if ligand cannot bind, the transactivation competence will be moot.

Below is a listing of the principal domains of nuclear receptor that hypothetically could contain polymorphisms along with a brief assessment

of whether such polymorphisms are present in human receptors. The impact on induction of specific genes will be discussed in Sec. 4.

- *Ligand Binding*. Polymorphisms that alter *affinity* for ligands are important determinants of the induction response in rodent models but no "affinity polymorphisms" have yet been discovered in human nuclear receptors (46). The *selectivity* for different ligands can vary substantially for the same receptor across different animal species. For example, there are striking differences among animal species regarding which ligands are the most effective agonists for PXR. Rifampicin is a potent PXR agonist in human, rhesus monkey, and rabbit but not in mice whereas dexamethasone is a good PXR ligand/CYP3A4 inducer in mice, rats, and rabbits but not in human cells (reviewed in Refs. 47 and 48).
- *Receptor Dimerization/Interaction with Chaperone Proteins.* No polymorphisms currently are known that affect this function.
- *Ability to Bind Specific DNA Response Elements*. No polymorphisms currently known that affect this function.
- *Transactivation Function*. Polymorphisms exist in AH receptor transactivation domains in rodents and in humans (see Sec. 4.1 for a discussion of the functional consequences).
- *Recruitment of Co-activators, Co-repressors*. No polymorphisms currently known that affect this function.

3.2.1.1. Polymorphisms in promoter regulatory sequences or introns of genes that are regulated: Several human DME genes contain polymorphisms in proximal promoter regions, the upstream 5′-flanking sequence, or introns; several of these alter up-regulation of the genes by xenobiotics. These regulatory sequence variants are described under particular enzymes in Section 4.

3.3. A Specific Suppressor of Xenobiotic-Induced Gene Expression: The AhRR

The AH receptor system encompasses a clever feedback mechanism that can dampen the response to xenobiotics that act via the AHR. One of the genes whose expression is highly induced by AHR ligands encodes the aryl-hydrocarbon-receptor-repressor. The AhRR protein is structurally similar to the AHR but does not bind ligand. The AhRR is able to recruit the AHRs dimerization partner, ARNT, and bind to AH response elements, thereby occluding the AHR and preventing the liganded AHR from enhancing transcription (49,50). The AhRR expression varies widely in different human tissues (51). It is induced by 3-methylcholanthrene in some human cells (the testis carcinoma line NEC8; mononuclear lymphocytes from adults) but not in other cells (mononuclear lymphocytes from infants or

from cord blood). The AhRR is very highly expressed in human adult testis and also is highly expressed in lung, ovary, spleen, and pancreas from adults but not in these tissues in the fetus (52). Polymorphisms exist in the human AhRR and epidemiologic studies have been conducted to determine if these polymorphisms are related to risk of reproductive disorders (53,54); however, no studies have yet assessed whether human AhRR polymorphisms influence induction of DMEs by the AHR.

The AhRR may stand as a unique repressor of DME induction by xenobiotics. No analogous "inducible repressor" is known to squelch DME induction by other nuclear receptors. However, members of the steroid receptor superfamily such as CAR, GR, LXR, PXR, etc., exhibit complex inhibitory or synergistic interactions because they alter each other's expression or because they compete for "generic" dimerization partners such as RXR.

3.4. Variation in Levels of Regulatory Components

The pharmacologic principles of ligand–receptor interaction predict that the level of receptor should affect dose–response curves for enzyme induction. In rodent systems, several endogenous and exogenous factors have been shown to alter levels of various receptors or other components in the induction pathway.

Levels of some receptors can be regulated, in part, by the receptor's own ligands. For example, basal levels of PPARγ in liver are low but agents that activate PPARγ also can up-regulate hepatic levels of this receptor (55). TCDD, the prototypical ligand for the AHR, causes short-term depletion of the AHR in rodent tissues in vivo and in rodent or human cells in culture (56,57) by provoking degradation of the receptor protein via the ubiquitin-proteasome system (58,59); this down-regulation of receptor levels can temporarily desensitize the cell to AHR-mediated events, including DME induction. With longer-term exposures, TCDD can modestly up-regulate AHR levels in rat liver (60) but this does not appear to substantially alter inducibility of CYP1A1.

Endogenous signals such as TGFβ down-regulate AHR in human cells in culture (61) whereas wnt/beta-catenin strongly up-regulates expression of AHR in human prostate cancer cells (62). Seven genetic variants have been detected in the 5′-flanking region of the human *AHR* gene but none of these variants appears to influence AHR expression as assessed in lymphocytes (63).

Interindividual differences in the levels of AHR mRNA in human peripheral blood lymphocytes have been reported to be small (64); however, little is known about population variation in AHR levels in other human tissues except for placenta where the receptor concentration is high in most subjects but not easily related to induction of CYP1A1 (33,65). The

AHR level in human lymphocytes was correlated positively with CYP1A1 inducibility by benz[a]anthracene in culture but negatively with CYP1B1 inducibility; however, AHR levels were correlated with basal CYP1B1 in untreated cells (66).

Glucocorticoids and glucocorticoid receptor have important modulatory effects on pathways mediated by other nuclear receptors. Most notably, glucocorticoids increase the levels of CAR, PXR, and RXRα in human hepatocyte cultures and thereby potentiate induction of CYP2B6, CYP2C8/9, and CYP3A4 (17,67).

It should be mentioned that another essential component in cytochrome P450 catalytic function, NADPH-cytochrome P450 oxidoreductase (68), is up-regulated at the transcript level by PPARα agonists in rodents but then appears to be down-regulated by post-transcriptional events (69). It remains to be seen if this phenomenon occurs in human cells and, if so, whether P450 reductase becomes rate limiting to the function of constitutive or induced P450s.

3.5. Pathological or Disease States

Inflammation and various disease states have been extensively studied for their impact on drug metabolism (70,71). With particular reference to DME induction, CAR expression has been shown to be decreased by proinflammatory cytokines as is subsequent phenobarbital induction of CYP genes (72). The acute endotoxin-induced inflammatory response leads to down-regulation of RXR, CAR, PXR, PPAR, FXR, and LXR in rat liver and the authors hypothesized that down-regulation of the relevant receptors is responsible for the suppression of CYP2C7, CYP2C23, and CYP3A1 observed in these experiments as well as for subsequent decreases in drug metabolism that occur during the acute-phase response (73). Studies of inflammation and infection have been conducted predominantly in animal models and in cells in culture; thus it is uncertain that modulation of nuclear receptor levels by inflammation has any important influence on induction of DMEs in humans in vivo.

4. VARIATION IN EXPRESSION OF SELECTED INDUCIBLE ENZYMES IN HUMANS: WHAT ARE THE EXPLANATIONS? THE NATURE/NURTURE INTERACTION

Table 2 illustrates the extraordinary range of human variation in the levels or activities of DMEs that are inducible. However, because factors other than induction also affect enzyme levels and activities, the ranges of variation listed in this table should not be taken as representing "fold-induction" nor should they be viewed as representing the range of inducibility for any particular enzyme. Some enzymes (e.g., CYP1A1) are virtually absent unless

the tissue has been exposed to a xenobiotic inducing agent. In cases where the denominator (lowest activity in the population) is zero or near zero, of course the apparent range of variation can appear very large, even infinite. Most of the inducible enzymes listed in this table (CYP1B1, CYP2B6, CYP2C9, CYP2C19, CYP3A4, UGT1, and GSTA-1) are expressed at significant constitutive levels in the absence of any known exposure to an inducing agent; hence the wide range of variation exhibited by these enzymes is not due to inflation by a minuscule denominator.

Some enzymes—notably CYP2C9 and CYP2C19—are encoded by genes which contain polymorphisms that alter structure and catalytic function of the enzyme rather than altering the level of expression; for 2C9 and 2C19 the range of catalytic variation in vivo is determined much more by the polymorphisms that alter enzyme structure than by the genetic and environmental factors that alter expression levels of the enzyme.

Space does not permit an in-depth analysis of the regulation of all of the inducible enzymes. The examples described are intended to give a sense of the variety of factors and mechanisms that contribute to variation in enzyme induction in humans.

4.1. CYP1A1, CYP1A2, CYP1B1, and CYP2S1—Factors That Affect Induction of This Battery of Enzymes

Each of these enzymes is regulated by the AH receptor; consequently they are induced by the same ligands and induction of each of the four enzymes may be influenced similarly by variation in shared regulatory factors such as the AHR itself. However, this does not mean that CYP1A1, CYP1A2, CYP1B1, and CYP2S1 universally are tightly co-regulated as a "package". For example, in humans CYP1A2 expression and CYP1A2 induction essentially are confined to liver whereas CYP1A1 expression and induction occur in virtually all cells and tissues but are low in human liver (74,75). CYP1B1 is expressed predominantly in non-hepatic tissues in humans such as breast where 1B1 expression is supported by estrogens (76). CYP2S1 is expressed and is inducible in human skin (77) and CYP2S1 mRNA is abundantly expressed in trachea, lung, small intestine, and spleen (78). Tissue-specific factors, most not yet identified, interact with AHR pathways to govern which cells respond to AHR ligands and which specific enzymes they will display.

4.1.1. AHR Polymorphisms

Variation in structure or function of the AHR would be expected to affect induction of all four of the above enzymes. We recently reviewed polymorphisms in the human AHR and their effects on AHR-mediated events (46). In brief, rodent models show clearly that polymorphisms which alter ligand affinity or transactivation function exert a major impact on induction

of CYP1 enzymes as well as on toxic responses to dioxin-like environmental contaminants (79).

Few polymorphisms have been detected in the human *AHR* gene. Thus far only one human *AHR* polymorphism has been shown to exert a substantial effect on receptor function; human *AHR* that encodes serine at codon 517, when in combination with isoleucine at 570 and lysine at 554 (a presumed haplotype found only in individuals of African descent), fails to sustain CYP1A1 induction in vitro even though this variant is competent to bind the inducer, TCDD, and to bind AH response elements derived from the *CYP1A1* gene (80).

The polymorphisms at codons 517, 554, and 570 in human AHR all reside in the transactivation domain rather than in the ligand-binding domain (46,81,82). In contrast, in inbred mice, the most prominent polymorphism affecting AHR function is a polymorphism in the ligand-binding domain which leads to a receptor phenotype wherein the affinity for TCDD is about 10-fold higher in the prototype AH "responsive" strain (C57BL/6) than in the prototype "non-responsive" strain (DBA/2). This difference is receptor affinity shifts the dose-response curve about one log to the right in DBA/2 (compared with C57BL/6) for multiple AHR-mediated responses including induction of CYP1 enzymes as well as numerous toxic endpoints (46,83–85). In humans, even small populations exhibit greater than a 10-fold variation in the affinity with which the AHR binds TCDD in placental cytosols; however, genetic examinations of individuals with the "high-affinity" phenotype vs. the "low-affinity" phenotype did not reveal any polymorphisms in the AHR itself, which would account for altered affinity (46,75). It also is not known whether this variation in affinity affects the capacity for induction of AHR-mediated enzymes. It is possible that some of the variation in the apparent affinity of binding resides at the level of chaperone proteins rather than the receptor itself but this remains to be tested.

4.1.2. CYP1A1

The CYP1A1 is highly induced in human placenta in response to cigarette smoke. Even within groups of smokers, the range of activity is as great as 400-fold (32,86). Twin studies suggest that the bulk of variation in placental CYP1A1 activity is environmental rather than genetic (87). However, the genetic constitution may be important in some individuals because placentas from a small fraction of smokers can be completely non-responsive, even to a heavy smoking stimulus (32,65).

CYP1A1 also is very highly induced (40–100-fold) in lung tissues of most smokers (88–90) with as much as a 180-fold range of interindividual variation among a smoking population (91). When attempting to understand individual variation, most examination of DNA focuses on polymorphisms. However, CYP1A1 illustrates another DNA-related mechanism by which constitutive expression or induction of DMEs might be regulated, that is

by promoter methylation. Hypermethylation of CpG islands, in promoters suppresses or silences expression of the gene. In a study by Anttila et al. (91) methylation of the *CYP1A1* promoter in normal lung tissue was more prevalent in populations of non-smokers or ex-smokers than in smokers and the degree of methylation increased within a few days after smoking ceased; similar patterns of methylation were observed in lung tumors as in normal lung tissue. The extent of methylation was not associated with the *CYP1A1*2A* or *CYP1A1*2B* variant alleles. CYP1A1 (EROD activity) tended to be higher in lung tissue of smokers with low methylation of the *CYP1A1* promoter but it remains to be seen if the methylation status is a major determinant of inducibility of lung CYP1A1 by smoking.

4.1.3. CYP1A2

The phenotype for human CYP1A2 activity can be conveniently measured in vivo using caffeine as a probe drug and measuring urinary or salivary metabolite profiles. The caffeine metabolic ratio is increased in populations of smokers (31,92) and wide ranges of CYP1A2 activity have been measured in many different human studies (Table 2).

Until recently it appeared that interindividual variation in CYP1A2 activity was predominantly due to environmental factors since no functionally significant polymorphisms could be found in the human *CYP1A2* gene or in factors that regulate 1A2 expression (93). However, subsequent studies revealed polymorphisms in intron 1 of the human *CYP1A2* gene that appear to account for differential expression and differential potential for induction, at least in certain individuals or ethnic groups.

Certain individuals exhibit an ultra-rapid CYP1A2 phenotype, most notably, some schizophrenic patients who are smokers (94–96). The ultra-rapid phenotype leads to sub-therapeutic plasma levels of the anti-psychotic drug, clozapine, which is cleared primarily by CYP1A2; thus it is important to determine the mechanistic basis for the ultra-rapid phenotype. Sachse et al. (97) discovered a polymorphism, -164C> A in intron 1 of the *CYP1A2* gene (now termed the *CYP1A2*1F* allele), that is related to increased caffeine metabolism in smokers and to increased inducibility of CYP1A2 by omeprazole (98). Genotyping of five schizophrenic patients who are ultra-rapid CYP1A2 metabolizers in two different laboratories showed that they were homozygous for the *CYP1A2*1F* "high-inducibility" allele (94–96). Ultra-rapid metabolism by the CYP1A2 pathway in these patients represents an example where environmental and genetic factors combine to produce an extreme phenotype. In these cases, heavy exposure to a strong inducing agent interacts with a genetically based highly responsive induction mechanism to produce an extreme phenotype, which interferes with pharmacotherapy of schizophrenia.

A further study by Sachse et al. (99) in a mixed population of smokers/non-smokers and colon cancer patients/controls, revealed a range of CYP1A2 activity in vivo (as measured by urinary caffeine phenotyping) that was at least 10-fold. In this study, no association was detected between any of the six most common *CYP1A2* genotypes and CYP1A2 hepatic activity. However, CYP1A2 activity was significantly higher in smokers than in non-smokers, accounting for some of the range of variation in the overall population.

Intron 1 of the human *CYP1A2* gene carries another polymorphism (−730C>T) that confers high-CYP1A2 activity in vivo and increased transcriptional response to the AHR agonist, TCDD, in cell culture (100). The allele frequencies of the −730C>T polymorphism, vary across different ethnic groups and may relate to the risk of adverse drug reactions from CYP1A2 substrates and to cancer risk from environmental contaminants such as aflatoxin (100). The mechanism by which the −730C>T polymorphism alters *CYP1A2* transcription appears to be related to the ability of a protein transcription factor to bind to an Ets element in intron 1; binding occurs when the element is the wild-type −730C but not with the −730T variant (100).

Very clearly the structure of regulatory regions of the *CYP1A2* gene itself may exert a profound influence on induction of its expression. Variation in factors outside of the *CYP1A2* gene also may contribute to the wide phenotypic variation seen for this enzyme in human populations. As mentioned in the general introduction to CYP1 enzyme regulation, polymorphisms in the AH receptor might be expected to alter induction of each of these four P450s. In the case of CYP1A2, we found that among healthy young women who smoke, the presence of at least one lysine-coding allele at codon 554 in the *AHR* gene was associated with higher caffeine metabolism than individuals who were homozygous for the arginine-coding allele at this position (Lam et al. in preparation; 46); polymorphism at codon 554 does not appear to significantly affect CYP1A1 activities in human lung, however (101).

The AHR must dimerize with ARNT in order to bind to AH response elements and alter gene transcription. Several polymorphic sites have been identified in the human *ARNT* gene but none of these appears to affect CYP1A2 activities in vivo as judged by caffeine phenotyping (102).

The CYP1A2 story also reminds us that there can be indirect mechanisms that strongly affect the induction response. For example, CYP1A2 is induced by omeprazole but only in CYP2D6 "poor metabolizers," perhaps because omeprazole has a longer half-life of in 2D6 poor metabolizers than those with the "rapid" 2D6 phenotype (103). In this instance, the genotype/phenotype for one species of P450 dictates the induction response of another P450 in a different family via pharmacokinetic effects on the inducer. Several current studies are devoted to determining how variation in expression

of drug transporters affects the induction of DMEs through modification of bioavailability and pharmacokinetics of the inducer.

One final twist in the regulation of CYP1A2 induction involves the nature of response elements in the 5′-flanking sequence of the human *CYP1A2* gene. Human CYP1A2 has long been suspected of being induced via regulatory sequences other than the canonical AH response element (AHRE; also known as DRE or XRE) that has been so thoroughly characterized for *CYP1A1* and many other genes in the AH gene battery (104). Very recently Sogawa et al. (105) identified a novel response element (termed XRE-II) in the 5′-flanking region of the rat *CYP1A2* gene. The nucleotide sequence of the XRE-II enhancer element is distinctly different from the conventional AHRE that mediates induction of CYP1A1 by AHR-ligands. For CYP1A1 induction, the ligand-AHR•ARNT complex binds directly to nucleotides in the AHRE. However, although AHR and ARNT are required for CYP1A2 induction, they do not bind directly to XRE-II. Rather, AHR and ARNT act essentially as co-activators by binding to a (currently unidentified) protein which itself binds directly to XRE-II (105). The human *CYP1A2* gene contains an XRE-II element, but it has not yet been proven that this element functions the same in human as in the rat (106). The presence of XRE-II provides yet another level of complexity and an opportunity for interindividual variation in expression and induction of CYP1A2.

4.1.4. CYP1B1

Aryl hydrocarbon hydroxylase (AHH) assays with benzo[a]pyrene as the substrate have been used for about three decades as a measure of CYP1A1 induction capacity in human leukocytes in vitro. However, recent experiments (43,75) indicate that the induced enzyme responsible for AHH activity in leukocytes probably is CYP1B1 rather than CYP1A1.

In fresh leukocytes harvested from workers exposed to dioxin-like substances at waste incinerators, levels of CYP1B1 mRNA appear to have a trimodal distribution (43). Although there is some disagreement about the nature of the distribution (107), there is no question that within a population exposed to similar levels of inducers, individuals can exhibit widely different CYP1B1 induction responses. Human leukocytes from different donors also exhibit a wide range of AHH inducibility after exposure to the prototype inducer, 3-methylcholanthrene, in culture. Cells from most individuals exhibit a "lower-inducibility" phenotype and only a few individuals demonstrate a "high-inducibility" response to 3-MC (75,108).

The mechanism responsible for high-inducibility vs. low-inducibility CYP1B1 phenotypes in vivo or in cell culture is not known. The codon 554 polymorphism in the human AHR (described in Sec. 4.1.3) was reported by Kawajiri et al. (81) not be associated with AHH inducibility in human lymphocytes from Japanese subjects in culture whereas Smart and Daly

(82) reported that induction of EROD in lymphocytes from Caucasian subjects was significantly higher in individuals with at least one lysine-coding allele at codon 554 than in subjects homozygous for arginine. The discrepancy between the findings of Kawajiri et al. vs. Smart and Daly might be attributable to ethnic differences between subjects in the two studies but it also may reflect differences due to the enzyme assays employed in the two studies. As mentioned above, AHH activity in leukocytes now appears to be predominantly carried out by CYP1B1. The EROD activity traditionally has been thought to be a highly selective marker for CYP1A1 catalytic function but this needs to be reassessed in leukocytes where, at an mRNA level, CYP1B1 appears to be the major induced CYP1 enzyme.

Several polymorphisms in the human *CYP1B1* gene are clinically associated with congenital glaucoma or have been epidemiologically evaluated for impact on cancer risk (75,109,110). Some coding-region polymorphisms in the *CYP1B1* gene reduce CYP1B1 catalytic function (111) whereas others produce a protein with "hyperactive" catalytic function (112); these polymorphisms alter enzyme function but are not related to CYP1B1 inducibility.

A search for a genetic explanation for varied regulation and expression of CYP1B1 revealed eight single-nucleotide polymorphisms (SNPs) in the *CYP1B1* promoter region but tests in vitro showed that none of these promoter sequence variations affected the ability of the promoter to respond to AHR ligands (113a). As was described for CYP1A1, it is possible that promoter methylation also might influence induction status. The *CYP1B1* is one of many genes that frequently become hypermethylated in breast cancers and there are individual differences in methylation that are associated with the status of estrogen receptor and/or progesterone receptor in the tumor (113b); the authors speculate that hypermethylation of the *CYP1B1* promoter will reduce expression of the gene but no experiments to confirm this prediction have been completed. Promoter methylation offers another level at which the capability for DME induction might be regulated and where interindividual differences in responsiveness might exist.

4.1.5. CYP2S1

The CYP2S1 is expressed in multiple human tissues (78) and, from studies in mouse, is known to be inducible by dioxins via the AH receptor (114). In human skin, CYP2S1 mRNA and protein can be induced by application of coal tar or *all-trans* retinoic acid to intact skin in vivo, but only 6 out of 13 subjects showed an induction response at the mRNA level (77). The basis for interindividual differences in response to inducers is not known for this recently discovered P450.

4.2. CYP2A6

The CYP2A6 was modestly up-regulated in primary cultures of human hepatocytes by the prototype CAR ligand, phenobarbital, and by the prototype PXR ligand, rifampicin. However, cells from one of the two donors in this study responded only to phenobarbital and not to rifampicin (16). The basis for this ligand-selective responsiveness is not known.

4.3. CYP2B6

Human hepatic CYP2B6 expression exhibits striking variation of nearly 300-fold at the mRNA (115) and protein (116) levels (Table 2). A large proportion of this variation could be due to varied levels of 2B6 induction. In laboratory animals, CYP2B enzymes are highly inducible by phenobarbital and related compounds. In human primary hepatocyte cultures, CYP2B6 is inducible by phenobarbital and by cyclophosphamide (117) and also by more than a dozen compounds that usually are considered to be CYP3A4 inducers (19). In vivo, CYP2B6 protein is elevated in several regions of brains from smokers and alcoholics (118). Human hepatic CYP2B6, as measured by bupropion hydroxylation, also is higher in livers from donors with a history of significant alcohol use than in donors without significant alcohol intake (116). Ethanol is known primarily for its ability to induce CYP2E1, so induction of CYP2B6 constitutes another example of the cross over between and among some traditional categories of inducers.

The mechanism by which CYP2B enzymes are induced in mammals became accessible due to pioneering experiments of Anderson and co-workers, which identified a phenobarbital response unit in the 5′-flanking sequence of the rat *CYP2B2* gene (119,120). Analogous response elements later were identified in the human *CYP2B6* gene (18,121), but there are substantial differences in mechanisms of receptor-mediated CYP2B regulation across animal species (24).

No functional polymorphisms have been reported in human CAR, the major nuclear receptor that mediates CYP2B6 induction. However, PXR, which also plays a significant role in CYP2B6 induction, exists as multiple allelic variants, some of which alter transactivation function of this receptor (122,123).

The CYP2B6 mRNA levels in human liver are correlated with mRNA levels of the nuclear receptors, CAR (115,124) and PXR (115). CYP2B6 expression also is strongly influenced by gender and ethnicity (124), but it is not clear if these factors affect primarily constitutive expression, induction, or both. The *CYP2B6* gene contains several non-synonymous exonic polymorphisms (124,125), many of which reduce or even abolish catalytic function (126). In addition, there are multiple *CYP2B6* polymorphisms that affect splicing (124). In the study by Miksys et al. (118), brain CYP2B6 levels in smoking alcoholics appeared to be influenced by the *CYP2B6*5*

polymorphism whereas in the study by Hesse et al. (116) the presence of the *CYP2B6*6B* haplotype reduced apparent induction by alcohol.

Despite recent extensive genotyping of human *CYP2B6*, none of the polymorphisms in coding or non-coding regions have been unequivocally associated with inducibility of the enzyme or have specific mechanisms that could account for variation in induction been identified; however, it has been suggested that nuclear receptors could play a role in regulating splicing (124).

4.4. CYP2C9 and CYP2C19

Much of the phenotypic variation in CYP2C9 and CYP2C19 activity is due to coding region polymorphisms that lead to defective catalytic function and a poor-metabolizer phenotype (39,127,128). However, CYP2C9 and CYP2C19 can be induced by various pharmaceuticals including phenobarbital, rifampicin, and dexamethasone. Hyperforin, a constituent of St. John's wort, induces CYP2C9 (129) and there is concern that such induction may interfere with pharmacotherapy. Case reports suggest that St. John's wort may accelerate metabolism of the anticoagulants, warfarin, and coumarins, possibly by inducing CYP2C9 (130).

The CYP2C9 and CYP2C19 are regulated by very similar mechanisms involving, mainly, the nuclear receptors, CAR and PXR. The human *CYP2C9* promoter contains a motif that binds both CAR and PXR and plays a role in basal expression of the gene in the absence of ligands; however, this motif does not appear to mediate induction of CYP2C9 by phenobarbital (131). The *CYP2C19* promoter also contains a CAR/PXR binding site that may regulate both constitutive expression and induction by xenobiotics (132). In addition to being regulated by CAR and PXR, expression of the *CYP2C19* gene also is significantly affected by glucocorticoids acting through the GR (132).

As described in Section 4.3, CYP2B6 is regulated by CAR and PXR, which also are the two key regulators of CYP2C9 and CYP2C19. Circumstances that alter induction of any one of these three enzymes are likely to also affect the expression of other enzymes since they share many of the same regulatory factors.

4.5. CYP2E1

The level of CYP2E1 traditionally has been viewed as being regulated primarily at post-transcriptional levels, mainly by protein stabilization (Fig. 1) (11,12). However, in rats exposed to high ethanol levels, increased *CYP2E1* transcription accounts for a significant fraction of the total increase in CYP2E1 protein levels and catalytic activity (133). CYP2E1 can be induced in laboratory animals by a wide variety of factors including obesity, overfeeding, and diabetes in addition to exposure to xenobiotic

chemicals such as ethanol and diverse hydrocarbons and solvents (12,13). Nicotine or other components of cigarette smoke induce CYP2E1 in animal models but the effect of smoking on CYP2E1 levels or activity in humans is uncertain (134).

In humans, CYP2E1 protein is elevated in hepatic tissues from patients who ingest large amounts of ethanol or those exposed to isoniazid (135–137). The human *CYP2E1* gene contains a polymorphism in the 5′-flanking region that is associated with increased CYP2E1 activity as assessed by use of the probe drug, chlorzoxazone in vivo (138). A repeat sequence in the *CYP2E1*1D* allele appears to disrupt a negative regulatory element and thereby increases *CYP2E1* expression (139). The frequency of the *CYP2E1*1D* allele differs widely across ethnic groups with a high of about 30% in Indo-Asian subjects to a low of about 2% in Caucasians, all living in Canada (140).

The range of human variation in hepatic CYP2E1 levels is not as great as that for most of the other inducible cytochrome P450 species (Table 2). At present, we possess little understanding of what proportion of interindividual variation is related to enzyme induction nor do we have any real understanding of what contribution is made by genetic factors. No known receptor mediates induction of CYP2E1 at the transcriptional level and there is a lingering mystery regarding the means by which small molecules such as ethanol are able to signal the transcriptional apparatus.

4.6. CYP3A4, CYP3A5, and CYP3A7

CYP3A4 is responsible for biotransformation of a larger number and broader range of therapeutic substrates than any other DME (7,141,142). CYP3A4's special importance in metabolism of therapeutic agents and as a site for undesirable drug–drug interactions prompted substantial efforts to determine the basis for its wide interindividual phenotypic variation.

The distribution of CYP3A4 activity in human populations tends to be unimodal, suggesting that multiple factors, both genetic and non-genetic, make important contributions to phenotypic variation (143). Studies in twins, comparisons of between-person and within-person variances for CYP3A4 activity (in several human populations) along with the method of repeated drug administration (RDA) (144) collectively indicate that as much as 90% of variation in human hepatic CYP3A4 activity is under genetic control. Genetic control encompasses possible genetic variation in catalytic function of the CYP3A4 enzyme per se as well as genetically based variation in regulation of its expression. Most evidence points to regulatory polymorphisms rather than to structural polymorphisms in the *CYP3A4* gene itself as an explanation for the wide interindividual variation (145–147).

The CYP3A4 activity in humans in vivo is highly susceptible to alteration by xenobiotic chemicals because CYP3A4 can be both induced by a

wide variety of chemicals and also can be inhibited by many xenobiotics. Several therapeutic agents induce human CYP3A4 as does hyperforin, a constituent of St. John's wort (148). The CYP3A4 substrate pocket accommodates an extraordinary range of xenobiotic chemicals; hence CYP3A4 catalytic activity can be competitively inhibited in individuals exposed to the myriad therapeutic agents that are CYP3A4 substrates or to certain dietary constituents (143,149).

Induction of CYP3A enzymes is primarily mediated by PXR. However, CAR also plays a prominent role because several ligands that bind PXR also bind CAR. There is further cross-regulation at the level of response elements for these two nuclear receptors (16,19,47,150,151). The abundance of PXR itself is correlated with the level of expression of CYP3A4, CYP3A5, and CYP transcripts and CYP3A proteins in human liver (152). Three additional nuclear receptors, GR, LXR, and VDR, exert modulatory roles in CYP3A induction (Table 3).

There are striking differences among animal species regarding which ligands are the most effective agonists for PXR. For example, rifampicin is a potent PXR agonist in human, rhesus monkey, and rabbit but not in mice whereas dexamethasone is a good PXR ligand/CYP3A4 inducer in mice, rats, and rabbits but not in human cells (reviewed in Refs. 47 and 48). The key amino acids that determine ligand specificity lie at the pore to the ligand-binding cavity (48). It is conceivable that genetic variants, which affect ligand selectivity of PXR might shift an individual's responsiveness to some CYP3A4 inducers, but not to others; however, it remains to be tested whether genetic variants exist in this critical domain in human PXR.

Expression of CYP3A4 is particularly high in liver. Recent experiments in mice demonstrate that both basal expression and full induction of CYP3A4 by xenobiotics require the participation of the orphan nuclear receptor, HNF4α (153). The nucleotide site that interacts with HNF4α has been identified about 7.7 kb up-stream in the human *CYP3A4* gene. This site is located within the xenobiotic-responsive enhancer module (XREM) that mediates both basal expression and induction by ligands for either PXR or CAR (153). It is possible that some of the variability in human expression of CYP3A4 is due to varied abundance of HNF4α (147).

The *CYP3A4* promoter contains an A→G nucleotide change at −290 (*CYP3A4*1B* variant) whose frequency varies dramatically among different ethnic groups (145,154). Molecular epidemiologic studies report association of the *CYP3A4*1B* allele with risk of some cancers but not others (155). The *CYP3A4*1B* appears to modestly elevate CYP3A4 basal expression but does not increase responsiveness to inducers such as rifampicin or dexamethasone (155).

Matsumura et al. (147) identified a novel enhancer module, CLEM4, far upstream (about 11 kb) in the human *CYP3A4* gene. CLEM4 binds multiple transcription factors, some of them liver-specific such as HNF-4.

CLEM4 appears to regulate constitutive expression of CYP3A4 rather than induction; this is reflected in its acronym, which stands for "constitutive liver enhancer module." CLEM4 is polymorphic in Caucasians; a TGT insertion disrupts a USF1 binding site and reduces enhancer activity about 35% in a reporter gene system in vitro. The TGT insertion allele is present at a frequency of about 3% in Caucasian subjects but was not found in 131 Japanese subjects (147).

A polymorphism in intron-3 of the human *CYP3A5* gene alters constitutive expression levels by affecting transcript splicing; only individuals with at least one *CYP3A5*1* allele express significant levels of CYP3A5 transcript and protein (156,157). This polymorphism seems to affect constitutive expression but not inducibility of CYP3A5.

Another fascinating variant in human *CYP3A* gene structure causes the promoter from *CYP3A4* to be substituted into the promoter of the *CYP3A7* gene (158). In most individuals CYP3A7 expression is prominent in fetal life but is lost in the adult. Individuals whose *CYP3A7* gene carries the *CYP3A4* promoter sequence continue to express high levels of CYP3A7 into adulthood. The segment of the *CYP3A4* promoter found in the *CYP3A7* gene in these cases contains a functional ER6 motif, a site that may mediate induction via PXR and CAR. However, it is not yet known whether inducibility, in response to xenobiotic exposure in vivo, is altered in such individuals.

As described in Sec. 4.3 (for CYP2B6), the important regulator, PXR, has recently been shown to exhibit allelic variation at multiple sites and some of these polymorphisms alter PXRs transactivation function (122,123). Potentially, PXR polymorphisms may explain not only part of the interindividual variation in expression of CYP3A enzymes but also CYP2B6, CYP2C9, and CYP2C19 (Table 1).

Despite diligent and exhaustive searches, polymorphisms that alter CYP3A4 protein structure have, until recently, been conspicuous by their rarity in the human species. Therefore, the explanation for the wide variation in CYP3A4 activity among humans seems to depend primarily on the level of enzyme that is expressed or on the extent to which enzyme activity is affected by inhibitors. It is obvious that regulation of CYP3 enzymes is multi-factorial and complex. The mechanistic basis for the striking variation in human CYP3A4 phenotype remains elusive but progress is being made on this topic, which is important from the viewpoint of improved therapeutics as well as for fundamental science.

4.7. UGT1A1

UGT1A1 is important because it is virtually the only enzyme in humans that can conjugate bilirubin for excretion (159). Human UGT1A1 induction is mediated by at least three nuclear receptors, CAR, PXR, and AHR (Table 3)

(159,160). Therefore, UGT1A1 is susceptible to induction by exposure to a wide variety of xenobiotic chemicals (Table 1) and its expression can be influenced by the variations in receptor levels or receptor structure described previously for other enzymes regulated by CAR, PXR, and AHR.

Human *UGT1A* genes contains multiple polymorphisms that alter enzyme structure (161) as well as a TATA-box promoter polymorphism, *UGT1A1*28,* that leads to reduced constitutive expression in conditions such as Gilbert's syndrome (162,163) and to decreased conjugation of 2-hydroxyestradiol (164). However, there is no evidence that the variant *UGT1A1*28* allele affects induction of UGT1A1.

4.8. GSTs

Human alpha-class glutathione S-transferases are highly variable in their expression among different individuals. These enzymes are inducible by many drugs, antioxidants, and dietary constituents (14). The *GSTA1* and *GSTA2* genes contain a large number of polymorphisms in their promoters (165–168). Some of these polymorphisms are associated with varied constitutive expression of the GST enzymes in human liver in vivo and some lead to varied expression when the promoters are experimentally ligated to reporter genes and tested for transcriptional activity by transfection into host cells. However, it has not yet been demonstrated that any of these promoter polymorphisms confer varied GST induction in response to chemical exposure in vivo or in human cells in culture.

4.9. NQO1

The NQO1 (previously known as DT-diaphorase) is important for its role in protecting from reactive oxygen species and quinones in addition to its ability to activate certain anti-neoplastic quinones such as mitomycin C into cytotoxic products (169,170). NQO1 activity and *NQO1* gene expression are substantially higher in many tumors (breast, lung, liver, and colon) than in their normal tissue counterpart (169,171–174). In normal human lung, NQO1 activity is slightly higher in smokers than in non-smokers but this possible enzyme induction by smoking evident in lung tumors because NQO1 expression already is elevated (169).

In rodents and in human cells, the transcription factor, Nrf2, can be activated by phenolic antioxidants to bind to the antioxidant response element (ARE), thereby enhancing transcription of genes encoding Phase II enzymes including *GST-1* and *NQO1* (14,175,176).

Interindividual variation in NQO1 expression is present in human populations, but the range of variation is relatively modest (Table 2). Some variation is due to coding-region polymorphisms that can nearly abolish catalytic function of the enzyme (177). Additional variation may be attributable to varied inducibility of the enzyme. The 5′-flanking region of the human

NQO1 gene contains at least five SNPs but none of these is present at a frequency of more than 2% (177). It is not known if any of these SNPs alters *NQO1* expression or induction by xenobiotic chemicals.

5. CONCLUSIONS

5.1. Adverse vs. Beneficial Effects of Enzyme Induction

5.1.1. Adverse Effects of Enzyme Induction

There are clinical circumstances in which induction of drug-metabolizing enzymes can compromise the success of pharmacotherapy or can increase the risk of toxicity from drugs or environmental chemicals. For example, the anti-tubercular drug, rifampicin is a potent inducer of CYP3A enzymes in humans as is hyperforin, a constituent of the herbal preparation, St. John's wort (148). Induction of CYP3A enzymes by St. John's wort has been associated in case reports with accelerated metabolism of cyclosporin, leading to failure to control transplant rejection (178) and with impaired HIV therapy by enhancing clearance of indinavir (179). As described under Section 4.1.3, induction of CYP1A2 by heavy smoking in some schizophrenic patients can lead to ultra-rapid elimination of the anti-psychotic drug, clozapine, and leave the patient under-medicated (94–96). A review of the clinical consequences of enzyme induction by Fuhr (180), however, concludes that DME induction in humans rarely leads to toxicologic effects with the notable exception of hepatotoxicity associated with induction by ethanol.

5.1.2. Beneficial Effects of Enzyme Induction

Drug-metabolizing enzymes such as P450s can generate toxic reactive intermediates from numerous xenobiotic chemicals. Nevertheless, high activity of P450s, in concert with increased Phase II enzymes appears, on balance, to be protective to laboratory animals exposed to xenobiotics in vivo (42,75). In most circumstances, the same probably is true for humans. Enzymes induced via the AHR are highly efficient at converting a wide spectrum of environmental chemicals into their ultimate mutagenic, carcinogenic, teratogenic, and cytotoxic derivatives in vitro. However, the suite of enzymes induced via the AHR in vivo accelerates clearance of the prototoxicants from the body by a first-pass clearance effect, thereby reducing overall exposure of vulnerable tissues and cells (42,75). PXR also is noteworthy as a front-line recognition site that mobilizes multiple Phases I and II enzymes to protect from xenobiotic chemicals by accelerating their biotransformation and clearance (Table 1).

Because induction of DMEs can enhance clearance of xenobiotic chemicals, specific agents are being tested for potential prevention of chemically induced toxicity, especially for chemoprevention of cancer. Oltipraz, originally developed as an anti-schistosomal drug, induces multiple Phase

II enzymes via the Nrf2/ARE pathway. In laboratory animals, oltipraz has been highly effective at inhibiting development of liver cancers in animals exposed to potent carcinogens such as aflatoxin B1. Oltipraz is undergoing clinical trials in China in populations who are at high risk for aflatoxin-induced liver cancers (181).

5.2. Our Present Understanding, Challenges, and Prospects for the Future

Despite extensive investigation over multiple decades, the plain fact remains: although we recognize wide phenotypic variation in the expression of inducible DMEs in humans, there are few cases where we can provide a cogent genetic or mechanistic explanation for that variation.

The receptors that regulate induction of xenobiotic-metabolizing enzymes appear to be remarkably conserved and contain relatively few polymorphisms. This stands in strong contrast to nuclear receptors that mediate the action of reproductive steroids. For example, more than two hundred polymorphisms and genetic variants have been described in the human androgen receptor gene, many of which lead to loss of response to androgens (182). The human estrogen receptor also displays many polymorphisms and splice variants (183).

In the foreseeable future, extensive SNP patterns or haplotype maps of the complete human genome may reveal multiple and complex genetic variants that function individually to modulate induction or that interact to influence induction of DMEs.

Until very recently, the approach to developing therapy optimized to the individual patient was focused to single-gene traits such as polymorphisms that alter structure and function of the drug-metabolizing enzymes themselves (184). It is not yet possible to predict the potential for inducibility of DMEs in individual patients based on genotypic information alone because induction mechanisms inherently involve multiple complex regulatory factors wherein we currently have only fragmentary understanding.

It would be of great value to be able to predict inducibility of key DMEs from genotype alone. However, currently there are very few reliable genotypic markers that predict the induction phenotype in humans for any of the common DMEs. Gashaw et al. (185) assessed CYP3A4 mRNA levels in human leukocytes before and after the subjects were treated in vivo with rifampicin. Rifampicin treatment increased clearance of the probe drug, alprazolam, by about fivefold, indicating substantial induction of hepatic CYP3A4. The CYP3A4 mRNA in leukocytes from the subjects was elevated nearly twofold but the large variation in mRNA levels in the population precluded any use of leukocyte mRNA levels to predict induction of hepatic CYP3A4. This is an example of a combined in vivo/molecular strategy that

may help to find the firmest experimental grounds for relating human phenotype to genotype.

An alternative approach is to create additional strains of "humanized mice" in which the human genes that regulate enzyme induction are "knocked-in" to the mouse genome. This has been accomplished for PXR (186,187), the AHR (188). Humanized mice allow important questions to be addressed such as how wild type vs. genetically variant forms of human receptors perform in vivo and whether species differences in response to various ligands reside in the receptor itself or in other components. For example, Cheung et al. (189), recently showed that mice humanized for PPARα do not exhibit hepatocellular proliferation in response to the PPARα ligand, Wy-14643, whereas mice with the normal wildtype mouse form of PPARα do develop hepatocellular proliferation. This species difference in receptor function may explain why hepatocarcinomas are readily induced in mice given fibrate drugs that act via PPARα whereas, fortunately, humans are resistant to the peroxisome proliferating effects of fibrates and do not develop liver cancers from these drugs (6). Despite the great promise of humanized mice, given the combinatorial complexity, overlap and interplay of regulatory pathways (briefly summarized in this review), the humanized mouse may not always faithfully predict how a specific regulatory molecule would actually function within its normal context in the human body.

For now, we can explain some of the interindividual variation in expression of inducible DMEs but the field is open for additional clever new ways to resolve the large proportion variation that remains unexplained.

ABBREVIATIONS

AHH	Aryl hydrocarbon hydroxylase
AHR	Aryl hydrocarbon receptor
AHRE	AH response element (also known as DRE = dioxin response element or XRE = xenobiotic response element)
AhRR	Aryl hydrocarbon receptor repressor
AlDH	Aldehyde dehydrogenase
CAR	Constitutive androstane receptor
CLEM	Constitutive liver enhancer module
DME	Xenobiotic-metabolizing enzyme
EROD	Ethoxyresorufin *O*-deethylase
FXR	Farnesoid X receptor
GR	Glucocorticoid receptor
GST	Glutathione S-transferase
HO	Heme oxygenase

LXR	Liver X receptor
Nrf2	NF-E2-related factor
NQO	NAD(P)H:quinone oxidoreductase
PCB	Polychlorinated biphenyl
PCN	Pregnenolone 16α-carbonitrile
PPAR	Peroxisome proliferator activated receptor
PXR	Pregnane X receptor (also known as SXR)
RXR	Retinoid X receptor
SNP	Single-nucleotide polymorphism
SULT	Sulfotransferase
TCDD	2,3,7,8-tetrachlorodibenzo-*p*-dioxin
UGT	UDPG-glucuronosyltransferase
VDR	Vitamin D receptor

ACKNOWLEDGMENTS

Research in the author's laboratory is supported by Grants from the National Cancer Institute of Canada (with funds from the Canadian Cancer Society) and from the Canadian Institutes of Health Research.

REFERENCES

1. Eaton DL, Gallagher EP, Bammler TK, Kunze KL. Role of cytochrome P4501A2 in chemical carcinogenesis: implications for human variability in expression and enzyme activity. Pharmacogenetics 1995; 5:259–274.
2. Evans WE, McLeod HL. Pharmacogenomics—drug disposition, drug targets, and side effects. N Engl J Med 2003; 348:538–549.
3. Weinshilboum R. Inheritance and drug response. N Engl J Med 2003; 348:529–537.
4. Dahl ML, Johansson I, Bertilsson L, Ingelman-Sundberg M, Sjoqvist F. Ultrarapid hydroxylation of debrisoquine in a Swedish population. Analysis of the molecular genetic basis. J Pharmacol Exp Ther 1995; 274:516–520.
5. Ingelman-Sundberg M. Genetic susceptibility to adverse effects of drugs and environmental toxicants. The role of the CYP family of enzymes. Mutat Res 2001; 482:11–19.
6. Pelkonen O, Maenpaa J, Taavitsainen P, Rautio A, Raunio H. Inhibition and induction of human cytochrome P450 (CYP) enzymes. Xenobiotica 1998; 28:1203–1253.
7. Guengerich FP. Cytochromes P450, drugs, and diseases. Mol Interv 2003; 3:194–204.
8. Conney AH. Enzyme induction and dietary chemicals as approaches to cancer chemoprevention: the Seventh DeWitt S. Goodman Lecture. Cancer Res 2003; 63:7005–7031.
9. Conney AH. Induction of microsomal enzymes by foreign chemicals and carcinogenesis by polycyclic aromatic hydrocarbons: G. H. A. Clowes Memorial Lecture. Cancer Res 1982; 42:4875–4917.

10. Conney AH. Induction of drug-metabolizing enzymes: a path to the discovery of multiple cytochromes P450. Annu Rev Pharmacol Toxicol 2003; 43:1–30.
11. Novak RF, Woodcroft KJ. The alcohol-inducible form of cytochrome P450 (CYP 2E1): role in toxicology and regulation of expression. Arch Pharm Res 2000; 23:267–282.
12. Lieber CS. Microsomal ethanol-oxidizing system (MEOS): the first 30 years (1968–1998)—a review. Alcohol Clin Exp Res 1999; 23:991–1007.
13. Caro AA, Cederbaum AI. Oxidative stress, toxicology, and pharmacology of CYP2E1. Annu Rev Pharmacol Toxicol 2004; 44:27–42.
14. Nguyen T, Sherratt PJ, Pickett CB. Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. Annu Rev Pharmacol Toxicol 2003; 43:233–260.
15. Denison MS, Fisher JM, Whitlock JP Jr. The DNA recognition site for the dioxin-Ah receptor complex. Nucleotide sequence and functional analysis. J Biol Chem 1988; 263:17221–17224.
16. Maglich JM, Stoltz CM, Goodwin B, Hawkins-Brown D, Moore JT, Kliewer SA. Nuclear pregnane X receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. Mol Pharmacol 2002; 62:638–646.
17. Pascussi JM, Gerbal-Chaloin S, Drocourt L, Maurel P, Vilarem MJ. The expression of CYP2B6, CYP2C9 and CYP3A4 genes: a tangle of networks of nuclear and steroid receptors. Biochim Biophys Acta 2003; 1619:243–253.
18. Wang H, Faucette S, Sueyoshi T, Moore R, Ferguson S, Negishi M, LeCluyse EL. A novel distal enhancer module regulated by pregnane X receptor/constitutive androstane receptor is essential for the maximal induction of CYP2B6 gene expression. J Biol Chem 2003; 278:14146–14152.
19. Faucette SR, Wang H, Hamilton GA, Jolley SL, Gilbert D, Lindley C, Yan B, Negishi M, LeCluyse EL. Regulation of CYP2B6 in primary human hepatocytes by prototypical inducers. Drug Metab Dispos 2004; 32:348–358.
20. Riddick DS, Lee C, Bhathena A, Timsit YE, Cheng PY, Morgan ET, Prough RA, Ripp SL, Miller KK, Jahan A, Chiang JY. Transcriptional suppression of cytochrome P450 genes by endogenous and exogenous chemicals. Drug Metab Dispos 2004; 32:367–375.
21. Riddick DS, Lee C, Bhathena A, Timsit YE. The 2001 Veylien Henderson Award of the Society of Toxicology of Canada. Positive and negative transcriptional regulation of cytochromes P450 by polycyclic aromatic hydrocarbons. Can J Physiol Pharmacol 2003; 81:59–77.
22. Kliewer SA, Lehmann JM, Willson TM. Orphan nuclear receptors: shifting endocrinology into reverse. Science 1999; 284:757–760.
23. Waxman DJ. P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR. Arch Biochem Biophys 1999; 369:11–23.
24. Handschin C, Meyer UA. Induction of drug metabolism: the role of nuclear receptors. Pharmacol Rev 2003; 55:649–673.
25. Tamasi V, Vereczkey L, Falus A, Monostory K. Some aspects of interindividual variations in the metabolism of xenobiotics. Inflamm Res 2003; 52: 322–333.

26. Kellermann G, Shaw CR, Luyten-Kellerman M. Aryl hydrocarbon hydroxylase inducibility and bronchogenic carcinoma. N Engl J Med 1973; 289:934–937.
27. Nebert DW, Eisen HJ, Negishi M, Lang MA, Hjelmeland LM, Okey AB. Genetic mechanisms controlling the induction of polysubstrate monooxygenase (P-450) activities. Annu Rev Pharmacol Toxicol 1981; 21:431–462.
28. Li W, Harper PA, Tang BK, Okey AB. Regulation of cytochrome P450 enzymes by aryl hydrocarbon receptor in human cells: CYP1A2 expression in the LS180 colon carcinoma cell line after treatment with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin or 3-methylcholanthrene. Biochem Pharmacol 1998; 56:599–612.
29. Drocourt L, Ourlin JC, Pascussi JM, Maurel P, Vilarem MJ. Expression of CYP3A4, CYP2B6, and CYP2C9 is regulated by the vitamin D receptor pathway in primary human hepatocytes. J Biol Chem 2002; 277:25125–25132.
30. Kalow W, Tang BK. Use of caffeine metabolite ratios to explore CYP1A2 and xanthine oxidase activities. Clin Pharmacol Ther 1991; 50:508–519.
31. Butler MA, Lang NP, Young JF, Caporaso NE, Vineis P, Hayes RB, Teitel CH, Massengill JP, Lawsen MF, Kadlubar FF. Determination of CYP1A2 and NAT2 phenotypes in human populations by analysis of caffeine urinary metabolites. Pharmacogenetics 1992; 2:116–127.
32. Manchester DK. Range of environmentally responsive monooxygenase activities in human placental microsomes determined by direct fluorescence techniques. Biochem Pharmacol 1981; 30:757–762.
33. Manchester DK, Gordon SK, Golas CL, Roberts EA, Okey AB. Ah receptor in human placenta: stabilization by molybdate and characterization of binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin, 3-methylcholanthrene, and benzo(a)-pyrene. Cancer Res 1987; 47:4861–4868.
34. Michnovicz JJ, Adlercreutz H, Bradlow HL. Changes in levels of urinary estrogen metabolites after oral indole-3-carbinol treatment in humans. J Natl Cancer Inst 1997; 89:718–723.
35. Michnovicz JJ, Bradlow HL. Induction of estradiol metabolism by dietary indole-3-carbinol in humans. J Natl Cancer Inst 1990; 82:947–949.
36. Michnovicz JJ, Bradlow HL. Altered estrogen metabolism and excretion in humans following consumption of indole-3-carbinol. Nutr Cancer 1991; 16:59–66.
37. MacLeod S, Sinha R, Kadlubar FF, Lang NP. Polymorphisms of CYP1A1 and GSTM1 influence the in vivo function of CYP1A2. Mutat Res 1997; 376:135–142.
38. Murray S, Lake BG, Gray S, Edwards AJ, Springall C, Bowey EA, Williamson G, Boobis AR, Gooderham NJ. Effect of cruciferous vegetable consumption on heterocyclic aromatic amine metabolism in man. Carcinogenesis 2001; 22: 1413–1420.
39. de Morais SM, Wilkinson GR, Blaisdell J, Nakamura K, Meyer UA, Goldstein JA. The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. J Biol Chem 1994; 269: 15419–15422.
40. Schwarz UI. Clinical relevance of genetic polymorphisms in the human CYP2C9 gene. Eur J Clin Invest 2003; 33(suppl 2):23–30.

41. Geyer HJ, Schramm KW, Feicht EA, Behechti A, Steinberg C, Bruggemann R, Poiger H, Henkelmann B, Kettrup A. Half-lives of tetra-, penta-, hexa-, hepta-, and octachlorodibenzo-*p*-dioxin in rats, monkeys, and humans—a critical review. Chemosphere 2002; 48:631–644.
42. Okey AB. Enzyme induction in the cytochrome P-450 system. Pharmacol Ther 1990; 45:241–298.
43. Toide K, Yamazaki H, Nagashima R, Itoh K, Iwano S, Takahashi Y, Watanabe S, Kamataki T. Aryl hydrocarbon hydroxylase represents CYP1B1, and not CYP1A1, in human freshly isolated white cells: trimodal distribution of Japanese population according to induction of CYP1B1 mRNA by environmental dioxins. Cancer Epidemiol Biomarkers Prev 2003; 12:219–222.
44. Ueda A, Hamadeh HK, Webb HK, Yamamoto Y, Sueyoshi T, Afshari CA, Lehmann JM, Negishi M. Diverse roles of the nuclear orphan receptor CAR in regulating hepatic genes in response to phenobarbital. Mol Pharmacol 2002; 61:1–6.
45. Puga A, Xia Y, Elferink C. Role of the aryl hydrocarbon receptor in cell cycle regulation. Chem Biol Interact 2002; 141:117–130.
46. Harper PA, Wong JY, Lam MS, Okey AB. Polymorphisms in the human AH receptor. Chem Biol Interact 2002; 141:161–187.
47. Quattrochi LC, Guzelian PS. Cyp3A regulation: from pharmacology to nuclear receptors. Drug Metab Dispos 2001; 29:615–622.
48. Tirona RG, Leake BF, Podust LM, Kim RB. Identification of amino acids in rat pregnane X receptor that determine species-specific activation. Mol Pharmacol 2004; 65:36–44.
49. Mimura J, Ema M, Sogawa K, Fujii-Kuriyama Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. Genes Dev 1999; 13:20–25.
50. Baba T, Mimura J, Gradin K, Kuroiwa A, Watanabe T, Matsuda Y, Inazawa J, Sogawa K, Fujii-Kuriyama Y. Structure and expression of the Ah receptor repressor gene. J Biol Chem 2001; 276:33101–33110.
51. Tsuchiya Y, Nakajima M, Itoh S, Iwanari M, Yokoi T. Expression of aryl hydrocarbon receptor repressor in normal human tissues and inducibility by polycyclic aromatic hydrocarbons in human tumor-derived cell lines. Toxicol Sci 2003; 72:253–259.
52. Yamamoto J, Ihara K, Nakayama H, Hikino S, Satoh K, Kubo N, Iida T, Fujii Y, Hara T. Characteristic expression of aryl hydrocarbon receptor repressor gene in human tissues: organ-specific distribution and variable induction patterns in mononuclear cells. Life Sci 2004; 74:1039–1049.
53. Watanabe T, Imoto I, Kosugi Y, Fukuda Y, Mimura J, Fujii Y, Isaka K, Takayama M, Sato A, Inazawa J. Human arylhydrocarbon receptor repressor (AHRR) gene: genomic structure and analysis of polymorphism in endometriosis. J Hum Genet 2001; 46:342–346.
54. Fujita H, Kosaki R, Yoshihashi H, Ogata T, Tomita M, Hasegawa T, Takahashi T, Matsuo N, Kosaki K. Characterization of the aryl hydrocarbon receptor repressor gene and association of its Pro185Ala polymorphism with micropenis. Teratology 2002; 65:10–18.

55. Park EY, Cho IJ, Kim SG. Transactivation of the PPAR-responsive enhancer module in chemopreventive glutathione S-transferase gene by the peroxisome proliferator-activated receptor-gamma and retinoid X receptor heterodimer. Cancer Res 2004; 64:3701–3713.
56. Prokipcak RD, Okey AB. Downregulation of the Ah receptor in mouse hepatoma cells treated in culture with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. Can J Physiol Pharmacol 1991; 69:1204–1210.
57. Pollenz RS, Santostefano MJ, Klett E, Richardson VM, Necela B, Birnbaum LS. Female Sprague-Dawley rats exposed to a single oral dose of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin exhibit sustained depletion of aryl hydrocarbon receptor protein in liver, spleen, thymus, and lung. Toxicol Sci 1998; 42:117–128.
58. Roberts BJ, Whitelaw ML. Degradation of the basic helix-loop-helix/Per-ARNT-Sim homology domain dioxin receptor via the ubiquitin/proteasome pathway. J Biol Chem 1999; 274:36351–36356.
59. Pollenz RS. The mechanism of AH receptor protein down-regulation (degradation) and its impact on AH receptor-mediated gene regulation. Chem Biol Interact 2002; 141:41–61.
60. Franc MA, Pohjanvirta R, Tuomisto J, Okey AB. Persistent, low-dose 2,3,7,8-tetrachlorodibenzo-p-dioxin exposure: effect on aryl hydrocarbon receptor expression in a dioxin-resistance model. Toxicol Appl Pharmacol 2001; 175:43–53.
61. Wolff S, Harper PA, Wong JM, Mostert V, Wang Y, Abel J. Cell-specific regulation of human aryl hydrocarbon receptor expression by transforming growth factor-beta(1). Mol Pharmacol 2001; 59:716–724.
62. Chesire DR, Dunn TA, Ewing CM, Luo J, Isaacs WB. Identification of aryl hydrocarbon receptor as a putative Wnt/beta-catenin pathway target gene in prostate cancer cells. Cancer Res 2004; 64:2523–2533.
63. Racky J, Schmitz HJ, Kauffmann HM, Schrenk D. Single nucleotide polymorphism analysis and functional characterization of the human Ah receptor (AhR) gene promoter. Arch Biochem Biophys 2004; 421:91–98.
64. Hayashi S, Watanabe J, Nakachi K, Eguchi H, Gotoh O, Kawajiri K. Interindividual difference in expression of human Ah receptor and related P450 genes. Carcinogenesis 1994; 15:801–806.
65. Okey AB, Giannone JV, Smart W, Wong JM, Manchester DK, Parker NB, Feeley MM, Grant DL, Gilman A. Binding of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin to AH receptor in placentas from normal versus abnormal pregnancy outcomes. Chemosphere 1997; 34:1535–1547.
66. Lin P, Hu SW, Chang TH. Correlation between gene expression of Aryl hydrocarbon receptor (AhR), hydrocarbon receptor nuclear translocator (Arnt), cytochromes P4501A1 (CYP1A1) and 1B1 (CYP1B1), and inducibility of CYP1A1 and CYP1B1 in human lymphocytes. Toxicol Sci 2003; 71:20–26.
67. Wang H, Faucette SR, Gilbert D, Jolley SL, Sueyoshi T, Negishi M, LeCluyse EL. Glucocorticoid receptor enhancement of pregnane X receptor-mediated CYP2B6 regulation in primary human hepatocytes. Drug Metab Dispos 2003; 31:620–630.
68. Henderson CJ, Otto DM, Carrie D, Magnuson MA, McLaren AW, Rosewell I, Wolf CR. Inactivation of the hepatic cytochrome P450 system by conditional

deletion of hepatic cytochrome P450 reductase. J Biol Chem 2003; 278: 13480–13486.
69. Fan LQ, Coley J, Miller RT, Cattley RC, Corton JC. Opposing mechanisms of NADPH-cytochrome P450 oxidoreductase regulation by peroxisome proliferators. Biochem Pharmacol 2003; 65:949–959.
70. Renton KW. Alteration of drug biotransformation and elimination during infection and inflammation. Pharmacol Ther 2001; 92:147–163.
71. Morgan ET, Li-Masters T, Cheng PY. Mechanisms of cytochrome P450 regulation by inflammatory mediators. Toxicology 2002; 181–182:207–210.
72. Pascussi JM, Dvorak Z, Gerbal-Chaloin S, Assenat E, Maurel P, Vilarem MJ. Pathophysiological factors affecting CAR gene expression. Drug Metab Rev 2003; 35:255–268.
73. Fang C, Yoon S, Tindberg N, Jarvelainen HA, Lindros KO, Ingelman-Sundberg M. Hepatic expression of multiple acute phase proteins and down-regulation of nuclear receptors after acute endotoxin exposure. Biochem Pharmacol 2004; 67:1389–1397.
74. Ding X, Kaminsky LS. Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. Annu Rev Pharmacol Toxicol 2003; 43:149–173.
75. Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem 2004; 279:23847–23850.
76. Spink DC, Katz BH, Hussain MM, Pentecost BT, Cao Z, Spink BC. Estrogen regulates Ah responsiveness in MCF-7 breast cancer cells. Carcinogenesis 2003; 24:1941–1950.
77. Smith G, Wolf CR, Deeni YY, Dawe RS, Evans AT, Comrie MM, Ferguson J, Ibbotson SH. Cutaneous expression of cytochrome P450 CYP2S1: individuality in regulation by therapeutic agents for psoriasis and other skin diseases. Lancet 2003; 361:1336–1343.
78. Rylander T, Neve EP, Ingelman-Sundberg M, Oscarson M. Identification and tissue distribution of the novel human cytochrome P450 2S1 (CYP2S1). Biochem Biophys Res Commun 2001; 281:529–535.
79. Okey AB, Franc MA, Moffat ID, Tijet N, Boutros PC, Korkalainen M, Tuomisto J, Pohjanvirta R. Toxicological implications of drug receptor polymorphisms: the case of the aryl hydrocarbon receptor. Toxicol Appl Pharmacol. In press.
80. Wong JM, Okey AB, Harper PA. Human aryl hydrocarbon receptor polymorphisms that result in loss of CYP1A1 induction. Biochem Biophys Res Commun 2001; 288:990–996.
81. Kawajiri K, Watanabe J, Eguchi H, Nakachi K, Kiyohara C, Hayashi S. Polymorphisms of human Ah receptor gene are not involved in lung cancer. Pharmacogenetics 1995; 5:151–158.
82. Smart J, Daly AK. Variation in induced CYP1A1 levels: relationship to CYP1A1, Ah receptor and GSTM1 polymorphisms. Pharmacogenetics 2000; 10:11–24.
83. Nebert DW. The Ah locus: genetic differences in toxicity, cancer, mutation, and birth defects. Crit Rev Toxicol 1989; 20:153–174.

84. Pohjanvirta R, Tuomisto J. Short-term toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in laboratory animals: effects, mechanisms, and animal models. Pharmacol Rev 1994; 46:483–549.
85. Okey AB, Riddick DS, Harper PA. The Ah receptor: mediator of the toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and related compounds. Toxicol Lett 1994; 70:1–22.
86. Nebert DW, Winker J, Gelboin HV. Aryl hydrocarbon hydroxylase activity in human placenta from cigarette smoking and nonsmoking women. Cancer Res 1969; 29:1763–1769.
87. Gottlieb K, Manchester DK. Twin study methodology and variability in xenobiotic placental metabolism. Teratog Carcinog Mutagen 1986; 6:253–263.
88. McLemore TL, Adelberg S, Liu MC, McMahon NA, Yu SJ, Hubbard WC, Czerwinski M, Wood TG, Storeng R, Lubet RA, Eggleston JC, Boyd MR, Hines RN. Expression of CYP1A1 gene in patients with lung cancer: evidence for cigarette smoke-induced gene expression in normal lung tissue and for altered gene regulation in primary pulmonary carcinomas. J Natl Cancer Inst 1990; 82:1333–1339.
89. Pasquini R, Sforzolini GS, Cavaliere A, Savino A, Monarca S, Puccetti P, Fatigoni C, Antonini G. Enzymatic activities of human lung tissue: relationship with smoking habits. Carcinogenesis 1988; 9:1411–1416.
90. Petruzzelli S, Camus AM, Carrozzi L, Ghelarducci L, Rindi M, Menconi G, Angeletti CA, Ahotupa M, Hietanen E, Aitio A. Long-lasting effects of tobacco smoking on pulmonary drug-metabolizing enzymes: a case-control study on lung cancer patients. Cancer Res 1988; 48:4695–4700.
91. Anttila S, Hakkola J, Tuominen P, Elovaara E, Husgafvel-Pursiainen K, Karjalainen A, Hirvonen A, Nurminen T. Methylation of cytochrome P4501A1 promoter in the lung is associated with tobacco smoking. Cancer Res 2003; 63:8623–8628.
92. Kalow W, Tang BK. Caffeine as a metabolic probe: exploration of the enzyme-inducing effect of cigarette smoking. Clin Pharmacol Ther 1991; 49:44–48.
93. Welfare MR, Aitkin M, Bassendine MF, Daly AK. Detailed modelling of caffeine metabolism and examination of the CYP1A2 gene: lack of a polymorphism in CYP1A2 in Caucasians [published erratum appears in Pharmacogenetics 1999 Dec; 9(6):782]. Pharmacogenetics 1999; 9:367–375.
94. Ozdemir V, Kalow W, Posner P, Collins EJ, Kennedy JL, Tang BK, Albers LJ, Reist C, Roy R, Walkes W, Afra P. CYP1A2 activity as measured by a caffeine test predicts clozapine and active metabolite steady-state concentrationin patients with schizophrenia. J Clin Psychopharmacol 2001; 21: 398–407.
95. Ozdemir V, Kalow W, Okey AB, Lam MS, Albers LJ, Reist C, Fourie J, Posner P, Collins EJ, Roy R. Treatment-resistance to clozapine in association with ultrarapid CYP1A2 activity and the C→A polymorphism in intron 1 of the CYP1A2 gene: effect of grapefruit juice and low-dose fluvoxamine. J Clin Psychopharmacol 2001; 21:603–607.

96. Eap CB, Bender S, Sirot EJ, Cucchia G, Jonzier-Perey M, Baumann P, Allorge D, Broly F. Nonresponse to clozapine and ultrarapid CYP1A2 activity: clinical data and analysis of CYP1A2 gene. J Clin Psychopharmacol 2004; 24:214–219.
97. Sachse C, Brockmoller J, Bauer S, Roots I. Functional significance of a C→A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. Br J Clin Pharmacol 1999; 47:445–449.
98. Han XM, Ouyang DS, Chen XP, Shu Y, Jiang CH, Tan ZR, Zhou HH. Inducibility of CYP1A2 by omeprazole in vivo related to the genetic polymorphism of CYP1A2. Br J Clin Pharmacol 2002; 54:540–543.
99. Sachse C, Bhambra U, Smith G, Lightfoot TJ, Barrett JH, Scollay J, Garner RC, Boobis AR, Wolf CR, Gooderham NJ. Polymorphisms in the cytochrome P450 CYP1A2 gene (CYP1A2) in colorectal cancer patients and controls: allele frequencies, linkage disequilibrium and influence on caffeine metabolism. Br J Clin Pharmacol 2003; 55:68–76.
100. Aklillu E, Carrillo JA, Makonnen E, Hellman K, Pitarque M, Bertilsson L, Ingelman-Sundberg M. Genetic polymorphism of CYP1A2 in Ethiopians affecting induction and expression: characterization of novel haplotypes with single-nucleotide polymorphisms in intron 1. Mol Pharmacol 2003; 64: 659–669.
101. Smith GB, Harper PA, Wong JM, Lam MS, Reid KR, Petsikas D, Massey TE. Human lung microsomal cytochrome P4501A1 (CYP1A1) activities: impact of smoking status and CYP1A1, aryl hydrocarbon receptor, and glutathione S-transferase M1 genetic polymorphisms. Cancer Epidemiol Biomarkers Prev 2001; 10:839–853.
102. Scheel J, Hussong R, Schrenk D, Schmitz HJ. Variability of the human aryl hydrocarbon receptor nuclear translocator (ARNT) gene. J Hum Genet 2002; 47:217–224.
103. Rost KL, Brosicke H, Brockmoller J, Scheffler M, Helge H, Roots I. Increase of cytochrome P450IA2 activity by omeprazole: evidence by the 13C-[*N*-3-methyl]-caffeine breath test in poor and extensive metabolizers of S-mephenytoin. Clin Pharmacol Ther 1992; 52:170–180.
104. Quattrochi LC, Vu T, Tukey RH. The human CYP1A2 gene and induction by 3-methylcholanthrene. A region of DNA that supports AH-receptor binding and promoter-specific induction. J Biol Chem 1994; 269:6949–6954.
105. Sogawa K, Numayama-Tsuruta K, Takahashi T, Matsushita N, Miura C, Nikawa J, Gotoh O, Kikuchi Y, Fujii-Kuriyama Y. A novel induction mechanism of the rat CYP1A2 gene mediated by Ah receptor-Arnt heterodimer. Biochem Biophys Res Commun 2004; 318:746–755.
106. Boutros PC, Moffat ID, Franc MA, Tijet N, Tuomisto J, Pohjanvirta R, Okey AB. Dioxin-responsive AHRE-II gene battery: identification by phylogenetic footprinting. Biochem Biophys Res Commun 2004; 321:707–715.
107. Landi MT, Baccarelli A. Correspondence re: K. Toide, et al. Aryl hydrocarbon hydroxylase represents CYP1B1, and not CYP1A1, in human freshly isolated white cells: trimodal distribution of Japanese population according to induction of CYP1B1 mRNA by environmental dioxins [2003; 12: 219–222]. Cancer Epidemiol Biomarkers Prev 2003; 12:1116–1117 (author reply 2003; 1118).

108. Petersen DD, McKinney CE, Ikeya K, Smith HH, Bale AE, McBride OW, Nebert DW. Human CYP1A1 gene: cosegregation of the enzyme inducibility phenotype and an RFLP. Am J Hum Genet 1991; 48:720–725.
109. Rylander-Rudqvist T, Wedren S, Granath F, Humphreys K, Ahlberg S, Weiderpass E, Oscarson M, Ingelman-Sundberg M, Persson I. Cytochrome P450 1B1 gene polymorphisms and postmenopausal breast cancer risk. Carcinogenesis 2003; 24:1533–1539.
110. McGrath M, Hankinson SE, Arbeitman L, Colditz GA, Hunter DJ, De Vivo I. Cytochrome P450 1B1 and catechol-*O*-methyltransferase polymorphisms and endometrial cancer susceptibility. Carcinogenesis 2004; 25:559–565.
111. Mammen JS, Pittman GS, Li Y, Abou-Zahr F, Bejjani BA, Bell DA, Strickland PT, Sutter TR. Single amino acid mutations, but not common polymorphisms, decrease the activity of CYP1B1 against (−)benzo[a]pyrene-7R-*trans*-7,8-dihydrodiol. Carcinogenesis 2003; 24:1247–1255.
112. Sasaki M, Tanaka Y, Kaneuchi M, Sakuragi N, Dahiya R. Alleles of polymorphic sites that correspond to hyperactive variants of CYP1B1 protein are significantly less frequent in Japanese as compared to American and German populations. Hum Mutat 2003; 21:652.
113a. Han W, Pentecost BT, Spivack SD. Functional evaluation of novel single nucleotide polymorphisms and haplotypes in the promoter regions of CYP1B1 and CYP1A1 genes. Mol Carcinog 2003; 37:158–169.
113b. Widschwendter M, Siegmund KD, Muller HM, Fiegl H, Marth C, Muller-Holzner E, Jones PA, Laird PW. Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. Cancer Res 2004; 64:3807–3813.
114. Rivera SP, Saarikoski ST, Hankinson O. Identification of a novel dioxin-inducible cytochrome P450. Mol Pharmacol 2002; 61:255–259.
115. Chang TK, Bandiera SM, Chen J. Constitutive androstane receptor and pregnane X receptor gene expression in human liver: interindividual variability and correlation with CYP2B6 mRNA levels. Drug Metab Dispos 2003; 31: 7–10.
116. Hesse LM, He P, Krishnaswamy S, Hao Q, Hogan K, von Moltke LL, Greenblatt DJ, Court MH. Pharmacogenetic determinants of interindividual variability in bupropion hydroxylation by cytochrome P450 2B6 in human liver microsomes. Pharmacogenetics 2004; 14:225–238.
117. Gervot L, Rochat B, Gautier JC, Bohnenstengel F, Kroemer H, de Berardinis V, Martin H, Beaune P, de Waziers I. Human CYP2B6: expression, inducibility and catalytic activities. Pharmacogenetics 1999; 9:295–306.
118. Miksys S, Lerman C, Shields PG, Mash DC, Tyndale RF. Smoking, alcoholism and genetic polymorphisms alter CYP2B6 levels in human brain. Neuropharmacology 2003; 45:122–132.
119. Trottier E, Belzil A, Stoltz C, Anderson A. Localization of a phenobarbital-responsive element (PBRE) in the 5′-flanking region of the rat CYP2B2 gene. Gene 1995; 158:263–268.
120. Paquet Y, Trottier E, Beaudet MJ, Anderson A. Mutational analysis of the CYP2B2 phenobarbital response unit and inhibitory effect of the constitutive

androstane receptor on phenobarbital responsiveness. J Biol Chem 2000; 275:38427–38436.
121. Makinen J, Frank C, Jyrkkarinne J, Gynther J, Carlberg C, Honkakoski P. Modulation of mouse and human phenobarbital-responsive enhancer module by nuclear receptors. Mol Pharmacol 2002; 62:366–378.
122. Hustert E, Zibat A, Presecan-Siedel E, Eiselt R, Mueller R, Fuss C, Brehm I, Brinkmann U, Eichelbaum M, Wojnowski L, Burk O. Natural protein variants of pregnane X receptor with altered transactivation activity toward CYP3A4. Drug Metab Dispos 2001; 29:1454–1459.
123. Zhang J, Kuehl P, Green ED, Touchman JW, Watkins PB, Daly A, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Wrighton SA, Hancock M, Kim RB, Strom S, Thummel K, Russell CG, Hudson JR Jr, Schuetz EG, Boguski MS. The human pregnane X receptor: genomic structure and identification and functional characterization of natural allelic variants. Pharmacogenetics 2001; 11:555–572.
124. Lamba V, Lamba J, Yasuda K, Strom S, Davila J, Hancock ML, Fackenthal JD, Rogan PK, Ring B, Wrighton SA, Schuetz EG. Hepatic CYP2B6 expression: gender and ethnic differences and relationship to CYP2B6 genotype and CAR (constitutive androstane receptor) expression. J Pharmacol Exp Ther 2003; 307:906–922.
125. Lang T, Klein K, Fischer J, Nussler AK, Neuhaus P, Hofmann U, Eichelbaum M, Schwab M, Zanger UM. Extensive genetic polymorphism in the human CYP2B6 gene with impact on expression and function in human liver. Pharmacogenetics 2001; 11:399–415.
126. Lang T, Klein K, Richter T, Zibat A, Kerb R, Eichelbaum M, Schwab M, Zanger UM. Multiple novel nonsynonymous CYP2B6 gene polymorphisms in Caucasians: demonstration of phenotypic null alleles. J Pharmacol Exp Ther 2004; 311:34–43.
127. De Morais SM, Wilkinson GR, Blaisdell J, Meyer UA, Nakamura K, Goldstein JA. Identification of a new genetic defect responsible for the polymorphism of (S)-mephenytoin metabolism in Japanese. Mol Pharmacol 1994; 46:594–598.
128. Goldstein JA. Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. Br J Clin Pharmacol 2001; 52:349–355.
129. Chen Y, Ferguson SS, Negishi M, Goldstein JA. Induction of human CYP2C9 by rifampicin, hyperforin, and phenobarbital is mediated by the pregnane X receptor. J Pharmacol Exp Ther 2004; 308:495–501.
130. Henderson L, Yue QY, Bergquist C, Gerden B, Arlett P. St John's wort (Hypericum perforatum): drug interactions and clinical outcomes. Br J Clin Pharmacol 2002; 54:349–356.
131. Ferguson SS, LeCluyse EL, Negishi M, Goldstein JA. Regulation of human CYP2C9 by the constitutive androstane receptor: discovery of a new distal binding site. Mol Pharmacol 2002; 62:737–746.
132. Chen Y, Ferguson SS, Negishi M, Goldstein JA. Identification of constitutive androstane receptor and glucocorticoid receptor binding sites in the CYP2C19 promoter. Mol Pharmacol 2003; 64:316–324.

133. Hu Y, Ingelman-Sundberg M, Lindros KO. Induction mechanisms of cytochrome P450 2E1 in liver: interplay between ethanol treatment and starvation. Biochem Pharmacol 1995; 50:155–161.
134. Schoedel KA, Tyndale RF. Induction of nicotine-metabolizing CYP2B1 by ethanol and ethanol-metabolizing CYP2E1 by nicotine: summary and implications. Biochem Biophys Acta 2003; 1619:283–290.
135. Wrighton SA, Thomas PE, Molowa DT, Haniu M, Shively JE, Maines SL, Watkins PB, Parker G, Mendez-Picon G, Levin W, Guzelian PS. Characterization of ethanol-inducible human liver *N*-nitrosodimethylamine demethylase. Biochemistry 1986; 25:6731–6735.
136. Ekstrom G, von Bahr C, Ingelman-Sundberg M. Human liver microsomal cytochrome P-450IIE1. Immunological evaluation of its contribution to microsomal ethanol oxidation, carbon tetrachloride reduction and NADPH oxidase activity. Biochem Pharmacol 1989; 38:689–693.
137. Girre C, Lucas D, Hispard E, Menez C, Dally S, Menez JF. Assessment of cytochrome P4502E1 induction in alcoholic patients by chlorzoxazone pharmacokinetics. Biochem Pharmacol 1994; 47:1503–1508.
138. McCarver DG, Byun R, Hines RN, Hichme M, Wegenek W. A genetic polymorphism in the regulatory sequences of human CYP2E1: association with increased chlorzoxazone hydroxylation in the presence of obesity and ethanol intake. Toxicol Appl Pharmacol 1998; 152:276–281.
139. Hu Y, Hakkola J, Oscarson M, Ingelman-Sundberg M. Structural and functional characterization of the 5′-flanking region of the rat and human cytochrome P450 2E1 genes: identification of a polymorphic repeat in the human gene. Biochem Biophys Res Commun 1999; 263:286–293.
140. Howard LA, Ahluwalia JS, Lin SK, Sellers EM, Tyndale RF. CYP2E1∗1D regulatory polymorphism: association with alcohol and nicotine dependence. Pharmacogenetics 2003; 13:321–328.
141. Nebert DW, Russell DW. Clinical importance of the cytochromes P450. Lancet 2002; 360:1155–1162.
142. Plant NJ, Gibson GG. Evaluation of the toxicological relevance of CYP3A4 induction. Curr Opin Drug Discov Dev 2003; 6:50–56.
143. Wilkinson GR. Cytochrome P4503A (CYP3A) metabolism: prediction of in vivo activity in humans. J Pharmacokinet Biopharm 1996; 24:475–490.
144. Ozdemir V, Kalow W, Tang BK, Paterson AD, Walker SE, Endrenyi L, Kashuba AD. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. Pharmacogenetics 2000; 10:373–388.
145. Westlind A, Lofberg L, Tindberg N, Andersson TB, Ingelman-Sundberg M. Interindividual differences in hepatic expression of CYP3A4: relationship to genetic polymorphism in the 5′-upstream regulatory region. Biochem Biophys Res Commun 1999; 259:201–205.
146. Sy SK, Ciaccia A, Li W, Roberts EA, Okey A, Kalow W, Tang BK. Modeling of human hepatic CYP3A4 enzyme kinetics, protein, and mRNA indicates deviation from log-normal distribution in CYP3A4 gene expression. Eur J Clin Pharmacol 2002; 58:357–365.

147. Matsumura K, Saito T, Takahashi Y, Ozeki T, Kiyotani K, Fujieda M, Yamazaki H, Kunitoh H, Kamataki T. Identification of a novel polymorphic enhancer of the human CYP3A4 gene. Mol Pharmacol 2004; 65:326–334.
148. Moore LB, Goodwin B, Jones SA, Wisely GB, Serabjit-Singh CJ, Willson TM, Collins JL, Kliewer SA. St. John's wort induces hepatic drug metabolism through activation of the pregnane X receptor. Proc Natl Acad Sci USA 2000; 97:7500–7502.
149. Dresser GK, Bailey DG. The effects of fruit juices on drug disposition: a new model for drug interactions. Eur J Clin Invest 2003; 33(suppl 2):10–16.
150. Moore LB, Parks DJ, Jones SA, Bledsoe RK, Consler TG, Stimmel JB, Goodwin B, Liddle C, Blanchard SG, Willson TM, Collins JL, Kliewer SA. Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. J Biol Chem 2000; 275:15122–15127.
151. Honkakoski P, Sueyoshi T, Negishi M. Drug-activated nuclear receptors CAR and PXR. Ann Med 2003; 35:172–182.
152. Westlind-Johnsson A, Malmebo S, Johansson A, Otter C, Andersson TB, Johansson I, Edwards RJ, Boobis AR, Ingelman-Sundberg M. Comparative analysis of CYP3A expression in human liver suggests only a minor role for CYP3A5 in drug metabolism. Drug Metab Dispos 2003; 31:755–761.
153. Tirona RG, Lee W, Leake BF, Lan LB, Cline CB, Lamba V, Parviz F, Duncan SA, Inoue Y, Gonzalez FJ, Schuetz EG, Kim RB. The orphan nuclear receptor HNF4alpha determines PXR- and CAR-mediated xenobiotic induction of CYP3A4. Nat Med 2003; 9:220–224.
154. Hashimoto H, Toide K, Kitamura R, Fujita M, Tagawa S, Itoh S, Kamataki T. Gene structure of CYP3A4, an adult-specific form of cytochrome P450 in human livers, and its transcriptional control. Eur J Biochem 1993; 218: 585–595.
155. Amirimani B, Ning B, Deitz AC, Weber BL, Kadlubar FF, Rebbeck TR. Increased transcriptional activity of the CYP3A4*1B promoter variant. Environ Mol Mutagen 2003; 42:299–305.
156. Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, Watkins PB, Daly A, Wrighton SA, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Venkataramanan R, Strom S, Thummel K, Boguski MS, Schuetz E. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. Nat Genet 2001; 27:383–391.
157. Hustert E, Haberl M, Burk O, Wolbold R, He YQ, Klein K, Nuessler AC, Neuhaus P, Klattig J, Eiselt R, Koch I, Zibat A, Brockmoller J, Halpert JR, Zanger UM, Wojnowski L. The genetic determinants of the CYP3A5 polymorphism. Pharmacogenetics 2001; 11:773–779.
158. Burk O, Tegude H, Koch I, Hustert E, Wolbold R, Glaeser H, Klein K, Fromm MF, Nuessler AK, Neuhaus P, Zanger UM, Eichelbaum M, Wojnowski L. Molecular mechanisms of polymorphic CYP3A7 expression in adult human liver and intestine. J Biol Chem 2002; 277:24280–24288.
159. Yueh MF, Huang YH, Hiller A, Chen S, Nguyen N, Tukey RH. Involvement of the xenobiotic response element (XRE) in Ah receptor-mediated induction of human UDP-glucuronosyltransferase 1A1. J Biol Chem 2003; 278: 15001–15006.

160. Sugatani J, Yamakawa K, Tonda E, Nishitani S, Yoshinari K, Degawa M, Abe I, Noguchi H, Miwa M. The induction of human UDP-glucuronosyltransferase 1A1 mediated through a distal enhancer module by flavonoids and xenobiotics. Biochem Pharmacol 2004; 67:989–1000.
161. Miners JO, Smith PA, Sorich MJ, McKinnon RA, Mackenzie PI. Predicting human drug glucuronidation parameters: application of in vitro and in silico modeling approaches. Annu Rev Pharmacol Toxicol 2004; 44:1–25.
162. Bosma PJ, Chowdhury JR, Bakker C, Gantla S, de Boer A, Oostra BA, Lindhout D, Tytgat GN, Jansen PL, Oude Elferink RP, Chowdhury NR. The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N Engl J Med 1995; 333:1171–1175.
163. Kohle C, Mohrle B, Munzel PA, Schwab M, Wernet D, Badary OA, Bock KW. Frequent co-occurrence of the TATA box mutation associated with Gilbert's syndrome (UGT1A1*28) with other polymorphisms of the UDP-glucuronosyltransferase-1 locus (UGT1A6*2 and UGT1A7*3) in Caucasians and Egyptians. Biochem Pharmacol 2003; 65:1521–1527.
164. Duguay Y, McGrath M, Lepine J, Gagne JF, Hankinson SE, Colditz GA, Hunter DJ, Plante M, Tetu B, Belanger A, Guillemette C, De Vivo I. The functional UGT1A1 promoter polymorphism decreases endometrial cancer risk. Cancer Res 2004; 64:1202–1207.
165. Coles BF, Morel F, Rauch C, Huber WW, Yang M, Teitel CH, Green B, Lang NP, Kadlubar FF. Effect of polymorphism in the human glutathione S-transferase A1 promoter on hepatic GSTA1 and GSTA2 expression. Pharmacogenetics 2001; 11:663–669.
166. Bredschneider M, Klein K, Murdter TE, Marx C, Eichelbaum M, Nussler AK, Neuhaus P, Zanger UM, Schwab M. Genetic polymorphisms of glutathione S-transferase A1, the major glutathione S-transferase in human liver: consequences for enzyme expression and busulfan conjugation. Clin Pharmacol Ther 2002; 71:479–487.
167. Ning B, Wang C, Morel F, Nowell S, Ratnasinghe DL, Carter W, Kadlubar FF, Coles B. Human glutathione S-transferase A2 polymorphisms: variant expression, distribution in prostate cancer cases/controls and a novel form. Pharmacogenetics 2004; 14:35–44.
168. Guy CA, Hoogendoorn B, Smith SK, Coleman S, O'Donovan MC, Buckland PR. Promoter polymorphisms in glutathione-S-transferase genes affect transcription. Pharmacogenetics 2004; 14:45–51.
169. Schlager JJ, Powis G. Cytosolic NAD(P)H:(quinone-acceptor)oxidoreductase in human normal and tumor tissue: effects of cigarette smoking and alcohol. Int J Cancer 1990; 45:403–409.
170. Rauth AM, Goldberg Z, Misra V. DT-diaphorase: possible roles in cancer chemotherapy and carcinogenesis. Oncol Res 1997; 9:339–349.
171. Belinsky M, Jaiswal AK. NAD(P)H:quinone oxidoreductase1 (DT-diaphorase) expression in normal and tumor tissues. Cancer Metastasis Rev 1993; 12: 103–117.
172. Smitskamp-Wilms E, Giaccone G, Pinedo HM, van der Laan BF, Peters GJ. DT-diaphorase activity in normal and neoplastic human tissues; an indicator for sensitivity to bioreductive agents?. Br J Cancer 1995; 72:917–921.

173. Marin A, Lopez de Cerain A, Hamilton E, Lewis AD, Martinez-Penuela JM, Idoate MA, Bello J. DT-diaphorase and cytochrome B5 reductase in human lung and breast tumours. Br J Cancer 1997; 76:923–929.
174. Sharp SY, Kelland LR, Valenti MR, Brunton LA, Hobbs S, Workman P. Establishment of an isogenic human colon tumor model for NQO1 gene expression: application to investigate the role of DT-diaphorase in bioreductive drug activation in vitro and in vivo. Mol Pharmacol 2000; 58: 1146–1155.
175. Noda S, Harada N, Hida A, Fujii-Kuriyama Y, Motohashi H, Yamamoto M. Gene expression of detoxifying enzymes in AhR and Nrf2 compound null mutant mouse. Biochem Biophys Res Commun 2003; 303:105–111.
176. Ma Q, Kinneer K, Bi Y, Chan JY, Kan YW. Induction of murine NAD(P)H: quinone oxidoreductase by 2,3,7,8-tetrachlorodibenzo-p-dioxin requires the CNC (cap 'n' collar) basic leucine zipper transcription factor Nrf2 (nuclear factor erythroid 2-related factor 2): cross-interaction between AhR (aryl hydrocarbon receptor) and Nrf2 signal transduction. Biochem J 2004; 377:205–213.
177. Nebert DW, Roe AL, Vandale SE, Bingham E, Oakley GG. NAD(P)H: quinone oxidoreductase (NQO1) polymorphism, exposure to benzene, and predisposition to disease: a HUGE review. Genet Med 2002; 4:62–70.
178. Ruschitzka F, Meier PJ, Turina M, Luscher TF, Noll G. Acute heart transplant rejection due to Saint John's wort. Lancet 2000; 355:548–549.
179. Piscitelli SC, Burstein AH, Chaitt D, Alfaro RM, Falloon J. Indinavir concentrations and St John's wort. Lancet 2000; 355:547–548.
180. Fuhr U. Induction of drug metabolising enzymes: pharmacokinetic and toxicological consequences in humans. Clin Pharmacokinet 2000; 38: 493–504.
181. Kensler TW, Curphey TJ, Maxiutenko Y, Roebuck BD. Chemoprotection by organosulfur inducers of phase 2 enzymes: dithiolethiones and dithiins. Drug Metab Drug Interact 2000; 17:3–22.
182. Gottlieb B, Pinsky L, Beitel LK, Trifiro M. Androgen insensitivity. Am J Med Genet 1999; 89:210–217.
183. Zheng SL, Zheng W, Chang BL, Shu XO, Cai Q, Yu H, Dai Q, Xu J, Gao YT. Joint effect of estrogen receptor beta sequence variants and endogenous estrogen exposure on breast cancer risk in Chinese women. Cancer Res 2003; 63:7624–7629.
184. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. Nature 2004; 429:464–468.
185. Gashaw I, Kirchheiner J, Goldammer M, Bauer S, Seidemann J, Zoller K, Mrozikiewicz PM, Roots I, Brockmoller J. Cytochrome p450 3A4 messenger ribonucleic acid induction by rifampicin in human peripheral blood mononuclear cells: correlation with alprazolam pharmacokinetics. Clin Pharmacol Ther 2003; 74:448–457.
186. Xie W, Barwick JL, Downes M, Blumberg B, Simon CM, Nelson MC, Neuschwander-Tetri BA, Brunt EM, Guzelian PS, Evans RM. Humanized xenobiotic response in mice expressing nuclear receptor SXR. Nature 2000; 406:435–439.

187. Rosenfeld JM, Vargas R Jr, Xie W, Evans RM. Genetic profiling defines the xenobiotic gene network controlled by the nuclear receptor pregnane X receptor. Mol Endocrinol 2003; 17:1268–1282.
188. Moriguchi T, Motohashi H, Hosoya T, Nakajima O, Takahashi S, Ohsako S, Aoki Y, Nishimura N, Tohyama C, Fujii-Kuriyama Y, Yamamoto M. Distinct response to dioxin in an arylhydrocarbon receptor (AHR)-humanized mouse. Proc Natl Acad Sci USA 2003; 100:5652–5657.
189. Cheung C, Akiyama TE, Ward JM, Nicol CJ, Feigenbaum L, Vinson C, Gonzalez FJ. Diminished hepatocellular proliferation in mice humanized for the nuclear receptor peroxisome proliferator-activated receptor alpha. Cancer Res 2004; 64:3849–3854.
190. Robertson GR, Field J, Goodwin B, Bierach S, Tran M, Lehnert A, Liddle C. Transgenic mouse models of human CYP3A4 gene regulation. Mol Pharmacol 2003; 64:42–50.
191. Saini SP, Sonoda J, Xu L, Toma D, Uppal H, Mu Y, Ren S, Moore DD, Evans RM, Xie W. A novel constitutive androstane receptor-mediated and CYP3A-independent pathway of bile acid detoxification. Mol Pharmacol 2004; 65:292–300.
192. Repa JJ, Mangelsdorf DJ. The liver X receptor gene team: potential new players in atherosclerosis. Nat Med 2002; 8:1243–1248.
193. Lu TT, Repa JJ, Mangelsdorf DJ. Orphan nuclear receptors as eLiXiRs and FiXeRs of sterol metabolism. J Biol Chem 2001; 276:37735–37738.
194. Handschin C, Podvinec M, Amherd R, Looser R, Ourlin JC, Meyer UA. Cholesterol and bile acids regulate xenosensor signaling in drug-mediated induction of cytochromes P450. J Biol Chem 2002; 277:29561–29567.
195. Savas U, Hsu MH, Johnson EF. Differential regulation of human CYP4A genes by peroxisome proliferators and dexamethasone. Arch Biochem Biophys 2003; 409:212–220.
196. Bertilsson G, Heidrich J, Svensson K, Asman M, Jendeberg L, Sydow-Backman M, Ohlsson R, Postlind H, Blomquist P, Berkenstam A. Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. Proc Natl Acad Sci USA 1998; 95:12208–12213.
197. Kliewer SA, Moore JT, Wade L, Staudinger JL, Watson MA, Jones SA, McKee DD, Oliver BB, Willson TM, Zetterstrom RH, Perlmann T, Lehmann JM. An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. Cell 1998; 92:73–82.
198. Lehmann JM, McKee DD, Watson MA, Willson TM, Moore JT, Kliewer SA. The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. J Clin Invest 1998; 102:1016–1023.
199. Xie W, Yeuh MF, Radominska-Pandya A, Saini SP, Negishi Y, Bottroff BS, Cabrera GY, Tukey RH, Evans RM. Control of steroid, heme, and carcinogen metabolism by nuclear pregnane X receptor and constitutive androstane receptor. Proc Natl Acad Sci USA 2003; 100:4150–4155.
200. Schmiedlin-Ren P, Thummel KE, Fisher JM, Paine MF, Watkins PB. Induction of CYP3A4 by 1 alpha,25-dihydroxyvitamin D3 is human cell line-specific

and is unlikely to involve pregnane X receptor. Drug Metab Dispos 2001; 29:1446–1453.
201. Bock KW. Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. Biochem Pharmacol 2003; 66:691–696.
202. Munzel PA, Schmohl S, Buckler F, Jaehrling J, Raschko FT, Kohle C, Bock KW. Contribution of the Ah receptor to the phenolic antioxidant-mediated expression of human and rat UDP-glucuronosyltransferase UGT1A6 in Caco-2 and rat hepatoma 5L cells. Biochem Pharmacol 2003; 66:841–847.
203. Venugopal R, Jaiswal AK. Nrf2 and Nrf1 in association with Jun proteins regulate antioxidant response element-mediated expression and coordinated induction of genes encoding detoxifying enzymes. Oncogene 1998; 17: 3145–3156.
204. Zhang W, Purchio A, Chen K, Burns SM, Contag CH, Contag PR. In vivo activation of the human CYP3A4 promoter in mouse liver and regulation by pregnane X receptors. Biochem Pharmacol 2003; 65:1889–1896.
205. Schweikl H, Taylor JA, Kitareewan S, Linko P, Nagorney D, Goldstein JA. Expression of CYP1A1 and CYP1A2 genes in human liver. Pharmacogenetics 1993; 3:239–249.
206. Chang TK, Chen J, Pillay V, Ho JY, Bandiera SM. Real-time polymerase chain reaction analysis of CYP1B1 gene expression in human liver. Toxicol Sci 2003; 71:11–19.
207. Kim JH, Sherman ME, Curriero FC, Guengerich FP, Strickland PT, Sutter TR. Expression of cytochromes P450 1A1 and 1B1 in human lung from smokers, non-smokers, and ex-smokers. Toxicol Appl Pharmacol 2004; 199: 210–219.
208. Ou-Yang DS, Huang SL, Wang W, Xie HG, Xu ZH, Shu Y, Zhou HH. Phenotypic polymorphism and gender-related differences of CYP1A2 activity in a Chinese population. Br J Clin Pharmacol 2000; 49:145–151.
209. Kirchheiner J, Klein C, Meineke I, Sasse J, Zanger UM, Murdter TE, Roots I, Brockmoller J. Bupropion and 4-OH-bupropion pharmacokinetics in relation to genetic polymorphisms in CYP2B6. Pharmacogenetics 2003; 13:619–626.
210. Code EL, Crespi CL, Penman BW, Gonzalez FJ, Chang TK, Waxman DJ. Human cytochrome P4502B6: interindividual hepatic expression, substrate specificity, and role in procarcinogen activation. Drug Metab Dispos 1997; 25:985–993.
211. Stresser DM, Kupfer D. Monospecific antipeptide antibody to cytochrome P-450 2B6. Drug Metab Dispos 1999; 27:517–525.
212. Chainuvati S, Nafziger AN, Leeder JS, Gaedigk A, Kearns GL, Sellers E, Zhang Y, Kashuba AD, Rowland E, Bertino JS Jr. Combined phenotypic assessment of cytochrome p450 1A2, 2C9, 2C19, 2D6, and 3A, *N*-acetyltransferase-2, and xanthine oxidase activities with the "Cooperstown 5+1 cocktail." Clin Pharmacol Ther 2003; 74:437–447.
213. Lapple F, von Richter O, Fromm MF, Richter T, Thon KP, Wisser H, Griese EU, Eichelbaum M, Kivisto KT. Differential expression and function of CYP2C isoforms in human intestine and liver. Pharmacogenetics 2003; 13:565–575.
214. Christensen M, Andersson K, Dalen P, Mirghani RA, Muirhead GJ, Nordmark A, Tybring G, Wahlberg A, Yasar U, Bertilsson L. The Karolinska

cocktail for phenotyping of five human cytochrome P450 enzymes. Clin Pharmacol Ther 2003; 73:517–528.
215. Kim RB, O'Shea D. Interindividual variability of chlorzoxazone 6-hydroxylation in men and women and its relationship to CYP2E1 genetic polymorphisms. Clin Pharmacol Ther 1995; 57:645–655.
216. Schuetz JD, Beach DL, Guzelian PS. Selective expression of cytochrome P450 CYP3A mRNAs in embryonic and adult human liver. Pharmacogenetics 1994; 4:11–20.
217. Shimada T, Yamazaki H, Mimura M, Inui Y, Guengerich FP. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. J Pharmacol Exp Ther 1994; 270:414–423.
218. Koch I, Weil R, Wolbold R, Brockmoller J, Hustert E, Burk O, Nuessler A, Neuhaus P, Eichelbaum M, Zanger U, Wojnowski L. Interindividual variability and tissue-specificity in the expression of cytochrome P450 3A mRNA. Drug Metab Dispos 2002; 30:1108–1114.
219. Fisher MB, Vandenbranden M, Findlay K, Burchell B, Thummel KE, Hall SD, Wrighton SA. Tissue distribution and interindividual variation in human UDP-glucuronosyltransferase activity: relationship between UGT1A1 promoter genotype and variability in a liver bank. Pharmacogenetics 2000; 10: 727–739.
220. Court MH, Hao Q, Krishnaswamy S, Bekaii-Saab T, Al-Rohaimi A, Von Moltke LL, Greenblatt DJ. UDP-glucuronosyltransferase (UGT) 2B15 pharmacogenetics: UGT2B15 D85Y genotype and gender are major determinants of oxazepam glucuronidation by human liver. J Pharmacol Exp Ther 2004; 310:656–665.
221. Ketterer B, Meyer DJ, Lalor E, Johnson P, Guengerich FP, Distlerath LM, Reilly PE, Kadlubar FF, Flammang TJ, Yamazoe Y, Beaune PH. A comparison of levels of glutathione transferases, cytochromes P450 and acetyltransferases in human livers. Biochem Pharmacol 1991; 41:635–638.
222. Vanden Heuvel JP, Kreder D, Belda B, Hannon DB, Nugent CA, Burns KA, Taylor MJ. Comprehensive analysis of gene expression in rat and human hepatoma cells exposed to the peroxisome proliferator WY14, 643. Toxicol Appl Pharmacol 2003; 188:185–198.
223. Porter TD, Coon MJ. Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. J Biol Chem 1991; 266:13469–13472.
224. Giguere V. Orphan nuclear receptors: from gene to function. Endocr Rev 1999; 20:689–725.

7

Pharmacogenetics and Cardiac Ion Channels

Dan M. Roden

Oates Institute for Experimental Therapeutics, Vanderbilt University School of Medicine, Nashville, Tennessee, U.S.A.

The use of drugs to treat cardiac arrhythmias is characterized by highly variable efficacy and serious toxicity. Both of these are difficult to predict in an individual patient. Studies of the mechanisms underlying this variability and unpredictability in drug action have been important platforms for defining the role of genetics in drug action and have also pointed to general mechanisms in the initiation and maintenance of abnormal cardiac rhythms. In practical terms, the past decade has seen a decline in the use of antiarrhythmic drugs, in part because of their unpredictable efficacy and toxicity. As well, "non-pharmacologic" techniques, including ablation of abnormal tissues underlying arrhythmias and implantable defibrillators, have matured and are increasingly used.

A common feature of most antiarrhythmic drugs is that they were developed without a clear understanding of the cellular and molecular mechanisms underlying cardiac arrhythmias. As a result, the molecular targets with which they interact to suppress (or occasionally exacerbate) arrhythmias and cause other forms of toxicity are only now being defined. Virtually all antiarrhythmic drugs target membrane proteins, including adrenergic receptors and ion channels, structures that generate ion-specific permeation pathways (pores) in response to changes in their environment; common examples of such changes include alterations in transmembrane

voltage (generated by other ion channels) or the presence of ligands such as acetylcholine. Studies defining the genetics of variable antiarrhythmic drug responses have not only pointed to new disease mechanisms, but also to strategies for development of new drugs lacking serious adverse effects and targeting underlying disease processes that culminate in arrhythmias. This review will first summarize the impact of variants in genes determining drug disposition on the effects of therapies used in the treatment of arrhythmias. The role of genetics in determining variable pharmacodynamics will then be considered.

1. THE GENETICS OF ANTIARRHYTHMIC DRUG DISPOSITION

1.1. CYP3A

Therapeutic drug monitoring was developed to optimize treatment with drugs that have a narrow margin between dosage required to produce efficacy and those producing adverse effects. Examples of such drug classes to which this approach has been applied include antibiotics, anticonvulsants, and antiarrhythmics. Initial descriptions of variable drug disposition as a consequence of disease or of drug interactions generally did not include a clear understanding of the specific molecular mechanisms underlying such variability. Thus, for example, anticonvulsants were recognized 30 years ago to strikingly lower concentrations of quinidine (1) and to reduce its therapeutic efficacy. In contemporary terms, this interaction reflects induction of CYP3A4 by anticonvulsant drugs, likely acting through orphan nuclear hormone receptors (2). While functionally important polymorphisms in the coding region of CPY3A4 have not been identified, there are such variants in CYP3A5, a closely related enzyme with overlapping substrate specificity and a prominent role in drug disposition in intestine (3). Given this view, variability in CYP3A expression becomes a logical candidate mechanism for modulating antiarrhythmic drug concentrations and hence effects; such variability could arise from polymorphisms in the promoter region of the CYP3A complex or in the genes encoding nuclear hormone receptors that mediate enzyme and transporter expression. Specific studies addressing these possibilities have not, however, been conducted.

1.2. CYP2D6

CYP2D6 is responsible for the biotransformation of a number of drugs used to control cardiac arrhythmias, notably some beta-adrenergic receptor antagonists (metoprolol, timolol, and carvedilol) as well as the sodium channel blocking antiarrhythmics encainide (no longer marketed), propafenone, and flecainide (4). The CYP2D6 poor metabolizers (PMs) making up 7% of Caucasian and African-American populations (and rare in Asian

populations) display higher plasma concentrations and greater pharmacologic effects during treatment with CYP2D6 substrate beta-blockers (5,6). The molecular basis of variable CYP2D6 activity is discussed in detail elsewhere in this volume (Chapter 3). The "third-generation" beta-blocker carvedilol is increasingly used in the treatment of heart failure and may be especially effective because of some vasodilator activity, possibly attributable to alpha-blockade. The extent of alpha- and beta-blockade by this drug does appear to be at least in part CYP2D6-mediated (6), although clinical trial data demonstrating that variable outcomes during carvedilol therapy in heart failure can be related to this polymorphism have not yet been developed.

Propafenone is another example of a drug that exerts multiple actions, and displays genetically determined variability in its clinical effects (7). In vitro, propafenone not only blocks sodium channels but also exerts beta-blocking activity (8,9). Initial reports of clinically significant adverse effects (such as bradycardia or bronchospasm) due to beta-blockade during propafenone therapy emphasized the unpredictable nature of this adverse effect (10). However, when CYP2D6 phenotype is considered, it becomes clear that PM subjects are at greater risk than extensive metabolizers (EMs) for clinically significant beta-blockade during treatment with the drug since they develop much higher concentrations of the parent drug than do EMs (9). It is also possible that consistent incorporation of beta-blockade into an antiarrhythmic molecule might improve its efficacy. Indeed, combining propafenone with low-dose quinidine (a CYP2D6 inhibitor) can result in phenocopying to the PM phenotype, and small trials have been undertaken to test the antiarrhythmic effects of the combination (11).

The antiarrhythmic drug encainide is biotransformed by CYP2D6 to a potent active metabolite, O-desmethyl encainide. Interestingly, this active metabolite, in turn, also undergoes CYP2D6-mediated biotransformation to a second active metabolite, 3-methoxy O-desmethyl encainide (12). Thus, variable CYP2D6 activity is a likely contributor to variable levels of the parent drug and its two active metabolites, and this may underlie some of the variability observed when the drug was used clinically. Flecainide, a compound with similar electrophysiologic properties, is bio-transformed to inactive metabolites by CYP2D6 and also undergoes renal excretion of parent drug. Thus, CYP2D6 phenotype has much less impact on flecainide plasma concentrations and effect, except in the rare patient with coexisting renal dysfunction (13).

1.3. CYP2C9

The anticoagulant warfarin is increasingly used to prevent thromboembolic complications in patients with atrial fibrillation. The drug is administered as racemate, and bio-inactivation of the active S-enantiomer is accomplished by CYP2C9. Relatively common variants in CYP2C9 that reduce its function have been described, and homozygotes for reduction of function alleles

appear to be at increased risk for bleeding complications with the drug (14–16).

1.4. *N*-Acetyltransferase

Plasma concentration monitoring of procainamide and its metabolite, *N*-acetylprocainamide (NAPA), led to the recognition that acetylator status is the major determinant of the development of anti-nuclear antibodies and the drug-induced lupus syndrome during procainamide therapy (17). This finding, in turn, pointed to the metabolism of procainamide by non-acetyl transferase pathways as an important modulator of this form of drug toxicity and also formed the rational for considering NAPA as a potentially less toxic antiarrhythmic entity (18). NAPA was found to exert electrophysiologic activity in vitro and to suppress some arrhythmias in patients (19,20); importantly, the clinical studies also showed that NAPA was not back-converted to procainamide and did not cause antinuclear antibodies. However, NAPA was not an especially potent antiarrhythmic, and it commonly caused non-cardiovascular adverse effects such as nausea. Interestingly, while both procainamide and NAPA prolong cardiac repolarization (manifest on the surface ECG as QT, interval prolongation, thought at the time to correlate with antiarrhythmic activity), only the parent drug blocked cardiac sodium channels. Thus, the NAPA structure did provide a starting point for the development of a series of new and highly potent QT-prolonging antiarrhythmic drugs that saw some enthusiasm for their use in the 1980s and 1990s (21,22).

Importantly, these clinical studies were executed at a time when the molecular basis of the slow and fast acetylator phenotypes were not well understood. We now know that there are two isoforms of the *N*-acetyltransferase enzyme arising from two different genes, NAT1 and NAT2. Constitutive expression of NAT1 accounts for basal enzyme function, and functionally important polymorphisms in NAT2, are thought to contribute to variability in overall enzymatic activity and thus to define the slow- and fast-acetylators phenotypes (23,24).

1.5. P-Glycoprotein

Digoxin is the prototypical substrate for transport by P-glycoprotein, the product of expression of the *MDR1* gene. Elevation of serum digoxin concentration and increased risk of serious toxicity result from co-administration of drugs that inhibit P-glycoprotein; these include quinidine, amiodarone, verapamil, itraconazole, erythromycin, and cyclosporine (25). MDR1 DNA polymorphisms that are linked to variability in serum digoxin concentrations have been described, although whether the polymorphisms are causative or in linkage disequilibrium to regulatory sites in the gene is not yet fully established (26,27).

2. ANTIARRHYTHMIC DRUG PHARMACODYNAMICS

Even at equivalent plasma concentrations of parent drug (and relevant active metabolites) the effect of antiarrhythmic therapies still vary considerably among individuals. There are a number of mechanisms that may underlie such variability. One is variable uptake or efflux transporter function, responsible for delivery to and removal from key intercellular sites of action. Drug effects may be different in normal vs. diseased (e.g., scarred or hypertrophied) hearts. Finally, a substantial body of knowledge has been accumulating over the past decade attesting to a prominent role of genetic factors in modulating normal and abnormal cardiac electrophysiology and, in turn, their responses to drug exposure.

An extraordinarily important starting point for this work has been identification of specific genes whose expression results in key proteins determining cardiac electrogenesis. Ion channels, pore-forming structures that respond to ligands or changes in voltage to permit transmembrane movement of specific ions, are the most important class of these proteins. One very important approach to identifying ion channel and related genes has been the study of rare monogenic (familial) arrhythmia syndromes (28).

Most currently available antiarrhythmic drugs were developed at a time when molecular mechanisms underlying arrhythmias were not appreciated. As a consequence, these drugs interact with a very limited number of molecular targets: the cardiac sodium channel, beta-1 adrenergic receptors, L-type calcium channels, and one specific potassium current (termed I_{Kr}). The genes whose expression results in these currents or receptors are now well understood. Many other genes have been identified whose expression generates or modulates ion currents in heart or otherwise prominently modulates overall cardiac electrical behavior. Indeed, this work has generated a lengthy list of potential ''new'' antiarrhythmic drug targets: these include potassium currents (I_{Ks} and I_{Kur}), pacemaker current (I_f), novel calcium currents (I_{Ca-T}; the A1D isoform of I_{Ca-L}), and cardiac connexins C×43 (29).

Study of the rare congenital arrhythmia syndromes has been important for ion channel biologists because it identifies genes whose expression plays a crucial role in normal cardiac electrophysiology. It has also been interesting to demonstrate that mutations in these genes may result not only in manifest congenital arrhythmia syndromes, but also subclinical phenotypes that can then be uncovered by drug administration. These are discussed next, followed by a consideration of how more common DNA polymorphisms, in these and other genes, might modulate drug responses.

3. THE PHARMACOGENETICS OF QT PROLONGATION

One important implication of this increasingly complex view of the molecular basis of cardiac electrophysiology is that variations in many genes may

modulate not only basal cardiac electrophysiology but also its response to drugs. One interesting and important example relates to the occasional development of marked prolongation of the QT interval on the surface electrocardiogram, a finding that is associated with a high risk of morphologically distinctive, potentially fatal, polymorphic ventricular tachycardia termed "torsades de pointes."

Torsades de pointes characteristically occurs in one of two clinical settings: in patients with a familial arrhythmia syndrome (the congenital Long QT Syndrome) and in patients exposed to drugs that have the potential to prolong the QT interval. The latter includes not only antiarrhythmic drugs, but a wide range of "non-cardiovascular" agents. Indeed, unexpected QT prolongation, torsades de pointes, and sudden death with the use of such drugs has been the single commonest reason for drug withdrawal in the United States over the past decade (30). Hence, evolving concepts with respect to the underlying molecular mechanisms of this distinctive arrhythmia syndrome have implications not only for management of patients with a relatively uncommon familial syndrome, but also have important implications for drug development in general.

The QT interval on the surface electrocardiogram represents the integrated behavior of a number of important ion currents active during the repolarization phase of a typical cardiac action potential. These include L-type calcium currents, a small contribution by sodium current, and several repolarizing potassium currents, notably the rapid component of the delayed rectifier (I_{Kr}), the slow component of the delayed rectifier (I_{Ks}), and the inward rectifier (I_{K1}). Studies of the congenital Long QT Syndrome in the mid-1990s identified mutations in the genes whose expression underlies I_{Kr}, I_{Ks}, and the cardiac sodium channel as the commonest causes of the syndrome (28). Importantly, cloning of the gene whose expression results in I_{Kr}, the Human Ether-a-go-go-Related Gene, or *HERG* (now known as *KCNH2*), was followed shortly thereafter by the recognition that virtually all drugs that prolong the QT interval do so by interacting with this particular ion channel protein (31). This work, then, provides an important link between a drug-induced syndrome and the congenital syndrome.

Another key observation made in the course of studying patients with the congenital syndrome was the identification of family members of probands in whom mutations could be identified but the QT interval appeared normal (32). This "incomplete penetrance" then raises the question of whether some or all of patients developing "idiosyncratic" drug-induced QT prolongation and torsades de pointes represent individuals with a subclinical form of the congenital syndrome. Addressing this issue is not straightforward, in part because several hundred mutations, in seven different genes, have now been identified in kindreds with this disease. Thus, unlike other common genetic diseases like cystic fibrosis or sickle cell anemia, there is a not a single predominant disease-associated mutation. As a result, any

DNA variant identified in a patient with drug-associated torsades de pointes may be a disease-associated mutation, a predisposing polymorphism, or an irrelevant polymorphism, and distinguishing among these may be difficult. Nevertheless, case reports (33,34) and one systematic survey (35) of approximately 100 individuals with drug-induced torsades de pointes do support the concept of a genetic predisposition increasing susceptibility to this arrhythmia syndrome in approximately 10% of subjects. Some of these individuals do, indeed, have variants that alter channel function in vitro and that are absent in large numbers of ethnically matched controls; such patients can therefore be viewed as having the congenital syndrome, clinically inapparent in absence of drug challenge. In fact, in some instances, other family members may have manifest QT-interval prolongation in the absence of drugs, further confirming the diagnosis of the congenital syndrome. DNA polymorphisms may also predispose to drug-induced torsades de pointes, as discussed further below.

Clinical studies have identified a multitude of risk factors for drug-induced torsades de pointes, including female gender, congestive heart failure, left ventricular hypertrophy, hypokalemia, bradycardia, and subclinical ion channel dysfunction. A unifying framework to account for how these multiple factors may modulate risk is that of "repolarization reserve," which suggests that the normal expression and function of multiple gene products—with potentially redundant function—ordinarily acts to maintain a QT interval within the normal range (36). Subtle dysfunction of such gene products may be clinically inapparent and only be exposed by administration of a QT-prolonging drug. Thus, an individual harboring a subclinical loss of function mutation or polymorphism in the gene encoding I_{Ks} may not display any clinical phenotype at baseline, because of a robust I_{Kr}. Administration of an I_{Kr} blocker to such an individual might leave them with very little repolarizing current and hence would cause marked QT-interval prolongation. The concept of "reduced repolarization reserve," developed to provide a unified approach to thinking about torsades de pointes risk, can be readily adapted to many other biological systems; in general, highly variable clinical responses to drug challenge may reflect subtle dysfunction of genes whose products modulate a highly complex phenotype like the QT interval.

4. ALTERED SODIUM CHANNEL FUNCTION

Mutations that result in a "gain of function" in the cardiac sodium channel gene (*SCN5A*) cause the long QT Syndrome (37). By contrast, mutations in the same gene may result in loss of function and a different congenital syndrome, the "Brugada Syndrome" (38,39). Individuals with Brugada Syndrome have structurally normal hearts, a high risk of ventricular fibrillation, and a morphologically distinctive electrocardiogram. As in the

Long QT Syndrome, mutation carriers who have a normal baseline electrocardiogram have been identified. In such cases, exposure to a sodium channel-blocking drug (e.g., flecainide or procainamide) is often used to unmask the Brugada Syndrome ECG phenotype. Indeed it was the observation that sodium channel block may unmask the ECG phenotype that led to consideration of *SCN5A* as a candidate in the Brugada Syndrome; in ~20% of probands, mutations can be found in this gene. Occasional patients treated clinically with sodium channel-blocking drugs such as drugs such as flecainide, procainamide, or tricyclic depressants do develop the typical Brugada Syndrome ECG pattern (40,41); whether such individuals actually harbor subclinical Brugada Syndrome or are otherwise predisposed to sudden death during administration of these agents is a logical possibility.

5. FROM RARE SYNDROMES TO COMMON POLYMORPHISMS

Cases of subclinical congenital arrhythmia syndromes exposed by drug challenge are quite rare, but nevertheless constitute an important "proof of concept" of a genetic basis for variable drug responses. On the other hand, intensive study of monogenic arrhythmia syndrome disease genes, as well as other genes whose expression contributes to normal electrophysiology, has identified polymorphisms, some quite common, that may also modulate arrhythmia susceptibility (Table 1). Population and association studies have identified such polymorphisms, with minor allele frequencies of 1.5–13% in control populations that appear to be over-represented among patients with drug-induced torsades de pointes. In such cases, in vitro studies have provided further support for the idea that the variant allele predisposes to the arrhythmia, especially in the presence of environmental triggers. The best-studied trigger for such arrhythmia susceptibility remains drug challenge, although other environmental stimuli, such as adrenergic activation or acute myocardial ischemia, could well play a role in other patients.

S1102Y in SCN5A is present in 13% of African Americans, but is absent or extremely rare in other ethnic groups (44). In vitro electrophysiologic studies showed that this variant does alter channel function, and clinical association studies suggest an overrepresentation of the Y allele in African-American subjects with a variety of arrhythmia syndromes, including drug-induced arrhythmias, compared to African-American controls. Similarly, Q9E was initially identified as a mutation in KCNE2 (encoding a function-modulating potassium channel subunit). The proband was an elderly African-American woman who displayed torsades de pointes on exposure to an I_{Kr}-blocking antibiotic (clarithromycin), and further study revealed that Q9E does, in fact, alter I_{Kr} function in vitro (45). More recently, Q9E has been recognized as a relatively common polymorphism

Table 1 Polymorphisms Associated with Variant Ion Channel Phenotypes

Gene	Function	Polymorphism	Minor allele frequency (%)	Function differs from wild-type in vitro (Ref.)	Clinical study suggesting effect of variant on ECG or drug response (Ref.)
SCN5A	Sodium channel	S1102Y	13	(44)	(44)
		H558R	20	(50)	
		C-92A (promoter region)	5	(52)	
KCNH2 (*HERG*)	I_{Kr} channel	K897T	10–20	(46,48)	(46,48,49)
KCNE1	I_{Ks} channel ancillary subunit	D85N	1–2	(42)	
KCNE2	I_{Kr} channel ancillary subunit	T8A	1.5	(43)	
		Q9E	5	(45)	(45)

(5% minor allele frequency), again detected only in African-Americans. Other polymorphisms have been described that appear to modulate normal cardiac electrophysiology, although none have yet been linked convincingly to increased susceptibility to drug-induced arrhythmias. Thus, for example, K897T in *HERG* has been associated with longer QT intervals, particularly among women (46–49). H558R in *SCN5A* does not appear to alter baseline sodium channel function, but does modulate the clinical and in vitro phenotype of a disease-associated mutation in the same channel (50). A promoter polymorphism in the gene encoding the connexin C×40, whose normal function underlies cell–cell communication especially in atrium, appears to reduce gene expression (51). This polymorphism thus becomes a logical candidate gene for modulating the development of the very common arrhythmia atrial fibrillation and thus its response to drugs. Polymorphisms in the promoter region of cardiac sodium channel have been described that modulate expression of the channel in vitro (52); whether such polymorphisms underlie variable expression of the channel in patients and thus variable responses to challenges such as sodium channel-blocking drugs or acute myocardial ischemia is not yet known. Finally, recent clinical and mechanistic

studies implicate activation of a number of key signaling pathways, such as the beta-adrenergic system, the renin-angiotensin-aldosterone system, oxidant stress, or inflammation, as potential arrhythmia triggers (53–58). Variants in these pathways then become new candidates for modulating arrhythmia susceptibility.

6. SUMMARY

Currently available arrhythmic drugs were developed at a time when the molecular basis of cardiac arrhythmias was not understood. Thus, therapy has been largely empiric, efficacy has been unpredictable, and serious side effects have been common. Some of variability in clinical action of antiarrhythmic drugs can be attributed directly to variable drug disposition, through specific pathways whose activity is now well recognized to be modulated by common DNA polymorphisms.

An in-depth understanding of the molecular basis or normal cardiac electrophysiology has come from a number of approaches, notably including the intensive study of families with rare monogenic arrhythmia syndromes. Variations in the complex physiologic signaling system that results in normal electrical activity now appears to be a proximate cause of most cardiac arrhythmias. DNA variants in genes encoding elements of this system may result in disease, or may more commonly modulate arrhythmia risk in the face of exogenous stressors, including drugs, in the susceptible patient. Further definition of these molecular mechanisms and the polymorphisms that underlie such susceptibility should lead to the development of drug therapies targeting underlying pathophysiologic mechanisms, and lacking common and serious adverse effects. As in many areas of pharmacogenetics, real advances require a partnership between clinical and basic investigators to precisely identify variable and important clinical phenotypes and then define mechanisms underlying that variability.

ACKNOWLEDGMENTS

Supported in part by grants from the U.S. Public Health Service (HL46681, HL49989, and HL65962). Dr. Roden is the holder of the William Stokes Chair in Experimental Therapeutics, a gift from the Dai-ichi Corporation.

REFERENCES

1. Data JL, Wilkinson GR, Nies AS. Interaction of quinidine with anticonvulsant drugs. N Engl J Med 1976; 294:699–702.
2. Tirona RG, Lee W, Leake BF, Lan LB, Cline CB, Lamba V, Parviz F, Duncan SA, Inoue Y, Gonzalez FJ, Schuetz EG, Kim RB. The orphan nuclear receptor

HNF4alpha determines PXR- and CAR-mediated xenobiotic induction of CYP3A4. Nat Med 2003; 9:220–224.
3. Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, Watkins PB, Daly A, Wrighton SA, Hall SD, Maurel P, Relling M, Brimer C, Yasuda K, Venkataramanan R, Strom S, Thummel K, Boguski MS, Schuetz E. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. Nat Genet 2001; 27:383–391.
4. Meyer UA, Zanger UM. Molecular mechanisms of genetic polymorphisms of drug metabolism. Annu Rev Pharmacol Toxicol 1997; 37:269–296.
5. Lennard MS, Silas JH, Freestone S, Ramsay LE, Tucker GT, Woods HF. Oxidation phenotype—a major determinant of metoprolol metabolism and response. N Engl J Med 1982; 307:1558–1560.
6. Zhou HH, Wood AJ. Stereoselective disposition of carvedilol is determined by CYP2D6. Clin Pharmacol Ther 1995; 57:518–524.
7. Siddoway LA, Thompson KAMKT, McAllister CB, Wang T, Wilkinson GR, Roden DM, Woosley RL. Polymorphism of propafenone metabolism and disposition in man: clinical and pharmacokinetic consequences. Circulation 1987; 75:785–791.
8. Thompson KAMKT, Iansmith DH, Siddoway LA, Woosley RL, Roden DM. Potent electrophysiologic effects of the major metabolites of propafenone in canine Purkinje fibers. J Pharmacol Exp Ther 1988; 244:950–955.
9. Lee JT, Kroemer HK, Silberstein DJ, Funck-Brentano C, Lineberry MD, Wood AJ, Roden DM, Woosley RL. The role of genetically determined polymorphic drug metabolism in the beta-blockade produced by propafenone. N Engl J Med 1990; 322:1764–1768.
10. Hill MR, Gotz VP, Harman E, McLeod I, Hendeles L. Evaluation of the asthmogenicity of propafenone, a new antiarrhythmic drug: comparison of spirometry with methacholine challenge. Chest 1986; 90:698–702.
11. Lau CP, Chow MS, Tse HF, Tang MO, Fan C. Control of paroxysmal atrial fibrillation recurrence using combined administration of propafenone and quinidine. Am J Cardiol 2000; 86:1327–1332.
12. Barbey JT, Thompson KAMKT, Echt DS, Woosley RL, Roden DM. Antiarrhythmic activity, electrocardiographic effects and pharmacokinetics of the encainide metabolites O-desmethyl encainide and 3-methoxy-O-desmethyl encainide in man. Circulation 1988; 77:380–391.
13. Evers J, Eichelbaum M, Kroemer HK. Unpredictability of flecainide plasma concentrations in patients with renal failure: relation to side effects and sudden death? Ther Drug Monit 1994; 16:349–351.
14. Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. Lancet 1999; 353:717–719.
15. Taube J, Halsall D, Baglin T. Influence of cytochrome P-450 CYP2C9 polymorphisms on warfarin sensitivity and risk of over-anticoagulation in patients on long-term treatment. Blood 2000; 96:1816–1819.
16. Higashi MK, Veenstra DL, Kondo LM, Wittkowsky AK, Srinouanprachanh SL, Farin FM, Rettie AE. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. JAMA 2002; 287:1690–1698.

17. Woosley RL, Drayer DE, Reidenberg MM, Nies AS, Carr K, Oates JA. Effect of acetylator phenotype on the rate at which procainamide induces antinuclear antibodies and the lupus syndrome. N Engl J Med 1978; 298:1157–1159.
18. Uetrecht J. Drug metabolism by leukocytes and its role in drug-induced lupus and other idiosyncratic drug reactions. Crit Rev Toxicol 1990; 20: 213–235.
19. Dangman KH, Hoffman BF. In vivo and in vitro antiarrhythmic and arrhythmogenic effects of *N*-acetyl procainamide. J Pharmacol Exp Ther 1981; 217: 851–862.
20. Roden DM, Reele SB, Higgins SB, Wilkinson GR, Smith RF, Oates JA, Woosley RL. Antiarrhythmic efficacy, pharmacokinetics and safety of *N*-acetylprocainamide in human subjects: comparison with procainamide. Am J Cardiol 1980; 46:463–468.
21. Cobbe SM. Modification of class III anti-arrhythmic activity in abnormal myocardium. Cardiovasc Res 1988; 22:847–854.
22. Hondeghem LM, Snyders DJ. Class III Antiarrhythmic agents have a lot of potential, but a long way to go: reduced effectiveness and dangers of reverse use-dependence. Circulation 1990; 81:686–690.
23. Blum M, Demierre A, Grant DM, Heim M, Meyer UA. Molecular mechanism of slow acetylation of drugs and carcinogens in humans. Proc Natl Acad Sci USA 1991; 88:5237–5241.
24. Grant DM, Hughes NC, Janezic SA, Goodfellow GH, Chen HJ, Gaedigk A, Yu VL, Grewal R. Human acetyltransferase polymorphisms. Mutat Res 1997; 376:61–70.
25. Fromm MF, Kim RB, Stein CM, Wilkinson GR, Roden DM. Inhibition of P-glycoprotein-mediated drug transport: a unifying mechanism to explain the interaction between digoxin and quinidine. Circulation 1999; 99:552–557.
26. Hoffmeyer S, Burk O, von Richter O, Arnold HP, Brockmoller J, Johne A, Cascorbi I, Gerloff T, Roots I, Eichelbaum M, Brinkmann U. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. Proc Natl Acad Sci USA 2000; 97:3473–3478.
27. Kim RB, Leake BF, Choo EF, Dresser GK, Kubba SV, Schwarz UI, Taylor A, Xie HG, McKinsey J, Zhou S, Lan LB, Schuetz JD, Schuetz EG, Wilkinson GR. Identification of functionally variant MDR1 alleles among European-Americans and African-Americans. Clin Pharmacol Ther 2001; 70:189–199.
28. Keating MT, Sanguinetti MC. Molecular and cellular mechanisms of cardiac arrhythmias. Cell 2001; 104:569–580.
29. Roden DM. Antiarrhythmic drugs: past, present, and future. J Cardiovasc Electrophysiol 2003; 14:1389–1396.
30. Roden DM. Drug-induced prolongation of the QT interval. N Engl J Med 2004; 350:1013–1022.
31. Sanguinetti MC, Jiang C, Curran ME, Keating MT. A mechanistic link between an inherited and an acquired cardiac arrhythmia: *HERG* encodes the I_{Kr} potassium channel. Cell 1995; 81:299–307.
32. Priori SG, Napolitano C, Schwartz PJ. Low penetrance in the long-QT syndrome: clinical impact. Circulation 1999; 99:529–533.

33. Donger C, Denjoy I, Berthet M, Neyroud N, Cruaud C, Bennaceur M, Chivoret G, Schwartz K, Coumel P, Guicheney P. KVLQT1 C-terminal missense mutation causes a forme fruste long-QT syndrome. Circulation 1997; 96:2778–2781.
34. Napolitano C, Schwartz PJ, Brown AM, Ronchetti E, Bianchi L, Pinnavaia A, Acquaro G, Priori SG. Evidence for a cardiac ion channel mutation underlying drug-induced QT prolongation and life-threatening arrhythmias. J Cardiovasc Electrophysiol 2000; 11:691–696.
35. Yang P, Kanki H, Drolet B, Yang T, Wei J, Viswanathan PC, Hohnloser SH, Shimizu W, Schwartz PJ, Stanton MS, Murray KT, Norris K, George ALJ, Roden DM. Allelic variants in Long QT disease genes in patients with drug-associated Torsades de Pointes. Circulation 2002; 105:1943–1948.
36. Roden DM. Taking the idio out of idiosyncratic—predicting torsades de pointes. PACE 1998; 21:1029–1034.
37. Wang Q, Shen J, Splawski I, Atkinson D, Li Z, Robinson JL, Moss AJ, Towbin JA, Keating MT. *SCN5A* mutations associated with an inherited cardiac arrhythmia, long QT syndrome. Cell 1995; 80:805–811.
38. Antzelevitch C, Brugada P, Brugada J, Brugada R, Shimizu W, Gussak I, Perez Riera AR. Brugada syndrome: a decade of progress. Circ Res 2002; 91:1114–1118.
39. Chen QY, Kirsch GE, Zhang DM, Brugada R, Brugada J, Brugada P, Potenza D, Moya A, Borggrefe M, Breithardt G, Ortizlopez R, Wang Z, Antzelevitch C, Obrien RE, Schulzebahr E, Keating MT, Towbin JA, Wang Q. Genetic basis and molecular mechanism for idiopathic—ventricular fibrillation. Nature 1998; 392:293–296.
40. Tada H, Sticherling C, Oral H, Morady F. Brugada syndrome mimicked by tricyclic antidepressant overdose. J Cardiovasc Electrophysiol 2001; 12:275.
41. Krishnan SC, Josephson ME. ST segment elevation induced by class Ic antiarrhythmic agents: underlying electrophysiologic mechanisms and insights into drug-induced proarrhythmia. J Cardiovasc Electrophysiol 1998; 9:1167–1172.
42. Wei J, Yang IC, Tapper AR, Murray KT, Viswanathan P, Rudy Y, Bennett PB, Norris K, Balser JR, Roden DM, George AL. *KCNE1* polymorphism confers risk of drug-induced long QT syndrome by altering kinetic properties of IKs potassium channels. Circulation 1999; I:495.
43. Sesti F, Abbott GW, Wei J, Murray KT, Saksena S, Schwartz PJ, Priori SG, Roden DM, George AL, Goldstein SA Jr. A common polymorphism associated with antibiotic-induced cardiac arrhythmia. Proc Natl Acad Sci USA 2000; 97:10613–10618.
44. Splawski I, Timothy KW, Tateyama M, Clancy CE, Malhotra A, Beggs AH, Cappuccio FP, Sagnella GA, Kass RS, Keating MT. Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia. Science 2002; 297:1333–1336.
45. Abbott GW, Sesti F, Splawski I, Buck ME, Lehmann MH, Timothy KW, Keating MT, Goldstein SAN. MiRP1 forms I_{Kr} potassium channels with HERG and is associated with cardiac arrhythmia. Cell 1999; 97:175–187.
46. Bezzina CR, Verkerk AO, Busjahn A, Jeron A, Erdmann J, Koopmann TT, Bhuiyan ZA, Wilders R, Mannens MM, Tan HL, Luft FC, Schunkert H, Wilde AA. A common polymorphism in KCNH2 (HERG) hastens cardiac repolarization. Cardiovasc Res 2003; 59:27–36.

47. Anson BD, Ackerman MJ, Tester DJ, Will ML, Delisle BP, CL Anderson BP, January CT. Molecular and functional characterization of common polymorphisms in HERG (KCNH2) potassium channels. Am J Physiol Heart Circ Physiol 2004; 286:H2434–H2441.
48. Paavonen KJ, Chapman H, Laitinen PJ, Fodstad H, Piippo K, Swan H, Toivonen L, Viitasalo M, Kontula K, Pasternack M. Functional characterization of the common amino acid 897 polymorphism of the cardiac potassium channel KCNH2 (HERG). Cardiovasc Res 2003; 59:603–611.
49. Pietila E, Fodstad H, Niskasaari E, Laitinen PP, Swan H, Savolainen M, Kesaniemi YA, Kontula K, Huikuri HV. Association between HERG K897T polymorphism and QT interval in middle-aged Finnish women. J Am Coll Cardiol 2002; 40:511–514.
50. Viswanathan PC, Benson DW, Balser JR. A common SCN5A polymorphism modulates the biophysical effects of an SCN5A mutation. J Clin Invest 2003; 111:341–346.
51. Groenewegen WA, Firouzi M, Bezzina CR, Vliex S, van Langen IM, Sandkuijl L, Smits JP, Hulsbeek M, Rook MB, Jongsma HJ, Wilde AA. A cardiac sodium channel mutation cosegregates with a rare connexin40 genotype in familial atrial standstill. Circ Res 2003; 92:14–22.
52. Yang P, Kupershmidt S, Roden DM. Cloning and initial characterization of the human cardiac sodium channel (SCN5A) promoter. Cardiovasc Res 2004; 61:56–65.
53. Rubart M, Zipes DP. NO hope for patients with atrial fibrillation. Circulation 2002; 106:2764–2766.
54. Aviles RJ, Martin DO, Apperson-Hansen C, Houghtaling PL, Rautaharju P, Kronmal RA, Tracy RP, Van Wagoner DR, Psaty BM, Lauer MS, Chung MK. Inflammation as a risk factor for atrial fibrillation. Circulation 2003; 108:3006–3010.
55. Carnes CA, Chung MK, Nakayama T, Nakayama H, Baliga RS, Piao S, Kanderian A, Pavia S, Hamlin RL, Mccarthy PM, Bauer JA, Van Wagoner DR. Ascorbate attenuates atrial pacing-induced peroxynitrite formation and electrical remodeling and decreases the incidence of postoperative atrial fibrillation. Circ Res 2001; 89:E32–E38.
56. Pedersen OD, Bagger H, Kober L, Torp-Pedersen C. Trandolapril reduces the incidence of atrial fibrillation after acute myocardial infarction in patients with left ventricular dysfunction. Circulation 1999; 100:376–380.
57. Madrid AH, Bueno MG, Rebollo JM, Marin I, Pena G, Bernal E, Rodriguez A, Cano L, Cano JM, Cabeza P, Moro C. Use of irbesartan to maintain sinus rhythm in patients with long-lasting persistent atrial fibrillation: a prospective and randomized study. Circulation 2002; 106:331–336.
58. Vermes E, Tardif JC, Bourassa MG, Racine N, Levesque S, White M, Guerra PG, Ducharme A. Enalapril decreases the incidence of atrial fibrillation in patients with left ventricular dysfunction insight from the studies of left ventricular dysfunction (SOLVD) trials. Circulation 2003; 107:2926–2931.

8

Interethnic Differences in Drug Response

Werner Kalow
Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada

1. INTRODUCTION

Attention to interethnic differences has become a major aspect of pharmacogenetics (1–9) stimulated by studies of drug response and toxicity in various human populations who differed in their response, even when taking usually well tolerated doses of some common therapeutic chemicals. The purpose of this chapter is to indicate essential elements of this broad topic, and to provide examples of some of its important aspects.

When pharmacogenetic studies started, very little was known about the genes responsible for either intra-individual or interethnic differences. In fact, nobody knew for sure whether or not most of the interethnic differences had a genetic or a cultural basis; within a population, family or twin studies can easily show that a trait is genetic. However, the evidence that a trait is genetically controlled in one population does not prove that differences between populations are genetic. Many people did simply not believe that population differences were genetic; they did not want to be racists. They used words like "cross-cultural" or similar terms as an expression of their beliefs. It is only the rise of modern methods to study genes that established clearly the genetic basis of many interethnic differences of pharmacogenetics.

In short, differences in drug response between human populations drew first attention 4 to 5 decades ago. However, because of the initial methodological shortcomings, the science of interethnic pharmacogenetics is in a strict sense only about 20 years old (10).

2. FUNDAMENTAL QUESTIONS

2.1. What Is Ethnicity?

At a time not too long ago, ethnic groups, or human races were defined by geography and appearence, most often simply by skin color, sometimes additionally by language. These factors are still considered by many people as main indicators, but at the present time, the scientifically defining factors are genetic differences between populations.

It is true that the majority of genes in different human populations are identical. Nevertheless, it would be wrong and inappropriate to take this overall similarity as an opportunity to disavow the medical importance of interethnic divisions of the human species.

A revolutionary book by Nei (11) entitled "Molecular Evolutionary Genetics" divided populations by their genetic diversities. More recently, this work has been greatly extended, mostly confirmed but with some changes, by Cavalli-Sforza et al. (12) in a book entitled "The History and Geography of Human Genes."

These authors tested 120 allele frequencies in 42 human populations. They calculated from these measurements genetic distances between the populations and estimated the times of their separation. They summarized their data, and thus constructed nine population clusters (Fig. 1). The greatest differences were between African and all non-African populations, in support of the theory that all present human beings derived from a wave of emigrants who left Africa approximately 100,000 years ago.

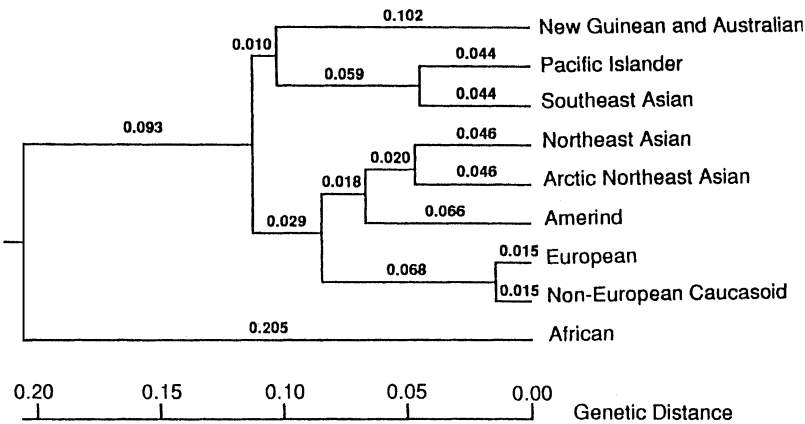


Figure 1 Linkage Tree. Analysis of nine population clusters, condensed from data obtained by studying 42 populations. The genetic distance 0.2 represents approximately 150,000 years. *Source*: Figure 2.3.3 from Ref. 12.

An interesting point in this figure is the indication that the separation between south–east and north–east Asians (appr. 50,000 year ago) is older than that between north–east Asians and Europeans (appr. 40,000 years). Most Chinese are north–east Asians while Japanese and Koreans are south–east Asians. Hence drug tests comparing Japanese and Koreans tend to be more similar than the equivalent comparisons between Japanese and Chinese.

It is largely a matter of economic and cultural factors that we know more about pharmacogenetic differences between Caucasian and Asian populations than between these and African populations. Data obtained with African-Americans must be interpreted as having a Caucasian admixture of approximately 30%; African populations are not a uniform group and often differ substantially from one another.

2.2. Compare Individual and Ethnic Differences in Drug Metabolism

As time went on, it became more and more clear that many genetic variations of drug-metabolizing enzymes that determine inter-individual differences of drug clearance, also show differences between populations. There are two possible kinds of difference: first, it is an almost universal observation that the frequency of a polymorphism is found to differ between populations. Of the numerous drug-metabolizing enzymes that show pharmacogenetic variation between individuals, at least 88% show an interethnic difference of variant frequency (Table 1). In most cases for which interethnic difference is on record, such difference has not yet been searched for. In any case, one can say with assurance, that inter-individual differences of drug-metabolizing enzymes are usually paralelled by interethnic differences in allele frequencies.

Second, there are often different variants in different populations; some phenotypic consequences of CYP2D6 variations are shown as an example in Figure 2 (1). There is a lower average enzyme activity in both, Chinese and Africans, in comparison to that in Europeans. These differences in activity represent structural changes of the CYP2D6 protein that affect enzyme function. However, the enzyme structures that cause the lower activities are different in the Chinese (13) and in the African populations (14). It is an independent fact that enzyme absence is more rare in Asians and Africans than in Europeans (4).

2.3. Not All Interethnic Differences in Drug Response Are Genetic

While this chapter aims to describe genetic factors which cause differences in drug response between human populations, one should not forget that population differences may also be caused by environmental influences. In

Table 1 Drug Metabolizing Enzymes Showing Genetic Variation[a]

Esterases	
1	+ Butyrylcholinesterase (106)
2	+ Paraoxonase/arylesterase (107)
Transferases	
3	+ *N*-acetyltransferases (Natl) (108)
4	+ *N*-acetyltransferases (Nat2) (108)
5	+ Catechol-*0*-methyltransferase (109)
6	0 Histamine methyltransferase (110)
7	0 Thiol methyltransferase (111)
8	+ Thiopurine methyltransferase (112)
9	+ Sulfotransferases (113)
10	+ Glutathione-*S*-Transferases (GSTM1) (114)
11	+ Glutathione-*S*-Transferases (GSTT1) (114)
12	0 Glutathione-*S*-Transferases (GSTM3) (115)
13	0 Glutathione-*S*-Transferases (GSTP1) (116)
14	+ Glucuronosyltransferase (UGT1A1) (117)
15	0 Glucuronosyltransferase (UGT2B4) (118)
16	0 Glucuronosyltransferase (UGT2B7) (119)
17	0 Glucuronosyltransferase (UGT2B15) (120)
18	+ Amobarbital-glucosyltransferase (121)
Reductases	
19	+ NAD (P) H:quinone oxidoreductase (122)
20	+ Glucose-6-phosphate dehydrogenase (123)
21	+ Epoxide hydrolase, microsomal (124)
Oxidases	
22	+ Alcohol dehydrogenase, class 1, ADH2 (B) (125)
23	+ Alcohol dehydrogenase, class 1, ADH3 ( ) (126)
24	+ Aldehyde dehydrogenase, mitochondrial (127)
25	+ Monoamine oxidase A (128)
26	+ Monoamine oxidase B (129)
27	+ Catalase (130)
28	+ Superoxide dismutase (131)
29	+ Trimethylamine *N*-oxidase (132)
30	+ Dihydropyrimidine dehydrogenase (133)
Cytochromes P450	
31	+ CYP1A1 (134)
32	+ CYP1A2 (135)
33	0 CYP1B1 (136)
34	+ CYP2A6 (137)
35	+ CYP2B6 (138)

(*Continued*)

Table 1 Drug Metabolizing Enzymes Showing Genetic Variation[a] (*Continued*)

36	0 CYP2C8 (139)
37	+ CYP2C9 (140)
38	+ CYP2C18 (141)
39	+ CYP2C19 (142)
40	+ CYP2D6 (143)
41	+ CYP2E1 (144)
42	+ CYP3A4 (145)
43	+ CYP3A5 (146)
44	0 CYP4B1 (147)

[a]A pharmacogenetic list of variable enzymes, marked + if the occurrence of interethnic differences in variant frequency is known ("0" usually means absence of interethnic comparisons).

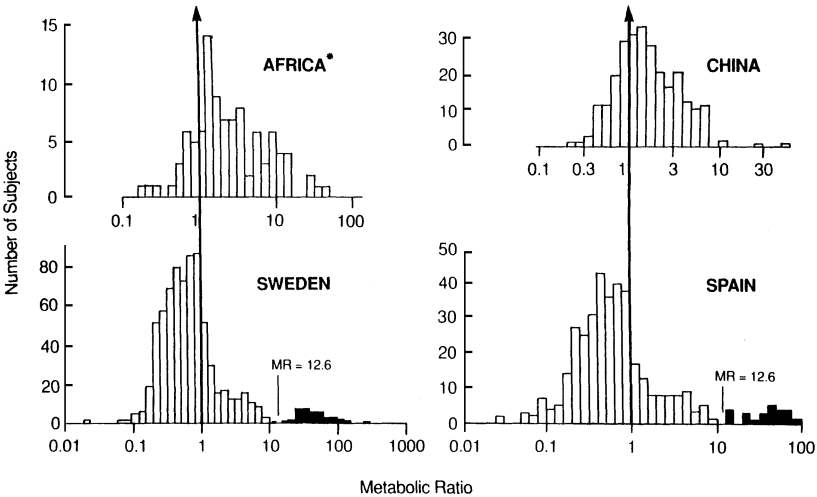


Figure 2 Frequency distributions of debrisoquine metabolic ratios (MR) in four populations. An alignment of data from four separate studies. The abscissa indicates on a logarithmic scale the metabolic ratio debrisoquine/4-OH-debrisoquine in urine after administration of a test dose of debrisoquine; these are conventional plots in which the increasing ratios reflect decreasing metabolism. The black bars indicate subjects classified as genetically poor metabolizers, usually defined as subjects with a metabolic ratio >12.6. Each of the four inserts represents an adaptation of a published illustration so that their abscissas are comparable and aligned for MR of unity. The insert marked "China" represents a study of 269 Han, "Africa" a study of 92 Venda, "Sweden" a study of 752 Swedes, and "Spain" a study of 377 Spaniards. The entry for Africa is marked with an asterisk because it represents tribal data of unknown generality. The measurements from China and Sweden were comparable, as ensured by controls. *Source*: Compiled from Refs. 1, 99–101.

some cases, environmental factors may be increased by genetic variants which should not be called "pharmacogenetic."

As pointed out by Anderson et al. (15), starvation, malnutrition, and protein deficiency may all cause differences in drug disposition and thereby differences of drug action. Even relatively minor food deficiencies as observed in Berlin after the second World War (16) caused some unusual reactions, e.g., death from injection of the old and generally safe local anesthetic drug procaine (17).

Climate affects food production and thereby is a factor that determines ethnic characterises of nutrition. Different food intakes may cause differences in drug response: some foods may cause enzyme induction (18), others enzyme inhibition (19). In this context, it may be worth mentioning that the P450 cytochrome CYP2A6, which metabolizes nicotine (20) and which varies between populations, does cause interethnic differences in smoking habits (21); since cigarette smoking causes enzyme induction, there must be population differences of drug-metabolizing capacity on that basis.

In some African and Polynesian populations, cardiovascular diseases often have different biochemical causes than in Caucasians. Different drug responses to these diseases may therefore have pathological rather than pharmacogenetic causes (22,23).

Also cultural differences exist. There are population differences in attitude towards disease and cure (24). These could affect compliance and perception of therapeutic benefits, and thereby psychologically influence the responsiveness to drug therapy. In western medicine, the occurrence of placebo effects are well documented; this is the name for imagined effects occurring when people are fooled into thinking that they had a drug while they were given only a little sugar or other fake.

Thus, while most interethnic differences in drug response are now known to have a genetic basis, it would be wrong to assume automatically that there must be a genetic cause.

3. SOME SPECIFIC EXAMPLES

A good understanding of the pharmacogenetic features of interethnic differences in drug-metabolizing enzymes should be obtainable by careful consideration of data obtained with some widely studied enzymes. We will therefore present some pertinent data of two P450 cytochromes in human liver, of the *N*-acetyltransferases, of G-6-PD deficiency enzymes, and of the ethanol-metabolizing enzymes. These should serve to illustrate principles and difficulties encountered by the student of interethnic pharmacogenetics, but there will be no attempt to provide a comprehensive overview over the many existing examples.

3.1. The P450 Cytochromes

As indicated in Table 1, differences between human populations are known for 11 of the 12 listed P450 cytochromes. Particularly well documented are the data for CYP2D6 and for CYP2C19. These will be briefly reviewed in the following paragraphs; they show instructive differences and similarities. CYP2D6 is the most investigated of all enzymes in pharmacogenetics; the medline quoted 2223 publications on this topic in 2003.

3.1.1. CYP2D6

Mutations in CYP2D6 (25) may lead to three kinds of functional changes: absence, decrease, or increase of enzyme activity. Lack of enzyme activity indicates in most cases a frameshift mutation or a splicing defect, but there is also a case of totally missing enzyme formation. Decreases of enzyme activity indicate a change of enzyme structure; these can be clinically treacherous since the decrease may differ in severity for different sustrates. Enhanced enzyme activity indicates usually gene duplication, sometimes multiplication; a case of 13-fold multiplication has been observed. Everyone of these enzyme changes may differ between human populations.

Marez et al. (26) described 48 different DNA variants in human CYP2D6, leading to 53 different alleles, which means that some alleles contain several DNA changes. Daly et al. (27) developed a rational nomenclature for CYP2D6 variants, and they also indicated the biochemical changes characterizing each variant. Lack of enzyme activity occurs with a high frequency of about 7% in all Caucasian populations, and a frequency usually less than 1% in all Asian populations. In about 75% of the Caucasian cases, the enzyme lack represents the frameshift mutation of CYP2D6*4. This mutation is absent in Asians, accounting for the low incidence of enzyme lack. On the other hand, the splicing mutant CYP2D6*5 occurs with low but similar frequency in all major population groups and accounts for the fact that poor metabolizers are occurring everywhere, even when rarely so.

Table 2 gives a list of some CYP2D6 mutations and their occurrences in major human populations. The list shows that the gene may carry as many as seven mutations in Europe, up to four mutations in China, while only three were seen in Japan. Only one nucleotide change (G4268C) occurs in all populations. One must assume that this is a very old mutation which occurred before *homo sapiens* left Africa, and therefore is seen everywhere. Mutations specific for any given population are probably relatively recent events. Thus, the time of mutation seems to explain at least some population differences.

The deficiency of CYP2D6 was discovered in England by testing debrisoquine as a substrate (28), in Germany by testing sparteine (29) . The fact that the deficiency of a single enzyme was responsible for both deficiencies was demonstrated later (30); the correlation coefficient of the rates of

Table 2 Mutations in Cytochrome CYP2D6

	Nucleotide changes and their locations											
CYP2D6 allele	C > T	C > A	A > G	C > G	C > T	C > T	G > C	G > C	G > A	C > T	G > C	People from
∗2								1749		2938	4268	Europe
∗4	188	1062	1072	1085				1749	1934		4268	—
∗4B	188	1062	1072	1085					1934		4268	—
∗10A	188							1749			4268	Japan
∗10B	188				1127			1749			4268	China
∗10C	188				1127			1749			4268+	—
17						1111	1726			2938	4268	Africa

The variants *4A and *4B represent enzymes without activity, all others have reduced activity compared to the wildtype. The + sign at CUP2D6*10C indicates exon 9 conversion into that of pseudogene CYP2D7.

Source: The data are compiled from information on CYP2D6 nomenclature by Daly et al. (27), which details the numerous original sources of the data, and on the overview by Marez et al. (26).

metabolism of the two substrates was found to be $r = 0.91$ in Europe. The presence or absence of enzyme activity is a straight-forward phenomenon, but problems may be encountered if structural variants occur. Woolhouse et al. (31) observed in a Ghanaian population that many poor metabolizers of debrisoquine could readily metabolize sparteine. Masimirembwa et al. (14) made a similar observation comparing the metabolisms of debrisoquine and metropolol (another CYP2D6 substrate) in Zimbabwe. It is a general truth that structural enzyme variants do not always have the same binding affinities for all substrates.

Many African data were due to the variant CYP2D6*17 which has reduced activity. This variant occurred with an allele frequency of 37% in Zimbabwe, of 17% in Tanzania, and in 9% in Ethiopia (8). The reduced enzyme activity in Asia is due to a different variant, CYP2D6*10 characterized by C188T, which has in China a frequency of 51%. Besides their generally reduced activity, the Asian and African variants have no similarity. The activities of these variants towards different substrates requires much additional investigation.

Gene duplication (or multiplication) representing abnormally high enzyme activity occurred in about 1–2% of the Swedish population, was 3.7% in Germany, 7–10% in Spain, 20% in Saudi Arabia, and 29% in Ethopia (8). Europeans with gene duplication are ultrarapid metabolizers of debrisoquine. The enzyme activity in Ethiopia is not as high as one might expect from the number of duplicants; the reason is not entirely clear. Perhaps a CYP2D6 variant with reduced activity is duplicated. The relatively high duplication frequency in Spain is almost certainly a remnant of the historical invasion of Spain by Arabs.

The large number of nucleotide changes in CYP2D6 raises questions: perhaps this fact points to a relative lack of any physiological importance of CYP2D6, allowing the enzyme to be eliminated without danger in a drug-free environment. On the other hand, this enzyme occurs in human brain and its absence seems to cause a difference of mental attitudes (32).

3.1.2. CYP2C19

The drug that led to the discovery of the polymorphic variation of CYP2C19 was mephenytoin (33,34), an anti-convulsant agent that is no longer in regular therapeutic use. It is a racemic drug of which the R-enantiomer is generally demethylated at a slow rate, while the S-enantiomer is hydroxylated at a very fast rate in some people and slow in others (for review see Ref. 35). Wrighton et al. (36) and Goldstein and de Morais (37) identified the P450 cytochrome CYP2C19 as the variable enzyme that caused this metabolic difference between people.

Some observations which carried this topic to prominence may be summarized by the statement that slow metabolizers of S-mephenytoin were found with a freqency of 2.8% in several Caucasian populations (38), but

with considerable higher frequencies (20–30%) in Asians (8). Measurements of mephenytoin metabolism in Africa indicate generally low frequencies of the slow metabolizer genotypes (38). The clinical significance of CYP2C19 polymorphism has been recently summarized (39).

Besides the wildtype CYP2C19*1, there are two deficiency alleles, CYP2C19*2 and *3 (formerly called CYP2C19m_1 and m_2). Both alleles cause a lack of enzyme formation and any person homozygous for these two alleles or with their combination will register as having the metabolic deficiency; wildtype heterozygotes are as a rule not phenotypically distinguished from wildtype homozygotes. The higher frequency of the deficiency in Asians is mostly due to CYP2C19*3. On Vanuato and other pacific islands this latter allele occurred in 70% of the populations (40).

Substrates other than mephenytoin which have caused clinical disturbances in people with deficient CYP2C19 alleles include omeprazol (41), proguanil (42), and citalopram (43). Additional substrates listed (8) include clomipramine, imipramine, diazepam, and propanolol.

Different investigators indicated interethnic differences in the kinetics of d-iazepam, omeprazol, and clomipramine, all known to be CYP2C19 substrates (8). The pharmacokinetic differences affecting these substrates could not be fully explained by the absence of CYP2C19, but structural changes and consequent race-dependent substrate selectivities (as known for CYP2D6) are not known for CYP2C19. Hence, the interethnic variations in the metabolism of these CYP2C19 substrates must mean, that other enzymes, which marginally participate in their metabolism show interethnic variability. This marginal participation becomes visible in the absence but is not noted in the presence of CYP2C19 activity.

3.2. *N*-Acetyltransferases

The introduction of isoniazid into medicine was an exciting advance because of its revolutionary capability to cure tuberculosis, a kind of cure never seen before (44) . When it was found that the drug caused tingling in hands or feet and other neurological disturbances in some but not all subjects, the cause was intensely investigated. It turned out, that people with this side effect did not properly metabolize isoniazid (45,46). The faulty drug-metabolizing enzyme was identified as a form of *N*-acetyltransferase. This was a major discovery in support of the then young science of pharmacogenetics.

Not very long after the discovery of the deficiency of isoniazid acetylation, it turned out that the frequency of this phenotypic fault also differed much between the world's populations (47). Today, the reason for this particular interethnic variability is still unknown; an influence of geographical latitude and climate is suspected. Such scientific uncertainty is common in this field of research, and it contrasts with the clear-cut findings relating malaria resistance and G-6-PD deficiency (see below).

Later it was found that there are two *N*-acetyltransferases, designated NATl and NAT2 (48). Isoniazid is metabolized only by NAT2, its variability was thus early recognized. Vatsis and Weber (49) reported evidence of genetic variation of NATl.

Grant et al. (50) described seven nucleotide changes in the NAT2 gene which gave rise to 15 variant alleles; of these, four resulted in rapid and nine in slow acetylator phenotypes, the remaining two are yet undetermined. In vitro tests indicated that most slow acetylations were due to instability of the enzyme, only one case showed decreased V_{max}. They listed the frequency of three selected slow acetylator alleles in 16 populations with together over 4000 subjects.

Their set of data may be summarized by the statements that the $T^{341}C$ allele were the cause of slow acetylation in most Caucasians and Africans (with frequencies of 0.450 and 0.350, respectively), but was much rarer in all other populations. The slow acetylator allele $G^{590}A$ occurred in all populations exept in Amerindians with the relatively high average frequency of $0.275 +/- 73$, and thereby was most often responsible for slow acetylation in Asians.

Grant et al. (50) also reported 10 nucleotide changes in the *NAT1* gene which formed eight allelic variants. The functional activities of most of these are still unknown, but NAT1*14 with the amino acid change $Arg^{187}Gln$ has reduced substrate affinity, and NATl*15 ($Arg^{187} -> Stop$) has a stop codon that prevents enzyme formation. Both have been found with low but significant frequencies (0.01 and 0.03, respectively) in Caucasian populations. However, Dhaini and Levy (51) reported a frequency of 0.238 for NAT1*14 in a Lebanese population. Other interethnic comparisons do not yet seem to exist.

Commonly used drugs (other than isoniazid) affected by NAT2 polymorphism were procainamide, hydralazine, dapsone, and sulfonamides with an increase of side effects in all cases. A selective substrate of NAT1 is *p*-aminosalicylic acid (PAS), but its genetic variation was never clinically important (52). Because of such lack of importance, more attention is often paid to the fact that various industrial chemicals with carcinogenic potential, and mutagenic heterocyclic amines, are substrates of both *N*-acetyltransferases (53). The presence or absence of these transferases will determine some incidences of cancer (54). Attempts have been made to ascribe cancer incidences in different populations to acetyltransferase differences (55).

3.3. Glucose-6-Phosphate Dehydrogenase (G-6-PD)

In the historical introduction to this book (Chapter 1, Sec. 2.1), I mentioned primaquine hemolysis and Glucose-6-Phosphate Dehydrogenase (G6PD) deficiency during the second world war in American soldiers. I did not

indicate the remarkable fact, that virtually only black soldiers developed this disease. The explanation for the interethnic difference was found later (56,57). The hemolysis-associated variant of the G-6-PD enzyme protected its carriers from malaria infection with *Plasmodium falciparum;* it therefore occurred mainly in populations that came from countries in which malaria was prominent (58).

The G-6-PD is a large structure with many genetic variants perhaps more than any other human protein (59–62). It is an almost ubiquitous cytosolic enzyme which catalyzes the first step in the hexose monophosphate pathway (62). Its most essential function is to produce the NADPH required to maintain the concentration of reduced glutathione (GSH) in the face of oxidative stress. The GSH together with catalase and glutathione peroxidase represent the defence against hydrogen peroxide, and this is particularly true in red blood cells.

There are different kinds of genetic variants of G-6-PD. Some of these generate serious disease in the form of chronic hemolytic anemia and therefore tend to be rare. This is not the pharmacogenetic problem; this results from the fact that hydrogen peroxide is often a byproduct arising during drug oxidation, and thus some drugs tend to produce destruction of the red cells, that is hemolysis (63). These deficiencies are the ones which are frequent in some populations because they protect against malaria. They tend to be symptomless unless a drug or an infection overwhelms the protective function of the enzyme; persons with these variants may go through life without ever realizing that they have G-6-PD deficiency. On the other hand, there may be pathological consequences for infants (see below). In addition to the life-threatening and the pharmacogenetic variants, there are neutral variants; Caucasians usually have a G-6-PD form called B+, while A+ is frequent in Africa. Approximately one tenth of the world population has one or other of the 300 odd functionally different variants of G-6-PD.

The gene *Gd* producing G-6-PD is located on the X chromosome at Xq28 (61). It means that the deficiency is sex linked, and it is mostly the males who show the deficiency. Males either do or do not have the deficiency, but females with their two X chromosomes may be heterozygous. In females during early embryogenesis, each cell eliminates at random one of the two X chromosomes so that in heterozygotes, approximately half the cells are normal and the other half are G-6-PD deficient. The protective effect of G-6-PD deficiency against malaria operates mostly via the heterozygous females.

However, the mechanism of the malaria protection is complex (64). It probably involves a sufficient survival time of particularly the malaria-infected heterozygous infant girls to allow them to develop immunological defences against malaria. This could mean that after puberty, their better state of health might render them more fertile than their less protected sisters. Since the occurrence of G-6-PD deficiency, of sickle cell anemia,

and of thalassemia, all seem to be related to the incidence of malaria, combinations of these traits are not uncommon.

About 20% of the X-chromosomes in American Blacks contain A+. It has about the same enzymatic activity as B+ but has higher electrophoretic mobility because aspartate replaces asparagine due to a A− > G mutation at nucleotide (nt) 376 (65). The G-6-PD deficiency typical of Blacks is A−; this has always one amino acid substitution in addition to the mutation which characterizes A+. The most common second mutation in A− is G− > A at nt 202. However, there may be mutations at nt 680 or nt 968 instead. Thus, at the molecular level, there are three A− variants instead of one. On the other hand, a few separately recorded variants were at first erroneously thought to be distinct from A−. In any case, all African deficiency variants are descendants of A+ (66).

One of the more common variants in the Orient is G-6-PD Canton, which has a mutation at nt 1376 (67); the same mutation turned out to be present in three other variants which had been thought to be independent. An equivalent situation occurred in Europe. G-6-PD Mediterranean is characterized by a C− > T substitution at nt 563. The four variants formerly distinguished as G-6-PD Cagliari, Dallas, Birmingham, and Sassari turned out to be identical with G-6-PD Mediterranean.

However, there is a peculiar observation in that virtually all persons from Southern Europe and the Near East who have G-6-PD Mediterranean also have the non-coding mutation at nt 1311. In addition, this noncoding mutation is common in the area and present in 20% of the Mediterranean population. Since subjects from India with G-6-PD Mediterranean did not have the 1311 mutation, Beutler (67) suggested that G-6-PD Mediterranean arose as independent mutations in Europe and in India.

Some pathological effects of G-6-PD deficiency are neonatal jaundice and hemolysis. Neonatal jaundice can be due to overproduction or undersecretion of bilirubin and thus can have many causes, but the likelihood of its occurrence is much enhanced in the presence of G-6-PD deficiency. Neonatal jaundice on the basis of G-6-PD deficiency is a relatively prominent problem in China and in Greece. The prevalence of G-6-PD deficiency in the Chinese and Malayan population of Singapore is in the order of 1.4% but about 25% of newborns with jaundice have the deficiency. The relative magnitude of these figures varies: In Greece with an adult deficiency rate of slightly less than 5%, the rate is about 15% of cases with neonatal jaundice. Such population differences are compatible with the recently provided evidence by Kaplan et al. (68) that it is the combination of deficiencies of G-6-PD and of glucuronyltransferase UDPGT1 (Gilbert syndrome) that causes neonatal jaundice.

A much investigated but not yet fully understood happening is the cause of favism, the hemolytic episode occurring in some Mediterranean people after eating fava beans; favism occurs only in persons with

G-6-PD deficiency but not in all persons with this deficiency (69). It seems that the fate of the glycoside divicine, a component of the fava been, varies metabolically or immunologically between persons.

3.4. Ethanol-Metabolizing Enzymes

In the current context, we will use the term "alcohol" not in its chemical sense in which there are many different alcohols with various (often poorly investigated) metabolic fates. This overview will be confined to the variable elements of ethanol metabolism.

Ethanol is mainly metabolized by alcohol dehydrogenase (ADH) to acetaldehyde which in turn is oxidized by aldehyde dehydrogenase (ALDH) to acetic acid. Genetic variation of ADH was discovered in 1964 by Von Wartburg et al. (70), that of aldehyde dehydrogenase in 1978 by Harada et al. (71). In the years since these discoveries, both variations have become targets of many investigations, and topics of specialized books (72–74). Alcohol dependence is influenced by many factors other than the dehydrogenases (75–77). Hence, the clinical significance of ADH variants is still unraveling. On the other hand, aldehyde dehydrogenase genotypes have been known for many years to influence drinking behavior in populations and alcohol-induced liver damage.

4. ALCOHOL DEHYDROGENASE

There are seven different ADH classes (gene families) which are expessed in different but overlapping sets of tissues (78). Here we will be concerned only with Class I ADH, consistig of three genes *ADH1*, *ADH2*, and *ADH3*, located on chromosome 4 between 4q21 and 4q24. Their products are referred to as alpha, beta, and gamma ADH, respectively (79). All these dehydrogenases are expressed in adult liver, beta, and gamma also in kidney.

The ADH Class I molecules are dimers, composed of two subunits. They may be homodimers composed of alpha/alpha, beta/beta, gamma/gamma units or they are the heterodimers alpha/beta, alpha/gamma, and beta/gamma. There is genetic variation of the beta and gamma subunits. There are three beta variants $beta_1$ $beta_2$, and $beta_3$. If for instance somebody is heterozygous having $beta_1$ and $beta_2$, the ADH is expected to consist of 10 dimers formed by random association. If this person were also heterozygous at the ADH_3 locus to produce $gamma_1$ and $gamma_2$ subunits, his or her alcohol dehydrogenase should consist of 18 different dimers. In vitro, the "atypical" ($beta_2$) variant is the better catalyst than is $beta_1$ (80), while $beta_3$ may be less efficient than $beta_1$.

The three subunits alpha or ADH1, beta or ADH2, and gamma or ADH3, are structurally similar. Each consists of 374 amino acids (81). In the context of ethnic comparisons, the most important difference is between

the "typical" subunit $beta_1$ which is predominant in Caucasians and the "atypical" $beta_2$ subunit which is prominent in Orientals. The occurrence of $beta_3$ in American Blacks is a comparatively new observation.

If the re-oxidation of NADH to cofactor NAD is the rate-limiting step of ethanol oxidation in vivo (82), the different capabilities of alcohol dehydrogenase may not be fully reflected in the elimination rate of ethanol. There are also the possibilities that the effect of the variants depends on ethanol concentration, or that $beta_2$ causes an initial spurt of ethanol oxidation in Orientals which is not seen in Caucasians who have the $beta_1$ allele and in Blacks who have the $beta_3$ allele.

Some functional clarification of these problems has been provided by recent comparative genotyping of alcoholics and non-alcoholics. Studies in Chinese (83–85) and in Japanese (86,87), supported by meta-analysis (88), indicated significantly reduced frequencies of $beta_2$ and $gamma_1$ in alcoholics, besides the usually present aldehyde dehydrogenase deficiency (see below). It means that the high rate of ethanol conversion to the toxic and unpleasant acetaldehyde tends to reduce ethanol consumption, particularly if acetaldehyde is slowly metabolized.

In short, the interethnic differences in the structure of ADH appear to have sufficient effects upon the fate of ethanol to be one of the determinants of alcoholism. Furthermore, one should not exclude the possibility that variation of ADH matters, for the fate of endogenous (89,90) or exogenous substrates (91,92). Ethanol is also metabolized by CYP2E1, but this enzyme is quantitatively less important than are the ADH (93).

5. ALDEHYDE DEHYDROGENASE

The clinical importance of ALDH2 deficiency for alcohol ingestion rests on the chemical reactivity and therefore toxicity of the ethanol-derived substrate acetaldehyde. It may produce facial flushing and a drop in blood pressure with tachycardia (94), i.e., effects which are perceived as an unpleasant sensation. These tend to occur after ethanol intake if the enzymatic removal of acetaldehyde is not fast enough. The unpleasantness, or even an embarassed reaction to the visual flushing, have been deterrents of excessive ethanol consumption and thereby of alcoholism. In Japan however, the deterrent effect of these sensations has been claimed to be gradually diminishing (95).

Traditionally established as being of clinical significance is variation of the mitochondrial aldehyde dehydrogenase referred to as ALDH2 (96). There are additional and different aldehyde dehydrogenases which are cytosolic enzymes and which seem to be also relevant in the present context (97), but this review will be confined to ALDH2, the best investigated enzyme of this group. Virtually all recent studies of alcoholism in Asians include data on ALDH2. The gene for this enzyme is located on chromosome 12 (98).

ALDH2 occurs mostly in liver and kidney and takes the form of a tetramer which consists normally of four identical subunits.

This tetrameric composition is important: There is an inactive genetic variant of ALDH2 which represents a point mutation. Glutamic acid at the 14th position from the C-terminus is substituted in the deficient enzyme by lysine (99). Even if the tetramer contains only one genetically inactive subunit, the whole tetramer is inactive. This means that there is enzyme deficiency even in the heterozygote, or in other words, ALDH2 deficiency is inherited as a dominant trait (100). Most published population comparisons therefore simply list the percentages of deficiency subjects.

Goedde and Agarwal (96) list test results from 29 different populations and a total of 3248 subjects. The data can be summarized by the statements that Central Asian, East Asian, and South-East Asian populations showed deficiencies in the order of 30%. The deficiency was absent in European, Near-East, and African populations. North American Indians showed deficiency rates of 2–5%, South American Indians of 40–45%. O'Dowd et al. (101) have shown that the functional enzyme deficiency in South American Indians must be due to a different mutation than the deficiency in Asians. This observation raises interesting questions regarding the biological significance of the mitochondrial aldehyde dehydrogenase.

6. CONCLUSIONS

More than inter-individual comparisons, interethnic studies require the use of genotyping rather than phenotyping methodology. The DNA analysis has shown that many, perhaps most, differences in drug response between populations have a genetic basis. At the same time, there are many non-pharmacogenetic or environmental factors capable of producing such differences, so that a genetic cause of interethnic differences cannot be automatically assumed.

Whenever inter-individual pharmacogenetic differences are observed, a proper search usually showed the existence of equivalent interethnic differences. It is not entirely clear whether this rule applies specifically to pharmacogenetics, or whether it affects all genetic variations to the same extent.

The number of mutant forms of an enzyme in a population varies widely from enzyme to enzyme. In most cases of pharmacogenetics, the reason for such enzyme differences is not clear. This is a striking fact, since the variability of drug-metabolizing enzymes often seems to have non-functional consequences in the absence of drugs.

In some cases, biological reasons for ethnic differences in drug-metabolizing capacity have been established, but usually they are not understood.

Genetic variation of a drug-metabolizing enzyme can have divergent functional consequences. Easiest to deal with are usually cases of absence of enzyme activity, even if it may occur as a dominant or recessive feature. Most complicated tend to be structural enzyme variants which may affect different ligands differently. Hence, fuctional predictions for a new population can often not be made on the basis of established experiences.

The same experiences as obtained with drugs in different populations apply principally to environmental chemicals. This opens new prospects for the investigation of environmentally caused intoxications, including cancer.

REFERENCES

1. Kalow W. Interethnic variation of drug metabolism. Trends Pharmacol Sci 1991; 12:102–107.
2. Wood AJ. Ethnic differences in drug disposition and response. Ther Drug Monit 1998; 20:525–526.
3. Kudzma EC. Drug response: all bodies are not created equal. Am J Nursing 1992; 92:48–50.
4. Lin KM, Poland RE, Nakasaki G (eds). Pshchopharmacology and Psychobiology of Ethnicity. Washington, DC, London, England:American Psychiatri Press Inc., 1993; 1–276.
5. Kalow W, Bertilsson L. Interethnic factors affecting drug response. Adv Drug Res 1994; 25:1–59.
6. Smith MW, Mendoza RP. Ethnicity and pharmacogenetics. Mt Sinai J Med 1996; 63:285–290.
7. Weber WW. Pharmacogenetics. Oxford: Oxford University Press, Inc. 1997; 1–400.
8. Bertilsson L, Kalow W. Interethnic differences in drug disposition and effects. In: Pacifici GM, Pelkonen O, eds. Interindividual Variability in Drug Metabolism in Humans. London: Taylor & Francis Ltd., 2001:15–75.
9. McCarthy LC, Davies KJ, Campbell DA. Pharmacogenetics in diverse ethnic populations—implications for drug discovery and development. Pharmacogenomics 2002; 3:493–506.
10. Kalow W. Ethnic differences in drug metabolism. Clin Pharmacokinet 1982; 7:373–400.
11. Nei M. Molecular Evolutionary Genetics New York: Columbia University Press, 1987; 1–512.
12. Cavalli-Sforza LL, Manozzi P, Piazza A. The history and geography of human genes. Princeton, NJ: Princeton University Press, 1994.
13. Johansson I, Oscarsson M, Yue Q-Y, Bertilsson L, Sjoqvist F, Ingelman-Sundberg M. Genetic analysis of the Chinese CYP2D locus. Characterization of variant CYP2D6 genes present in subjects with diminished capacity for debrisoquine hydroxylation. Mol Pharmacol 1994; 46:452–459.
14. Masimirembwa CM, Hasler JA, Bertilsson L, Johansson I, Ekberg O, Ingelman-Sundberg M. Phenotype and genotype analysis of debrisoquine hydroxylase (CYP2D6) in a black Zimbabwean population. Reduced enzyme

activity and evaluation of metabolic correlation of CYP2D6 probe drugs. Eur J Clin Pharmacol 1996; 51:117–122.
15. Anderson KE, Conney AH, Kappas A. Nutrition as an environmental influence on chemical metabolism in man. In: Kalow W, Goedde HW, Agarwal DP, eds. Ethnic Differences in Reactions to Drugs and Xenobiotics. New York: Alan R. Liss Inc., 1986; 39–54.
16. Herken H. Untersuchungen an Serumproteinen bei Eiweißmamgelernährung. Ärztl Wochenschr 1949; 4:297–302.
17. Herken H, Kalow W. Photometrische Bestimmung der Enzymatischen Novocain-Hydrolyse. Klin Wochenschr 1951; 29:90–91.
18. Schuetz EG. Induction of cytochromes P 450. Curr Drug Metab 2001; 2: 139–147.
19. Fuhr U, Strobl G, Manaut F, Anders EM, Sorgel F, Lopez-de-Brinas E, Chu DT, Pernet AG, Mahr G, Sanz F, et al. Quinolone antibacterial agents: relationship between structure and in vitro inhibition of the human cytochrome P450 isoform CYP1A2. Mol Pharmac 1993; 43:191–199.
20. Messina ES, Tyndale RF, Sellers EM. A major role for CYP2A6 in nicotine C-oxidation by human liver microsomes. J Pharmacol Exp Ther 1997; 282: 1608–1614.
21. Sellers EM. Pharmacogenetics and ethnoracial differences in smoking. JAMA 1998; 280:179–180.
22. Richardson AD, Piepho RW. Effect of race on hypertension and antihypertensive therapy. Int J Clin Pharmacol Ther 2000; 38:75–79.
23. Cushman WC, Reda DJ, Perry HM, Williams D, Abdellatif M, Materson BJ. Regional and racial differences in response to antihypertensive medication use in a randomized controlled trial of men with hypertension in the United States. Department of Veterans Affairs Cooperative Study Group on Antihypertensive Agents. Arch Intern Med 2000; 160:825–831.
24. Lefley HP. Culture and chronic mental illness. Hosp Commun Psychiatry 1990; 41:277–286.
25. Kalow W. Interethnic differences in drug response. In: Kalow W, Meyer UA, Tyndale R, eds. Pharmacogenomics. New York, Basel: Marcel Dekker Inc., 2001; 109–134.
26. Marez D, Legrand M, Sabbagh N, Lo Guidice JM, Spire C, Lafitte JJ, Meyer UA, Broly F. Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles, their frequencies and evolution. Pharmacogenetics 1997; 7:193–202.
27. Daly AK, Brockmoller J, Broly F, Eichelbaum M, Evans WE, Gonzalez FJ, Huang JD, Idle JR, Ingelman-Sundberg M, Ishizaki T, Jacqz-Aigrain E, Meyer UA, Nebert DW, Steen VM, Wolf CR, Zanger UM. Nomenclature for human CYP2D6 alleles. Pharmacogenetics 1996; 6:193–201.
28. Mahgoub A, Dring LG, Idle JR, Lancaster R, Smith RL. Polymorphic hydroxylation of debrisoquine in man. Lancet 1977; 2:584–586.
29. Eichelbaum M, Spannbrucker N, Steincke B, Dengler HJ. Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. Eur J Clin Pharmacol 1979; 16:183–187.

30. Eichelbaum M, Bertilsson L, Sawe BJ, Zekorn C. Polymorphic oxidation of sparteine and debrisoquine: related pharmacogenetic entities. Clin Pharmacol Ther 1982; 31:184–186.
31. Woolhouse NM, Eichelbaum M, Oates NS, Idle JR, Smith RL. Dissociation of co-regulatory control of debrisoquin/phenformin and sparteine oxidation in Ghanaians. Clin Pharmacol Ther 1985; 37:512–521.
32. Bertilsson L, Alm C, De Las Carreras C, Widen J, Edman G, Schalling D. Debrisoquine hydroxylation polymorphism and personality. Lancet 1989:555.
33. Kupfer A, Preisig R. Pharmacogenetics of mephenytoin: a new drug hydroxylation polymorphism in man. Eur J Clin Pharmacol 1984; 26:753–759.
34. Wedlund PJ, Aslanian WS, McAllister CB. Mephenytoin hydroxylation deficiency in Caucasians: frequency of a new oxidative drug metabolism polymorphism. Clin Pharmacol Ther 1984; 36:773–780.
35. Wilkinson GR, Guengerich FP, Branch RA. Genetic polymorphism od S-mephenytoin hydroxylation. In: Kalow W, ed. Pharmacogenetics of Drug Metabolism. New York: Pergamon Press Inc., 1992; 657–685.
36. Wrighton SA, Stevens JC, Becker GW. Isolation and characterization of human liver cytochrome P4502C19: correlation between 2C19 and S-mephenytoin 4′-hydroxylation. Arch Biochem Biophys 1993; 306:240–245.
37. Goldstein J, de Morals S. Biochemistry and molecular biologyof the human CYP2C subfamily. Pharmacogenetics 1994; 4:285–299.
38. Xie HG, Kim RB, Stein CM, Wilkinson GR, Wood AJ. Genetic polymorphism of (S)-mephenytoin 4′-hydroxylation in populations ofAfrican descent. Br J Clin Pharmacol 1999; 48:402–408.
39. Desta Z, Zhao X, Shin JG, Flockhart DA. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. Clin. Pharmacokinet 2002; 41: 913–958.
40. Kaneko A, Lum JK, Yaviong L, Takahashi N, Ishizaki T, Bertilsson L, Kobayakawa T, Bjorkman A. High and variable frequenciesof CYP2C19 mutations: medical consequences of poor drug metabolism in Vanuatu and other Pacific islands. Pharmacogenetics 1999; 5:581–590.
41. Andersson T, Regardh CG, Lou YC, Zhang Y, Dahl ML, Bertilsson L. Polymorphic hydroxylation of S-mephenytoin and omeprazole metabolism in Caucasian and Chinese subjects. Pharmacogenetics 1992; 2:25–31.
42. Ward SA, Helby NA, Skjelbo E. The activation of the biguanide antimalarial proguanil co-segragates with the mephenytoin oxidation polymorphism—a panel study. Br J Clin Pharmacol 1991; 31:689–692.
43. Sindrup SH, Brosen K, Hansen MG, Aaes-Jorgensen T, Overo KF, Gram LF. Pharmacokinetics of citalopram in relation to the sparteine and the mephenytoin oxidation polymorphisms. Ther Drug Monit 1993; 15:11–17.
44. Robitzek EH, Selikoff IJ, Ornstein GG. Chemotherapy of human tuberculosis with hydrazine derivatives of isonicontinic acid. Q Bull Sea View Hosp 1952; 13:27–51.
45. Bonicke R, Reif W. Enzymatische inaktivierung von Isonicotinsaure-hydrazid im menschlichen und tierischen Organisms. Arch exp Path u Pharmakol 1953; 220:321–333.

46. Hughes HB, Biehl JP, Jones AP, Schmidt LH. Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. Am Rev Tuberc 1954; 70:266–273.
47. Sunahara S, Urano M, Ogawa M. Genetical and geographical studies on isoniazid inactivation. Science 1961; 134:1530–1531.
48. Grant DM, Blum M, Beer M, Meyer UA. Monomorphic and polymorphic human arylamine *N*-acetyltransferases: a comparison of liver isozymes and expressed products of two cloned genes. Mol Pharm 1991; 39:184–191.
49. Vatsis KP, Weber WW. Structural heterogeneity of Caucasian *N*-acctyltransferase at the NAT1 gene locus. Arch Biochem Biophys 1993; 301:71–76.
50. Grant DM, Hughes NC, Janezic SA, Goodfellow GII, Chen IIJ, Gaedigk A, Yu VL, Grewal R. Human acetyltransferase polymorphisms. Mutat Res 1997; 376:61–70.
51. Dhaini HR, Levy GN. Arylamine *N*-acetyltransferase 1 (NAT1) Genotypes in a Lebanese population. Pharmacogenetics 2000; 10:79–83.
52. Hughes NC, Janezic SA, McQueen KL, Jewett MA, Castranio T, Bell DA, Grant DM. Identification and characterization of varian talleles of human acetyltransferase NAT1 with defective function using *p*-aminosalicylate as an in-vivo and in-vitro probe. Pharmacogenetics 1998; 8:55–66.
53. Hirvonen A. Chapter 20: polymorphic NATs and cancer predisposition. IARC Sci Publ 1999; 148:251–270.
54. Butcher NJ, Boukouvala S, Sim E, Minchin RF. Pharmacogenetics of the arylamine *N*-acetyltransferases. Pharmacogenomics J 2002; 2:30–42.
55. Brockton N, Little J, Sharp L, Cotton SC. *N*-acetyltransferase polymorphisms and colorectal cancer: a HuGE review. Am J Epidemiol 2000; 151:846–861.
56. Motulsky AG. Metabolic polymorphisms and the role of infectious diseases in human evoluation. Hum Biol 1960; 32:28.
57. Allison AC. Glucose 6-phosphate dehydrogenase deficiency in red blood cells of East Africans. Nature 1960; 186:531.
58. Vulliamy T, Mason P, Luzzatto L. The molecular basis of glucose-6-phosphate dehydrogenase deficiency. Trends Genet 1992; 8:138–143.
59. Yoshida A, Beutler E. Glucose-6-Phosphate Dehydrogenase. Orlando, San Diego: Academic Press Inc., 1986.
60. Luzzatto L. Glucose-6-phosphate dehydrogenase and other genetic factors -interacting with drugs. In: Kalow W, Goedde HW, Agarwal D, eds. Ethnic Differences in Reactions to Drugs and Xenobiotics. New York: Liss, 1986; 385–399.
61. Beutler E. Study of glucose-6-phosphate dehydrogenase: History and molecular biology. Am J Hematol 1993; 42:53–58.
62. Luzzatto L, Mehta A. Glucose 6-Phosphate Dehydrogenase Deficiency. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW, Vogelstein B, eds. The Metabolic and Molecular Bases of Inherited Disease. Vol. 111. New York, St. Louis, San Francisco, Auckland, Bogota, Lisbon, London, Madrid: McGraw-Hill Inc., 2001; 4517–4553.
63. Beutler E, Vulliamy TJ. Hematologically important mutations: Glucose-6-phosphate dehydrogenase. Blood Cells Mol Dis 2002; 28:93–103.

64. Usanga EA, Luzzatto L. Adaptation of plasmodium falciparum to glucose 6-phosphate dehydrogenase-deficient host red cells by production of parasite-encoded enzyme. Nature 1985; 313:793–795.
65. Hirono A, Beutler E. Molecular cloning and nuicleotide sequenceof cDNA for human glucose-6-phosphate dehydrogenase variant A(-). Proc Natl Acad Sci USA 1988; 85:3951–3954.
66. Kay AC, Kuhl W, Prchal J, Beutler E. The origin of glucose-6-phosphate-dehydrogenase (G-6-PD) polymorphisms in African-Americans. Am J Hum Genet 1992; 50:394–398.
67. Beutler E. The molecular biology of G-6-PD variants and other red cell enzyme defects. Annu Rev Med 1992; 43:47–59.
68. Kaplan R, Becker M, Broadman L. Isolated masseter muscle spasm versus generalized rigidity? Anesth Analg 1992; 141:156
69. Kattamis C. Favism: epidemiological and clinical aspects. In: Yoshida A, Beutler E, eds. Glucose-6-Phosphate Dehydrogenase. Orlando, San Diego, New York: Academic Press Inc., 1986:25–43.
70. Von Wartburg JP, Bethune JL, Vallee BL. Human liver alcohol dehydrogenase. Kinetic and physicochemical properties. Biochemistry 1964; 3:1775–1782.
71. Harada S, Agarwal DP, Goedde HW. Isozyme variations in acetaldehyde dehydrogenase (EC1.2.1.3.) in human tissues. Hum Genet 1978; 44:181–185.
72. Goedde HW, Agarwal DP. Aldehyde dehydrogenase polymorphism: molecular basis and phenotypic relationship to alcohol sensitivity. Alcohol Alcohol 1987 (suppl 1):47–54.
73. Crabbe JC, Harris RA (eds). The Genetic Basis of Alcohol and Drug Actions. New York, London: Plenum Press, 1991.
74. Weiner H, Lindahl R, Crabb DW, Flynn TG. Enzymology and Molecular Biology of Carbonyl Metabolism 6. New York, London: Plenum Press, 1997; 1–602.
75. Ferguson RA, Goldberg DM. Genetic markers of alcohol abuse. Clin Chim Acta 1997; 257:199–250.
76. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, Hesselbrock V, Schuckit MA, Bucholz K, Porjesz B, Li TK, Conneally PM, Nurnberger JIJR, Tischfield JA, Crowe RR, Cloninger CR, Wu W, Shears S, Carr K, Crose C, Willig C, Begleiter H. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet 1998; 81:207–215.
77. Eriksson CJ, Fucunaga T, Sarkola T, Chen WJ, Chen CC, Ju JM, Cheng AT, Yamamoto H, Kohlenberg-Muller K, Kimura M, Murayama M, Matsushita S, Kashima H, Higuchi S, Carr L, Viljoen D, Brooke L, Stewart T, Foroud T, Su TK, Whitfield JB. Functional relevance of human adh polymorphism. Alcohol Clin Exp Res 2001; 25 (supply 5):157–163.
78. Edenberg HJ, Brown CJ, Hur MW, Kotagiri S, Li M, Zhang L, Zhi X. Regulation of the Seven Human Alcohol Dehydrogenase Genes. In: Weiner H, Lindahl R, Crabb DW, Flynn TG, eds. Enzymology and Molecular Biology of Carbonyl Metabolism 6. New York: Plenum Press, 1996:339–345.
79. Jornvall H, Hoog JO. Nomenclature of alcohol dehydrogenases. Alcohol Alcohol 1995; 30:153–161.

80. Hurley TD, Ehrig T, Edenberg HJ, Bosron WF. Characterization of human alcohol dehydrogenases containing substitutions at amino acids 47 and 51. In: Weiner H, Wermuth B, Crabb DW, eds. Enzymology and Molecular Biology of Carbonyl Metabolism 3. New York: Plenum Press, 1990:271–275.
81. Bosron WF, Li TK. Genetic polymorphism of human liver alcohol and aldehyde dehydrogenases, and their relationship to alcohol metabolism and alcoholism. Hepatology 1986; 6:502–510.
82. Suddendorf RF. Research on alcohol metabolism among Asians and its implications for understanding causes of alcoholism. Public Health Rep 1989; 104:615–620.
83. Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li TK, Wang SP, Lin YT, Lu RB, Yin SJ. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. Am J Hum Genet 1991; 48:677–681.
84. Chao YC, Liou SR, Chung YY, Tang HS, Hsu CT, Li TK, Yin SJ. Polymorphism of alcohol and aldehyde dehydrogenase genes and alcoholic cirrhosis in Chinese patients. Hepatology 1994; 19:360–366.
85. Shen YC, Fan JH, Edenberg HJ, Li TK, Cui YH, Wang YF, Tian CH, Zhou CF, Wang RL, Zhou J, Zhao ZL, Xia GY. Polymorphism of ADH and ALDH genes among four ethnic groups in China and effects upon the risk for alcoholism. Alcohol Clin Esp Res 1997; 21:1272–1277.
86. Higuchi S. Polymorphisms of ethanol metabolizing enzyme genes and alcoholism. Alcohol Alcohol 1994 (Suppl 2):29–34.
87. Shiratori F, Tanaka Y, Yokosuka O, Imazeki F, Tsukada Y, Omata M. High incidence of ADH2*1/ALDH2*1 genes among Japanese alcohol dependents and patients with alcoholic liver disease. Hepatology 1996; 23:234–239.
88. Whitfield JB. Meta-analysis of the effects of alcohol dehydrogenase genotype on alcohol dependence and alcoholic liver disease. Alcohol Alcohol 1997; 32:613–619.
89. Strasser F, Huyng MN, Plapp BV. Activity of liver alcohol dehydrogenases of steroids. In: Weiner H, Lindahl R, Crabb DW, Flynn TG, eds. Enzymology and Molecular Biology of Carbonyl Metabolism 6. New York: Plenum Press, 1996; 313–320.
90. Kedishvili NY, Stone CL, Popov KM, Chernoff EAG. Role of alcohol dehydrogenases in steroid and retinoid metabolism. In: Weiner H, Lindahl R, Crabb DW, Flynn TG, eds. Enzymology and Molecular Biology of Carbonyl Metabolism 6. New York: Plenum Press, 1996; 321–329.
91. Kassam JP, Tang BK, Kadar D, Kalow W. In vitro studies of human liver alcohol dehydrogenase variants using a variety of substrates. Drug Metab Dispos 1989; 17:567–572.
92. Burnell JC, Bosron WF. Genetic polymorphism of human liver alcohol dehydrogenase and kinetic properties of the isoenzymes. In: Crow KE, Batt RD, eds. Human Metabolism of Alcohol. Vol. 2. Boca Raton, Florida: CRC Press, Inc., 1989; 65–75.
93. Maezawa Y, Yamauchi M, Toda G, Suzuki H, Sakurai S. Alcohol-metabolizing enzyme polymorphisms and alcoholism in Japan. Alcohol Clin Esp Res 1995; 19:951–954.

94. Higuchi S, Parrish KM, Dufour MC, Towle LH, Harford TC. The relationship between three subtypes of the flushing response and DSM-III alcohol abuse in Japanese. J Stud Alcohol 1992; 53:553–560.
95. Hasumura Y, Takeuchi J. Alcoholic liver disease in Japanese patients: a comparison with Caucasians. J Gastroenterol Hepatol 1991; 6:520–527.
96. Goedde HW, Agarwal DP. Pharmacogenetics of aldehydo dehydrogenase. In: Kalow W, ed. Pharmacogenetics of Drug Metabolism. Int Encycl Pharmacol Ther Sect 137. New York, Oxford: Pergamon Press, 1992:281–311.
97. Sophos NA, Vasiliou V. Aldehyde dehydrogenase superfamily: the 2002 update. Chem Biol Interact 2003; 144:5–22.
98. Vasiliou V. Appendix B, aldehyde dehydrogenase Genes. In: Weiner H, Lindahl R, Crabb DW, Flynn TG, eds. Enzymology and Molecular Biology of Carbonyl Metabolism 6. New York, London: Plenum Press, 1997:595–600.
99. Hempel J, Kaiser R, Jörnvall H. Mitochondrial aldehyde dehydrogenase from human liver: primary structure, differences in relation to the cytosolic enzyme and functional correlations. Eur J Biochem 1985; 153:13–28.
100. Singh S, Fritze G, Fang B, Harada S, Paik YK, Eckey R, Agarwal DP, Goedde HW. Inheritance of mitochondrial aldehyde dehydrogenase: genotyping in Chinese, Japanese and South Korean families reveals dominance of the mutant allele. Hum Genet 1989; 83:119–121.
101. O'Dowd BF, Rothhammer F, Israel Y. Genotyping of mitochondrial aldehyde dehydrogenase locus of native American Indians. Alcohol Clin Exp Res 1990; 14:531–533.
102. Lou YC, Ying L, Bertilsson L, Sjoqyist F. Low frequency of slow debrisoquine hydroxylation in a native Chinese population. Lancet 1987; 2:852–853.
103. Sommers DK, Moncrieff J, Avenant J. Non-correlations between debrisoquine and metoprolol polymorphism in the Venda. Hum Toxicol 1989; l8: 365–368.
104. Steiner E, Bertilsson L, Sawe J, Bertling I, Sjoqvist F. Polymorphic debrisoquine hydroxylation in 757 Swedish subjects. Clin Pharmacol Ther 1988; 44:431–435.
105. Benitez J, Llerena A, Cobaleda J. Debrisoquin oxidation polymorphism in a Spanish population. Clin Pharmacol Ther 1988; 44:74–77.
106. Primo-Parmo SL, Bartels CF, Wiersema B, van der Spek AF, Innis JW, LaDu BN. Characterization of 12 silent alleles of the human butyrylcholinesterase (BCHE) gene. Am J Hum Genet 1996; 58:52–64.
107. Deakin SP, James RW. Genetic and environmental factors modulating serum concentrations and activities of the anti-oxidant enzyme, paraoxanase-1. Clin Sci Lond 2004.
108. Grant DM, Hughes NC, Janezic SA, Goodfellow GH, Chen HJ, Gaedigk A, Yu VL, Grewal R. Human acetyltransferase polymorphisms. Mutat Res 1997; 376:61–70.
109. Schwartz GL, Turner ST. Pharmacogenetics of antihypertensive drug responses. Am J Pharmacogenomics 2004; 4(3):151–160.
110. Weinshilboum RM, Otterness DM, Szumlanski CL. Methylation Pharmacogenetics: catechol *O*-methyltransferase, thriopurine methyltransferase, and histamine *N*-methyltransferase. Annu Rev Pharmacol Toxicol 1999; 39:19–52.

111. Moore JW, Babidge WJ, Millard SH, Roediger WE. Thiolmethyl transferase activity in the human colonic mucosa: implications for ulcerative colitis. Gastroenterol Hepatol 1997; 12:678–684.
112. Alves S, Rocha J, Amorim A, Prata MJ. Tracing the origin of the most common thiopurine methyltransferase (TPMT) variants: preliminary data from the patterns of haplotype association with two CA repeats. Ann Human Genet 2004; 68:313–323.
113. Thomae BA, Rifki OF, Theobald MA, Eckloff BW, Wieben ED, Weinshilboum RM. Human catecholamine sulfotransferase (SULTIA3) pharmacogenetics: functional genetic polymorphism. J Neurochem 2003; 87(4):809–819.
114. Hengstler JG, Kett A, Arand M, Oesch-Barthlomowicz B, Oesch F, Pilch H, Tanner B. Glutathione S-transferase Tl and Ml gene defects in ovarian carcinoma. Cancer Lett 1998; 14:43–48.
115. Jourenkova-Mironova N, Voho A, Bouchardy C, Wikman H, Dayer P, Benhamou S, Hirvonen A. Glutathione *S*-transsferase GSTM1, GSTM3, GSTPl and GSTT1 genotypes and the risk of smoking-related oral and pharyngeal cancers. Int J Cancer 1999; 31:44–48.
116. Harris MJ, Coggan M, Langton L, Wilson SR, Board PG. Polymorphism of the Pi class glutathione *S*-transferase in normal populations and cancer patients. Pharmacogenetics 1998; 8:27–31.
117. Guillemette C. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. Pharmacogenomics J 2003; 3(3):136–158.
118. Lampe JW, Bigler J, Bush AC, Potter JD. Prevalence of polymorphisms in the human UDP-glucuronosyltransferase 2B family: UGT2B4(D458E), UGT2B7 (H268Y), and UGT2B15(D85Y). Cancer Epidemiol Biomarkers Prev 2000; 9(3): 329–333.
119. Patel M, Tang BK, Kalow W. (S)oxazepam glucuronidation is inhibited by ketoprofen and other substrates of UGT2B7. Pharmacogenetics (a) 1995; 5:43–49.
120. Court MH, Hao Q, Krishnaswamy S, Bekaii-Saab T, Al-Rohaimi A, Von Moltke LL, Greenblatt DJ. UDP-glucuronosyltransferase (UGT)2B15 pharmacogenetics: UGT2B15 D85Y genotype and gender are major determinants of oxazepam glucuronidation by human liver. J Pharmacol Exp Ther 2004; 310(2):656–665.
121. Kalow W, Tang BK, Kadar D, Inaba T. Distinctive patterns of amobarbital metabolites in man. Clin Pharmacol Ther 1978; 24:576–582a.
122. Morin S, Bodin L, Loriot MA, Thijssen HH, Robert A, Strabach S, Verstuyft C, Tregouet DA, Dubert L, Laurent-Puig P, Funck-Brentano C, Jaillon P, Beaune PH, Becquemont L. Pharmacogenetics of acenocoumarol pharmacodynamics. Clin Pharmacol Ther 2004; 75(5):403–414.
123. Beutler E, Vulliamy TJ. Hematologically important mutations: glucose-6-phosphate dehydrogenase. Blood Cells Mol Dis 2002; 28(2):93–103.
124. Laurenzana EM, Hassett C, Omiecinski CJ. Post-transcriptional regulation of human microsomal epoxide hydrolase. Pharmacogenetics 1998; 8:157–167.
125. Crabb DW, Dipple KM, Thomasson HR. Alcohol sensitivity, alcohol metabolism, risk of alcoholism, and the role of alcohol and aldehyde dehdrogenase genotypes. J Lab Clin Med 1993; 122:234–240.

126. Osier MV, Pakstis AJ, Soodyall H, Comas D, Goldman D, Odunsi A, Okonofua F, Parnas J, Schulz LO, Bertranpetit J, Bonne-Tamir B, Lu RB, Kidd JR, Kidd KK. A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. Am J Hum Genet 2002; 71(1):84–99.
127. Novoradovsky A, Tsai SJ, Goldfarb L, Peterson R, Long JC, Goldman D. Mitochondrial aldehyde dehydrogenase polymorphism in Asian and American Indian populations: detection of new ALDH2 alleles. Alcohol Clin Exp Res 1999; 19:1105–1110.
128. Balciuniene J, Syvanen AC, McLeod HL, Petterson U, Jazin EE. The geographic distribution of monoaminr oxidase haplotypes supports a bottleneck during the disperrsion of modern humans from Africa. J Mol Evol 2001; 52:157–163.
129. Garcia CC, Blair HJ, Seager M, Coulthard A, Tennant S, Buddles M, Curtis A, Goodship JA. Identification of a mutation in Sunapsin I, a synaptic vesicle protein, in a family with epilepsy. J Med Genet 2004; 41:183–186.
130. Calabrese EJ, Canada AT. Catalase: its role in xenobiotic detoxification. In: Kalow W, ed. Pharmacogenetics of Drug Metabolism. New York: Pergamon Press, 1992:397–410.
131. Canada AT, Calabrese EJ. Superoxide dismutase: its role in xenobiotic detoxification. In: Kalow W, ed. Pharmacogenetics of Drug Metabolism. New York: Pergamon Press Inc., 1992; 383–396.
132. Mitchell SC, Zhang AQ, Barrett T, Ayesh R, Smith RL. Studies on the discontinuous *N*-oxidation of trimethylamine among Jordanian, Ecuadorian and New Guinean populations. Pharmacogenetics 1997; 7:45–50.
133. Fernandez-Salguero PM, Sapone A, Wei X, Holt JR, Jones S, Idle JR, Gonzalez FJ. Lack of correlation between phenotype and genotype for the polymorphically expressed dihydropyrimidine dehydrogenase in a family of Pakistani origin. Pharmacogenetics 1997; 7:161–163.
134. Munoz S, Vollrath V, Vallejos MP, Miquel JF, Covarrubias C, Raddatz A, Chianale J. Genetic polymorphisms of CYP2D6, CYP1A1 and CYP2E1 in the South-Amerindian population in Chile. Pharmacogenetics 1998; 8: 343–351.
135. Nakajima M, Yokoi T, Mizutani M, Kinoshita M, Funayama M, Kamataki T. Genetic polymorphism in the 5′-flanking region of human CYP1A2 gene: effect on the CYP1A2 inducibility in humans. Nat Biotechnol 1999; 125: 803–808.
136. Saintot M, Malaveille C, Hautefeuille A, Gerber M. Interaction between genetic polymorphism of cytochrome P450-1B1 and environmental pollutants in breast cancer risk. Eur J Cancer Prev 2004; 13(1):83–86.
137. Pitarque M, von Richter O, Rodriguez-Antona C, Wang J, Oscarson M, Ingelman-Sundberg M. A nicotine *C*-oxidase gene (CYP2A6) polymorphism important for promoter activity. Hum Mutat 2004; 23(3):258–266.
138. Jacob RM, Johnstone EC, Neville MJ, Walton RT. Identification of CYP2B6 sequence variants by use of multiplex PCR with allele-specific genotyping. Clin Chem 2004.

139. Tanaca T, Kamiguchi N, Okuda T, Yamamoto Y. Characterization of the CYP2C8 active site by homology modeling. Chem Pharm Bull 2004; 52: 836–841.
140. Lee CR, Goldstein JA, Pieper JA. Cytochrome P450 2C9 polymorphisms: a comprehensive review of the in-vitro and human data. Pharmacogenetics 2002; 12:251–263.
141. Komai K, Sumida K, Kaneko H, Nakatsuka I. Identification of a new non-functional CYP2C18 allele in Japanese: substitution of T204 to A in exon2 generates a premature stop codon. Pharmacogenetics 1996; 6:117–119.
142. Daniel HI, Edeki TI. Genetic polymorphism of S-mephenytoin 4′-hydroxylation. Psychopharmacol Bull 1996; 32:219–230.
143. Meyer UA, Zanger UM. Molecular mechanisms of genetic polymorphisms of drug metabolism. Annu Rev Pharmacol Toxicol 1997; 37:269–296.
144. Howard LA, Ahluwalia JS, Lin SK, Sellers EM, Tyndale RF. CYP2E1*1D regulatory polymorphism: Association with alcohol and nicotine dependence. Pharmacogenetics 2003; 13:321–328.
145. Sata F, Sapone A, Elizondo G, Stocker P, Miller VP, Zheng W, Raunio H, Crespi CL, Gonzalez FJ. CyP3A4 allelic variants with amino acid substitutions in exon 7 and 12: evidence for an allelic variant with altered catalytic activity. Clin Pharmacol Ther 1999 in press.
146. Xie HG, Wood AJ, Kim RB, Stein CM, Wilkinson GR. Genetic variability in CYP3A5 and its possible consequences. Pharmacogenomics 2004; 5(3): 243–272.
147. Lo-Guidice JM, Allorge D, Cauffiez C, Chevalier D, Lafitte JJ, Lhermitte M, Broly F. Genetic polymorphism of the human cytochrome P450 CYP4B1: evidence for a non-functional allelic variant. Pharmacogenetics 2002; 12(5): 367–374.

9

Clinical Perspectives

Urs A. Meyer

Division of Pharmacology/Neurobiology, Biozentrum, University of Basel, Basel, Switzerland

David A. Flockhart

Division of Clinical Pharmacology, Indiana University School of Medicine, Indianapolis, Indiana, U.S.A.

The eventual clinical implications of the large efforts in pharmacogenomic research throughout the world described in other chapters may be profound and widespread, but the actual utility of pharmacogenetic knowledge in clinical practice to date remains limited and largely untested. Changes in clinical practice represented by changes in dose or in the drug administered that result in real changes in health outcomes would be important measures of progress toward the ''personalized medicine'' or ''precision prescriptions'' so frequently predicted. Such outcomes might include the avoidance of a specific toxicity or the achievement of a specific therapeutic effect, but, as with all measures of success in medicine, must also include real clinical outcomes.

It seems inevitable that significant improvements will continue to be made in the proportion of patients that are treated well, as is heralded by the recent reports of the effectiveness of a generally *ineffective* drug, gefitinib (Iressa™) in lung cancer in patients who carry sensitizing mutations of the tyrosin kinase domain of the epidermal growth factor receptor (EGFR) (1,2). We would not be surprised at this progress, if we recognized that such improvements are actually part of a continuum of improvement in the

quality and individualization of patient care. Individualization of therapy to an individual patient is not a new concept in medicine, but one central to its practice since the beginning. The writings of Hippocrates, Garrod, Jenner, and Osler all emphasize the centrality of treating the patient as an individual. There are multiple recent successes in targeting therapy to individual patients, supported by outcomes data and in wide clinical use already. To take the field of breast cancer as an example, such advances would include the evolution from the simple individualization of dose of chemotherapy by weight and body surface area to the use of estrogen and progesterone receptor status to target endocrine therapy by specific estrogen receptor modulators or aromatase inhibitors (3), to the use of the Gail index to identify patients most worth of preventive treatment (4), and to the use of BRCA1 mutations and HER2 status to target trastuzumab (Herceptin™) therapy (Table 1). We should not delude ourselves either by thinking that pharmacogenetic testing will usher in a new revolution in therapy or that the need for it represents an insult to our collective ability to individualize treatments in the past. ''Personalized medicine'' is nothing new. We have been doing it all along.

That said we have to admit that we have been doing it when we had time and when it was possible in busy clinical practice environments and

Table 1 Examples of Genetic Polymorphisms of Drug-Metabolizing Enzymes and Drug Targets for Which Genetic Tests May Affect Drug Choice or Dose and Reduce Adverse Reactions

Enzyme/target	Drug	Unwanted effect to be avoided
NAT2	Isoniazid	Hepatotoxicity
CYP2C9	Warfarin, acenocoumarol	Hemorrhage
CYP2D6	Antipsychotic drugs (e.g., thioridazine)	Extrapyramidal symptoms, cardiotoxicity
	Antidepressants (e.g., venlafaxine, TCAs)	ADRs in poor metabolizers, lack of response in ultrarapid metabolizers
	Codeine	Lack of analgesic effect
TPMT	Mercaptopurine, azathioprine	Myelosuppression, radiotherapy-related secondary tumors
UGT1A1	Irinotecan	Diarrhea, myelosuppression
HER2	Trastuzumab	Ineffective therapy
HCV genome	Interferon, ribivirin	Unnecessary prolongation of therapy

Abbreviations: NAT2, N-acetyltransferase 2; TPMT, thiopurine S-methyltransferase; UGT1A1, UDP-glucuronosyltransferase 1A1; HER2, human epidermal growth factor receptor 2 (ErbB2); HCV, hepatitis C virus; TCAs, tricyclic antidepressants; ADRs, adverse drug reactions.

that the pressures to adopt a "one dose fits all" approach are real. Quality has always mattered in health care, but now the quality of our performance as effective prescribers is being measured by hospitals, by health care systems and by government agencies. Pharmacogenomics should be seen as a stimulus to quality prescribing, a salve to the impersonalization of mass health care, and an invitation to advance the quality of our care *beyond* what is possible when we have time to do it and less medication to choose from. It should be designed as a "tool for quality improvement" along with the legions of administrators monitoring the number of new and return patients we see, the multiple digital assistants marketed to us, and the digitizing of our medical records.

A large number of potential pharmacogenetic tests have been proposed, and so it is necessary to have some simple means of separating, which tests are likely to be most valuable.

In guiding drug therapy, pharmacogenetic tests should help to prevent serious adverse reactions, reduce hospitalizations and mortality, and should thereby reduce health care costs, but also avoid the prescribing of drugs to patients who are likely not to respond. In fact, preventing ineffective treatment possibly is as effective in reducing costs of health care as adjusting doses to minimize adverse effects and improve efficacy. While tests that are robust, reliable and cheap to perform will inevitably have an edge over those that are not, it is also important that a test provides added value above what is currently available. If we can measure LDL-cholesterol as a metric to follow statin efficacy, or if the physician of a patient with hypertension can tell at the next weekly visit whether a diuretic or an angiotensin converting enzyme (ACE) inhibitor is working, it makes less sense to develop a pharmacogenetic test to predict the effects of statins and anti-hypertensive drugs, than to search for one that predicts the response to an antidepressant or cancer treatment, where our current predictive powers are more limited. However, we constantly learn and better understand the reasons for variability in drug response. For instance, the somewhat smaller reduction in total cholesterol and LDL cholesterol in some patients treated with pravastatin recently was explained by a common variant of the HMG-CoA reductase gene (5).

We should not be naïve to the fact that our scientific forbearers have developed numerous useful predictors of treatment response. To return to the example of breast cancer treatment: no pharmacogenetic test is likely to have value unless it can improve on the currently widely used and useful predictive clinical parameters: the number of lymph nodes, grade of tumor, etc. An excellent example of a rigorous approach has been provided by the group at the Netherlands Cancer Institute, who showed that a customized tumor gene expression array was a more powerful predictor of the outcome of disease in young patients with breast cancer than standard systems based on clinical and histologic criteria (6). Last, but NOT least, a pharmacogenetic

test, as any other diagnostic procedure, must be economically viable for companies that make test kits and for laboratories that do the testing. With these criteria in mind we have reviewed the approximately 45 pharmacogenetic situations or genes that have been associated with drug response in more than one clinical study (7–10). Monogenic traits were considered in regard to their possible clinical impact or the potential that they may consistently affect the choice and/or dose of a drug treatment. We come to the conclusion that the present data are convincing enough to consider genetic testing in order to guide therapy in the situations listed in Table 1.

Of the seven genes and situations, the role of CYP2D6 in the use of neuroleptics and antidepressants is discussed in other chapters, and the testing for the presence of HER2 in breast cancer or for variants of HCV is well established.

1. *N*-ACETYLTRANSFERASE (NAT2) AND ISONIAZID

One of the first pharmacogenetic traits to be recognized more than 50 years ago was the slow acetylation of the antituberculosis drug isoniazid now known as the polymorphism of *N*-acetyltransferase 2 (NAT2) and inherited as an autosomal recessive trait (reviewed in Refs. 11 and 12).

Isoniazid is the treatment of choice for latent tuberculosis infection and is included in most first-line therapy regimens in combination with rifampicin, ethambutol and pyrazinamide. However, acute or chronic hepatitis frequently develops in patients receiving these drugs, with an incidence of 1–36%, depending on different regimens and how one defines hepatic injury, from transient elevations of liver function tests to serious injury or even death (13). Of the various drugs used in the combination therapy, isoniazid appears to be the main drug to induce hepatotoxicity (14). Additional risk factors are alcohol consumption, advanced age and pre-existing chronic liver disease. In numerous studies, slow acetylators treated with isoniazid and rifampicin had a higher risk of hepatotoxicity than rapid acetylators and among patients with hepatotoxicity, slow acetylators had significantly higher serum aminotransferase levels (e.g., Refs. 13 and 15). Additional genetic risk factors were a homozygous ''wild-type'' genotype for CYP2E1 (CYP2E1c1/c1) conferring high activity to this enzyme (14). In other studies, the acetylator genotype was a good predictor of isoniazid plasma levels and isoniazid-induced hepatotoxicity (Ohno et al., 2004). These data suggest that genotype-derived dosage regimens, e.g., 400, 300, and 200 mg per day for slow (homozygous for two defective alleles), intermediate (heterozygous) or rapid (homozygous for two active alleles) acetylators, should be considered in future prospective studies.

The resurgence of tuberculosis in many countries as a serious threat because of a growing prevalence of drug resistance and its association with high risk patients such as HIV-seropositive individuals, convicts, homeless, or

drug users has reemphasized the role of genetic risk factors for the hepatotoxicity associated with antituberculous regimens.

The incidence of NAT2 slow acetylators can vary from 5% to 95%, depending on the geographic/ethnic origin of the population studied (16). In addition to isoniazid, the NAT2 polymorphism affects the pharmacokinetics of a wide variety of arylamine and hydrazine drugs and chemical carcinogens. They include sulfonamides such as salazosulfapyridine, the antiarrhythmic procainamide and the anticancer drug amonafide, for all of which dose adjustments according to the acetylator genotype or phenotype have been recommended, but are not part of common practice.

2. CYP2C9 AND WARFARIN

Warfarin is a commonly used anticoagulant that requires careful clinical management to balance the potentially lethal risks of over-anticoagulation and bleeding with the equally daunting risks of under anticoagulation and clotting. It is a legitimate target for a pharmacogenetic test because the surrogate for clinical effect, the international normalized ratio (INR) of the prothrombin time takes several days to reach its steady state, and improvements in both adverse outcomes and efficacy would be obtained if a pharmacogenetic test could more accurately predict dose.

The CYP2C9 is known to be the primary catalyst of the human metabolism of the S-enantiomer of warfarin (17), and the scientific association is sufficiently strong that warfarin was the substrate chosen for the crystallization of a human cytochrome CYP2C9 with its bound substrate (18). Multiple retrospective studies from investigators all over the world have shown that variant alleles of the CYP2C9 gene encoding the *2 [Arg144Cys] and *3 [Ile359Leu] alleles increase the anticoagulant effect of warfarin and decrease the mean daily dose required to maintain the INR of the prothrombin time within the target therapeutic range (19). These studies are based not only on persuasive case reports as illustrated in Figure 1, which demonstrates the difficulty encountered in establishing a stable warfarin dose in a patient who carried a CYP2C9 homozygous variant genotype. It is clear that the cost of caring for such a patient would likely outweigh the cost of a pharmacogenetic test. This anecdotal observation is supported by other case reports (20,21), and by a number of trials. The more persuasive of these are studies that take into account the other clinically available predictive factors, including specifically Vitamin K intake. Aithal et al. (22), the first group to report a clinical association between CYP2C9 genotype and warfarin dose (22) went on to study the contribution of CYP2C9 genotype, age, body size, and vitamin K and lipid status to warfarin dose requirements (23). The multiple linear regression models for warfarin dose indicated significant contributions from age ($r = 0.41$, $p < 0.001$), genotype ($r = 0.24$, $p < 0.005$), and age and genotype

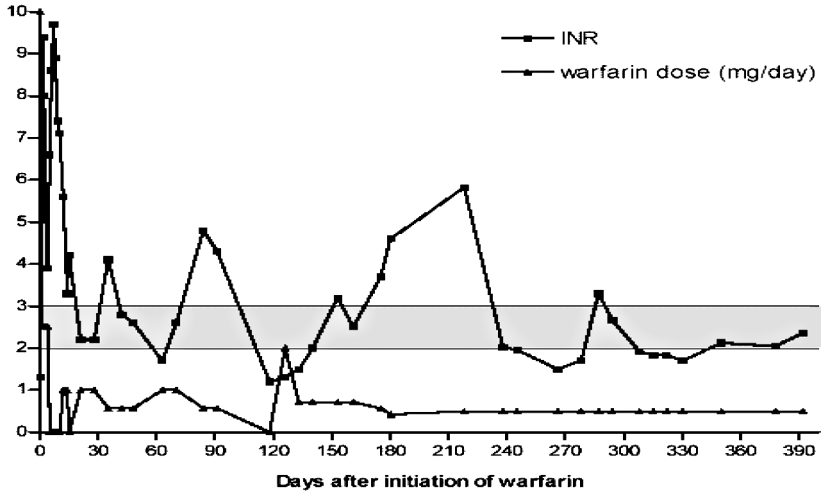


Figure 1 The INR values and warfarin doses of a patient with *CYP2C9*3* allelic variant requiring over 1 year to maintain a therapeutic INR.

together ($r = 0.45$, $p < 0.005$). The CYP2C9 genotype had a significant effect on S-warfarin clearance ($r = 0.34$, $p < 0.0001$) but none on R-warfarin clearance (23). In addition, Khan et al. studied the influence of dietary Vitamin K on dosage requirements and showed that, while dietary vitamin K had no effect, CYP2C9 genotype ($p = 2\%$) and age ($p < 1\%$) significantly contributed to inter-patient variability in warfarin dose requirements (20).

While data have been available for some time in relation to maintenance dose (24), the effect of CYP2C9 polymorphisms on dose requirements during the induction phase (25), when the danger of bleeding complications likely is greatest, has also been studied: patients with 2C9*2 or 2C9*3 variant alleles more frequently had INR values above the upper limit of the target range (3.0) (65% for 2C9*2/- and 66% for 2C9*3/- Vs. 33% for 2Cp*1/1; $p = 0.006$ and .012, respectively).

In elderly patients, a genetic influence on response to warfarin does exist as in younger patients. In work carried out by Siguret et al. (26). The CYP2C9 genotype was performed in 126 patients, with a mean age 87 ± 6 years. The mean daily dose of warfarin was 3.0 ± 1.4 mg, with 3.1 mg in patients with the wild type *1/*1 genotype ($n = 80$), 2.7 mg in *1/*2 heterozygotes ($n = 20$), 2.9 mg in *1/*3 heterozygotes ($n = 18$), 1.2 mg in *2/*2 homozygotes ($n = 2$), 2.3 mg in compound heterozygotes *2/*3 ($n = 6$).

Most importantly, the aggregated data prompted Higashi et al. (27) to design a trial to test whether CYP2C9 genotyping could predict outcomes of patients on warfarin therapy. In this retrospective cohort study, 200 patients receiving long-term warfarin therapy for various indications underwent

CYP2C9 genotyping and were evaluated by outcome measures, including anticoagulation status, measured by time to therapeutic INRs, rate of above-range INRs, and time to stable warfarin dosing or to serious or life-threatening bleeding events. They found that the mean maintenance dose varied significantly among the six-genotype groups ($^*1/^*1$ [$n = 127$], $^*1/^*2$ [$n = 28$], $^*1/^*3$ [$n = 18$], $^*2/^*2$ [$n = 4$], $^*2/^*3$ [$n = 3$], and $^*3/^*3$ [$n = 5$]) (by the Kruskall–Wallis test, $\chi^2(5) = 37.348$; $p < 0.001$). Compared with patients with the wild-type genotype, patients with at least one variant allele had an increased risk of above-range INRs of 1.40 (95% CI, 1.03–1.90). The variant group also required more time to achieve stable dosing (HR, 0.65; 95% CI, 0.45–0.94), with a median difference of 95 days ($p = 0.004$). In addition, patients with a variant genotype had a significantly increased risk of a serious or life-threatening bleeding event (HR, 2.39; 95% CI, 1.18–4.86).

More *prospective* studies are needed to test the feasibility and cost effectiveness of using algorithms based on these parameters for adjusting initial warfarin dose to meet individual needs, but the data at present make a strong case for the use of CYP2C9 genotyping testing prior to warfarin treatment, especially in the elderly.

The main alternatives to warfarin, the coumarin derivatives acenocoumerol and phenoprocoumon, are widely or exclusively used instead of warfarin in certain European countries (28). Again the S-enantiomers of these closely related chemicals are substrates of CYP2C9. The presence of even one copy of CYP2C9*3 profoundly decreases the metabolic clearance of S-acenocoumarol. S-acenocoumarol, which normally is clinically inactive will now exert the main anticoagulant activity. The *CYP2C9*3* allele therefore is related to low-dose requirements of racemic acenocoumerol, a higher frequency of over anticoagulation and an unstable anticoagulant response (29–31). Phenoprocoumon not significantly affected in its kinetics by the CYP2C9 polymorphism and appears to be a clinically useful alternative to warfarin in patients carrying *CYP2C9*2* and *3 alleles (32).

3. THIOPURINE S-METHYLTRANSFERASE AND MERCAPTOPURIN

The therapy of cancer almost always involves multiple drugs with considerable toxicity. Pharmacogenomic approaches to cancer therapy have already briefly been discussed at the beginning of this chapter in regard to breast cancer. In general, these strategies include variations in germline DNA (e.g., genetic polymorphisms), acquired somatic mutations in tumor cells (e.g., sensitizing mutations in the tyrosine kinase domain of the EGFR gene that is the target of genitifib) or variations in RNA expression. Perhaps one of the best studied examples for the application of pharmacogenomic strategies to prevent adverse drug reactions is the polymorphism of the thiopurine

S-methyltransferase (TPMT) gene, which has been extensively reviewed (e.g., Ref. 33). The TPMT catalyses the S-methylation of thiopurine drugs such as mercaptopurine and its prodrug azathioprine. These drugs are successfully used to treat acute lymphoblastic leukemia (ALL) of childhood. Moreover, gastroenterologists prescribe thiopurine drugs as second-line (off-label) therapy for Crohn's disease and ulcerative colitis (34,35). Because methylation by TPMT is the pre-dominant pathway for inactivation of thiopurines, patients with TPMT deficiency accumulate active thioguanine nucleotides and this can lead to severe and life-threatening hematological toxicity. The TPMT activity in erythrocytes is trimodally distributed among Europeans, European-Americans, and African-Americans, which corresponds well to the genotypes or the respective presence of 2, 1, or 0 functional TPMT alleles. In fact, the concordance rate between TPMT genotype and phenotype is >98% (36). Twenty mutant alleles of TPMT have been associated with low TPMT activity and three of these variants (TPMT*2, TPMT*3A, and TPMT*3C) account for approximately ~95% of low TPMT activity phenotypes (36). Approximately 1 in 150–300 individuals is homozygous for inactive TPMT alleles, approximately 10% of patients are heterozygous and have intermediate activity and ~90% are normal or high methylators in a Northern European Caucasian population. Interestingly, a subpopulation of ultrarapid TPMT metabolizers also was identified in these population studies (36).

Because of the genotype–phenotype concordance and the severe toxicity associated with high concentrations of thioguanine nucleotides, several cancer centers routinely genotype patients for TPMT mutant alleles and use genotype-derived algorithms for dosing. Intermediate metabolizers receive ~65% and poor metabolizers 5–10% of standard doses of mercaptopurine (37–39). Dose reductions in patients with variant TPMT alleles lead to similar or superior survival compared to patients with wild-type alleles. Expectedly, there is considerable variability in the frequency of TPMT alleles in populations of different ethnic origins.

Despite the obvious robustness and cost-effectiveness of TPMT genotyping (40), additional factors and "resistance of physicians to change" so far have limited genotyping to specialized cancer centers. The arguments that full TPMT activity does not preclude possible myelotoxicity with multi-drug regimens and that partial TPMT activity does not always mandate reduced doses appear not very convincing. The information from genotyping obviously is only one type of information to be used in making decisions on drug and dosage regimens in following a patient with ALL, but it has an important function of telling the physician which patients have to be monitored more closely.

Of major concern is a report of an increased incidence of secondary brain tumors after radiotherapy in children with decreased TPMT activity phenotypes and/or high concentrations of thioguanine nucleotides in blood

cells (37–39). The implications for therapeutic decisions regarding prophylactic radiotherapy in ALL therefore must be further investigated.

4. UDP-GLUCURONOSYLTRANSFERASE (UGT) 1A1 AND IRINOTECAN

Results from several recently published trials suggest that patients who are homozygous for a UGT gene variant known as UGT1A1*28 (the "7/7" genotype) are at greater risk for irinotecan-induced severe diarrhea or neutropenia (41). Irinotecan is a camptothecin analog and inhibits topoisomerase I as an antineoplastic principle. It is used to treat several solid tumors. The disposition of irinotecan is quite complex and involves numerous metabolic enzymes and transport proteins. The SN-38 is the active metabolite of irinotecan and is eliminated via UGT1A1 conversion to SN-38G, an inactive glucuronide cleared via biliary excretion.

Reduced activity of UGT1A1 is linked to an approximately fourfold increased risk of severe toxicity, including dose-limiting diarrhea and neutropenia. Significant correlations between patients carrying one or two copies of the *UGT1A1*28* allele and reduced UGT1A1 expression and reduced SW38 glucuronidation have been well documented.

More than 50 mutations in UGT1A1 have been reported by Tukey and Strassburg, 2002 (42) many of which are found in patients with Gilbert's syndrome, a form of mild non-hemolytic unconjugated hyperbilirubinemia. The most common mutant gene is *UGT1A1*28*, it contains seven dinucleotide repeats in the TATA box of the promoter [A(TA) $_7$TAA] instead of the normally six repeats, which leads to approximately 70% reduction of transcriptional activity. Many rare mutations also lead to Gilbert's syndrome, and individuals with this syndrome are pre-disposed to SN-38 initiated toxicity. Although, as always, a number of additional factors influence the toxicity of SN-38 in the intestine and bone narrow, assessment of the presence of the *UGT1A1*28* allele in patients prior to irinotecan treatment will allow to start with lower dose or change to alternative therapies.

5. CYP2D6 AND CODEINE

The clinical impact of the CYP2D6 polymorphism has been the subject of numerous reviews (43,44). It involves a large number of drugs, including many antidepressants and neuroleptics, antiarrhythmics, and many others commonly used drugs. In fact, recent recommendation for adjusting the doses for a number of antidepressant and neuroleptic drugs in the four genotypes/phenotypes poor metabolizers, extensive metabolizers, intermediate metabolizers, and ultrarapid metabolizers have been proposed (45). Prospective studies are planned to evaluate these recommendations, as genotyping tests for the multiple mutations of CYP2D6 are available.

Here we discuss the striking differences in the responses to opioids that are associated with the CYP2D6 polymorphism (46). Dextromethorphan, codeine, hydrocodone, oxycodone, ethylmorphine, and dihydrocodeine all are dealkylated by polymorphic CYP2D6. The polymorphic O-demethylation of codeine is of clinical importance when this drug is given as an analgesic. About 5% of codeine is O-demethylated to morphine (47), and this pathway is deficient in poor metabolizers. Poor metabolizers therefore experience little analgesic benefit from treatment with codeine. Similarly, respiratory, psychomotor, and pupillary effects of codeine are decreased in poor metabolizers compared with extensive metabolizers. Codeine is frequenctly recommended as a drug of first choice for treatment of chronic severe pain. Physicians must appreciate that no analgesic effect is to be expected in the 5–10% of Caucasians who are of the poor metaboliser phenotype, or who are extensive metabolizers receiving concomitant treatment with a potent inhibitor of CYP2D6. No morphine or morphine metabolites were detected in plasma when codeine was coadministered with quinidine (48,49). Although codeine may seem on the surface a poor candidate for a pharmacogenetic test, since the patient knows whether the medicine has worked or not, in fact there are many situations where analgesia is imperfect and even in situations where a patient can tell that codeine is having no analgesic benefit, his or her physician may not be aware, and self-reporting about pain is a notoriously variable and subjective phenomenon. It follows that the test may be valuable as a means of indicating which patients should not receive codeine as an analgesic, and who would most likely benefit.

6. THERAPEUTIC LESSONS

Pharmacogenetics has provided a number of therapeutic lessons that make us understand clinical drug response and it has influenced the drug development process. Among the therapeutic lessions are that most drug effects vary considerably from person to person and that all drug effects are influenced by genes. But it has also been realized that the most drug responses and toxicities are influenced by many genes interacting with environmental and behavioral factors. Genetic polymorphisms of single genes, including mutations in coding sequences, gene duplications, gene deletions, and regulatory mutations affect numerous drug-metabolizing enzymes. Several cytochrome-P450 enzymes, *N*-acetyltransferases 2 (NAT2), TPMT, and a UDP-glucuronosyltransferases (UDP-GT) are the examples discussed here. Individuals who possess these polymorphisms are at risk of experiencing more adverse drug reactions or inefficacy of drugs at usual doses. Genetic polymorphisms of drug targets and drug transporters also are increasingly recognized (receptors, ion channels, and growth factors) as causing variation in drug responses, but they have not been studied enough in regard to their clinical importance that would advocate genotype-based

dose adjustments. Several targets of cancer therapy, for example, the epidermal-growth-factor receptor, respond to treatment only in subgroups of patients who carry sensitizing mutations of these targets. Finally, the frequency of variation of drug effects, whether multifactorial, or genetic, varies considerably in populations of different ethnic origins.

7. FUTURE PERSPECTIVES

One of the major challenges in the future is the interpretation of multigenic and multifactorial influences on drug responses. Indeed, as already mentioned, most drug effects and treatment outcomes, or the individual risk for drug inefficacy or toxicity are due to complex interactions between genes and the environment. Environmental variables include nutritional factors, concomitantly administered drugs, disease, and many other factors including lifestyle influences such as smoking and alcohol consumption. These factors act in concert with several individual genes that code for pharmacokinetic and pharmacodynamic determinants of drug effects such as receptors, ion channels, drug-metabolizing enzymes, and drug-transporters. The challenge will be to define polygenic determinants of drug effects and to use a combination of genotyping and phenotyping tests to assess environmental influences.

The increasing use of the term pharmacogenomics reflects the evolution of pharmacogenetics into the study of the entire spectrum of genes that determine drug response, including the assessment of the diversity of the human genome sequence and its clinical consequences. Rapid sequencing and genotyping of SNPs will have a major role in associating sequence variations with heritable clinical phenotypes of drug or xenobiotic response. The SNPs occur approximately once every 300–3000 bp if one compares the genomes of two unrelated individuals and represent 90–95% of all variant DNA sites. Any two individuals thus differ at approximately 3 to 10 million base pairs.

How can we use this information to predict drug responses particularly with the view that in a few years technologies will be available to sequence an entire human genome in a few hours and at a reasonable prize? Once a large number of SNPs and their frequencies in different populations are known, they can be used to correlate an individual's genetic "fingerprint" with the probable individual drug response. It has been proposed that high density maps of SNPs or the so-called haplotype blocks in the human genome might allow the use of these SNPs as markers of xenobiotic responses even if the target remains unknown, providing a "drug–response profile" that is associated with contributions from multiple genes to a response phenotype. A recent "proof of concept" was provided by Xu et al. 2004 (50) for tranilast-induced hyperbilirubinemia. Whole genome screening for regions of linkage disequilibrium associated with this adverse

effect was used to identify three SNPs of *UGT1A1* gene to be responsible for this drug-induced adverse reaction. In practice, and because of the complexities of defining disease phenotypes and clinical outcomes, the validity of this concept is limited.

Genomic technologies also include methods to study the expression of large groups of genes and indeed the entire complement of products (mRNAs) of a genome. Most drug actions produce changes in gene expression in individual cells or organs. This provides a new perspective for the way in which drugs interact with the organism and also provides a measure of the drug's biological effects.

For instance, numerous drugs induce their own metabolism and the metabolism of other drugs by interacting with nuclear receptors such as arylhydrocarbon receptor (AhR), peroxisome proliferator activated receptor (PPAR), pregnane $\times$ receptor (P $\times$ R), and constitutive androstane receptor (CAR). These receptors act as "xenosensors" and transcription factors that activate a response that includes increased biotransformation of drugs (reviewed in Ref. 51). The phenomenon of induction has major clinical consequences such as altered kinetics, drug–drug interaction or changes in hormone and carcinogen metabolism. Genomics is providing the technology to better analyze these complex multifactorial situations and to obtain individual genotypic and gene expression information to assess the relative contributions of environmental and genetic factors to variations in drug response.

8. THE CLINICAL POTENTIAL OF PHARMACOGENETICS

Why is pharmacogenetics so rarely applied in clinical practice, in spite of well-established genetic polymorphisms and available genotyping methods? Numerous reasons for the slow acceptance of pharmacogenetic principles have been brought forward (9,10,52,53). The lack of large prospective studies to evaluate the impact of genetic variation on drug therapy is one reason for the slow acceptance of these principles. On the other hand, pharmacogenetic information is only reluctantly included in product information or drug data sheets alerting the physician to dosing problems. A recent search for pharmacogenetic information in the prescribing information available to physicians provided the following bleak results (54). Seventy-six drug package inserts (PIs) from the Physician's Desk Reference (PDR) contained pharmacogenomic data. The gene usually was either a drug-metabolizing enzyme or the information was related to the variability in viral genomes as predictors of response to antiviral therapy or drug resistance. Information to guide treatment decisions was found in only 25 PIs, representing 22 drugs. Of these four were ranked in the top 200 prescribed drugs (celexocib, fluoxetine, pantoprazole, and divalproex sodium).

Advice for treatment decisions based on specific genetic conditions were found in four PIs, namely that prolastin (α1-proteinase inhibitor) is not indicated in patients with certain α1-antitrypsin deficiency phenotypes, trastuzumab indicated only in patients with overexpression of the HER2 protein, tretinoin, and imatinib are to be given only in patients with either a specific subtype of acute myelogenous leukemia or Philadelphia chromosome-positive chronic myeloid leukemia, respectively.

Surprisingly, information on increased risk for potentially life-threatening adverse effects or treatment failures with conditional recommendations for genetic evaluation were found for only four drugs, namely recombinant factor X, somatotropin, divalproex sodium, and valproic acid. Only for one drug, thioridazine, did the PI contain a contraindication for a genetic subgroup, namely CYP2D6 poor metabolizers, which may develop QTc (prolonged heart-rate-corrected QT interval) interval prolongation in the electrocardiogram and develop ventricular arrhythmia. Clearly, PIs at present do not contain useful information for gene-guided dose adjustments or therapeutic decisions. In particular, there are no explicit recommendations for drug dosing in TPMT deficient patients in the PIs for mercaptopurine or azathioprine or in the PI for warfarin for patients with low activity alleles of CYP2C9. Similarly, the PI for irinotecan does not contain information on the risk of patients with UGT1A1 deficiency. A major effort is underway at the Food and Drug Administration to correct these obvious deficiencies in alerting physicians to potential problems, and this has included the first approval of a pharmacogenetic test: an oligonucleotide microarray test for CYP2D6 and CYP2C19 genotypes in December of 2004. In the future, *not* performing a pharmacogenetic test may have legal consequences.

Pharmacogenomics offers the potential to provide better health care through improved rational prescribing. We believe that is gradual acceptance in clinical practice will contribute to the education of health care professionals as prescribers. In addition, we must recognize an increasingly important source of pressure to improve pharmacotherapy: increasingly educated patients will come to expect the application of genomics and other technologies to drug selection and dosage when possible. The personalized medicine that many view as a new goal is actually what physicians always intended, and is becoming what patients, and the regulatory bodies that protect them, expect.

ACKNOWLEDGMENTS

Work in the authors' laboratories was supported in part by grants from the Swiss National Science Foundation (U.A. Meyer) and by the Pharmacogenetics Research Network Grant (U-01-GM061373) from the National Institute of General Medical Sciences, Bethesda, MD, U.S.A. (D.A. Flockhart).

The authors thank George A. Davis, University of Kentucky, Lexington, KY, U.S.A., for the data of Figure 1.

REFERENCES

1. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004; 350:2129–2139.
2. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004; 304:1497–1500.
3. Baum M. Tamoxifen—the treatment of choice. Why look for alternatives? Br J Cancer 1998; 78 (suppl 4):1–4.
4. Freedman AN, Graubard BI, Rao SR, McCaskill-Stevens W, Ballard-Barbash R, Gail MH. Estimates of the number of US women who could benefit from tamoxifen for breast cancer chemoprevention. J Natl Cancer Inst 2003; 95:526–532.
5. Chasman DI, Posada D, Subrahmanyan L, Cook NR, Stanton VP Jr, Ridker PM. Pharmacogenetic study of statin therapy and cholesterol reduction. Jama 2004; 291:2821–2827.
6. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002; 347:1999–2009.
7. Evans WE, McLeod HL. Pharmacogenomics–drug disposition, drug targets, and side effects. N Engl J Med 2003; 348:538–549.
8. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. Nat Rev Genet 2003; 4:937–947.
9. Nebert DW, Jorge-Nebert L, Vesell ES. Pharmacogenomics and ''individualized drug therapy': high expectations and disappointing achievements. Am J Pharmacogenomics 2003; 3:361–370.
10. Weinshilboum R, Wang L. Pharmacogenomics: bench to bedside. Nat Rev Drug Discov 2004; 3:739–748.
11. Evans DAP. Genetic Factors in Drug Therapy. Clinical and Molecular Pharmacogenetics. Cambridge: Cambridge University Press, 1993.
12. Weber WW. The acetylator genes and drug response. New York: Oxford University Press, 1987.
13. Huang YS, Chern HD, Su WJ, Wu JC, Lai SL, Yang SY, Chang FY, Lee SD. Polymorphism of the N-acetyltransferase 2 gene as a susceptibility risk factor for antituberculosis drug-induced hepatitis. Hepatology 2002; 35:883–889.
14. Huang YS, Chern HD, Su WJ, Wu JC, Chang SC, Chiang CH, Chang FY, Lee SD. Cytochrome P450 2E1 genotype and the susceptibility to antituberculosis drug-induced hepatitis. Hepatology 2003; 37:924–930.

15. Ohno M, Kubota R, Yasunaga M, Yokota S, Azuma J. Population pharmacokinetics and pharmacogenetics trial of isoniazid. Abstract PO-338, Clin and Exp Pharmacol and Physiol 2004; 31: (suppl.), A138.
16. Meyer UA, Zanger UM. Molecular mechanisms of genetic polymorphisms of drug metabolism. Annu Rev Pharmacol Toxicol 1997; 37:269–296.
17. Kunze KL, Wienkers LC, Thummel KE, Trager WF. Warfarin-fluconazole. I. Inhibition of the human cytochrome P450-dependent metabolism of warfarin by fluconazole: in vitro studies. Drug Metab Dispos 1996; 24:414–421.
18. Williams PA, Cosme J, Ward A, Angove HC, Matak Vinkovic D, Jhoti H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature 2003; 424:464–468.
19. Ablin J, Cabili S, Eldor A, Lagziel A, Peretz H. Warfarin therapy is feasible in CYP2C9*3 homozygous patients. Eur J Intern Med 2004; 15:22–27.
20. Khan T, Kamali F, Daly A, King B, Wynne HA. Warfarin sensitivity: be aware of genetic influence. Age Ageing 2003; 32:226–227.
21. Redman AR, Dickmann LJ, Kidd RS, Goldstein JA, Ritchie DM, Hon YY. CYP2C9 genetic polymorphisms and warfarin. Clin Appl Thromb Hemost 2004; 10:149–154.
22. Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. Lancet 1999; 353:717–719.
23. Kamali F, Khan TI, King BP, Frearson R, Kesteven P, Wood P, Daly AK, Wynne H. Contribution of age, body size, and CYP2C9 genotype to anticoagulant response to warfarin. Clin Pharmacol Ther 2004; 75:204–212.
24. Scordo MG, Pengo V, Spina E, Dahl ML, Gusella M, Padrini R. Influence of CYP2C9 and CYP2C19 genetic polymorphisms on warfarin maintenance dose and metabolic clearance. Clin Pharmacol Ther 2002; 72:702–710.
25. Peyvandi F, Spreafico M, Siboni SM, Moia M, Mannucci PM. CYP2C9 genotypes and dose requirements during the induction phase of oral anticoagulant therapy. Clin Pharmacol Ther 2004; 75:198–203.
26. Siguret V, Gouin I, Golmard JL, Geoffroy S, Andreux JP, Pautas E. Cytochrome P450 2C9 polymorphisms (CYP2C9) and warfarin maintenance dose in elderly patients. Rev Med Intern 2004; 25:271–274.
27. Higashi MK, Veenstra DL, Kondo LM, Wittkowsky AK, Srinouanprachanh SL, Farin FM, Rettie AE. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. JAMA 2002; 287:1690–1698.
28. Daly AK, King BP. Pharmacogenetics of oral anticoagulants. Pharmacogenetics 2003; 13:247–252.
29. Tassies D, Freire C, Pijoan J, Maragall S, Monteagudo J, Ordinas A, Reverter JC. Pharmacogenetics of acenocoumarol: cytochrome P450 CYP2C9 polymorphisms influence dose requirements and stability of anticoagulation. Haematologica 2002; 87:1185–1191.
30. Thijssen HH, Drittij MJ, Vervoort LM, de Vries-Hanje JC. Altered pharmacokinetics of R- and S-acenocoumarol in a subject heterozygous for CYP2C9*3. Clin Pharmacol Ther 2001; 70:292–298.

31. Thijssen HH, Verkooijen IW, Frank HL. The possession of the CYP2C9*3 allele is associated with low dose requirement of acenocoumarol. Pharmacogenetics 2000; 10:757–760.
32. Visser LE, van Vliet M, van Schaik RH, Kasbergen AA, De Smet PA, Vulto AG, Hofman A, van Duijn CM, Stricker BH. The risk of overanticoagulation in patients with cytochrome P450 CYP2C9*2 or CYP2C9*3 alleles on acenocoumarol or phenprocoumon. Pharmacogenetics 2004; 14:27–33.
33. Krynetski E, Evans WE. Drug methylation in cancer therapy: lessons from the TPMT polymorphism. Oncogene 2003; 22:7403–7413.
34. Nielsen OH, Vainer B, Rask-Madsen J. Review article: the treatment of inflammatory bowel disease with 6-mercaptopurine or azathioprine. Aliment Pharmacol Ther 2001; 15:1699–1708.
35. Sandborn W, Sutherland L, Pearson D, May G, Modigliani R, Prantera C. Azathioprine or 6-mercaptopurine for induction of remisssion in Crohn's disease. The Cochrane Database of Systemic Reviews 1998, Issue 3. Art No.: CD000545. DOI:10.1002/14651858.CD000545.
36. Schaeffeler E, Fischer C, Brockmeier D, Wernet D, Moerike K, Eichelbaum M, Zanger UM, Schwab M. Comprehensive analysis of thiopurine S-methyltransferase phenotype–genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. Pharmacogenetics 2004; 14:407–417.
37. Relling MV, Hancock ML, Boyett JM, Pui CH, Evans WE. Prognostic importance of 6-mercaptopurine dose intensity in acute lymphoblastic leukemia. Blood 1999; 93:2817–2823.
38. Relling MV, Hancock ML, Rivera GK, Sandlund JT, Ribeiro RC, Krynetski EY, Pui CH, Evans WE. Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. J Natl Cancer Inst 1999; 91:2001–2008.
39. Relling MV, Rubnitz JE, Rivera GK, Boyett JM, Hancock ML, Felix CA, Kun LE, Walter AW, Evans WE, Pui CH. High incidence of secondary brain tumours after radiotherapy and antimetabolites. Lancet 1999; 354:34–39.
40. Phillips KA, Veenstra DL, Ramsey SD, Van Bebber SL, Sakowski J. Genetic testing and pharmacogenomics: issues for determining the impact to healthcare delivery and costs. Am J Manag Care 2004; 10:425–432.
41. Iyer L, Das S, Janisch L, Wen M, Ramirez J, Karrison T, Fleming GF, Vokes EE, Schilsky RL, Ratain MJ. UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. Pharmacogenomics J 2002; 2:43–47.
42. Tukey RH, Strassburg CP, Mackenzie PI. Pharmacogenomics of human UDP-glucuronosyltransferases and irinotecan toxicity. Mol Pharmacol 2002; 62: 446–450.
43. Meyer UA. Pharmacogenetics–five decades of therapeutic lessons from genetic diversity. Nat Rev Genet 2004; 5:669–676.
44. Zanger UM, Raimundo S, Eichelbaum M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. Naunyn Schmiedebergs Arch Pharmacol 2004; 369:23–37.

45. Kirchheiner J, Nickchen K, Bauer M, Wong ML, Licinio J, Roots I, Brockmoller J. Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. Mol Psychiatry 2004; 9:442–473.
46. Sindrup SH, Brosen K. The pharmacogenetics of codeine hypoalgesia. Pharmacogenetics 1995; 5:335–346.
47. Yue QY, Hasselstrom J, Svensson JO, Sawe J. Pharmacokinetics of codeine and its metabolites in Caucasian healthy volunteers: comparisons between extensive and poor hydroxylators of debrisoquine. Br J Clin Pharmacol 1991; 31:635–642.
48. Caraco Y, Sheller J,Wood AJ. Impact of ethnic origin and quinidine coadministration on codeine's disposition and pharmacodynamic effects. J Pharmacol Exp Ther 1999; 290:413–422.
49. Desmeules J, Gascon MP, Dayer P, Magistris M. Impact of environmental and genetic factors on codeine analgesia. Eur J Clin Pharmacol 1991; 41:23–26.
50. Xu CF, Lewis KF, Yeo AJ, McCarthy LC, Maguire MF, Anwar Z, Danoff TM, Roses AD, Purvis IJ. Identification of a pharmacogenetic effect by linkage disequilibrium mapping. Pharmacogenomics J 2004; 4:374–378.
51. Handschin C, Meyer UA. Induction of Drug Metabolism: the role of nuclear receptors. Pharmacol Rev 2003; 55:649–673.
52. Holtzmann NA. Clinical utility of pharmacogenetics and pharmacogenomics. In: Rothstein MA, ed. Pharmacogenomics. Social, ethical, and clinical dimensions. HobokenNew Jersey: John Wiley & Sons, Hoboken, New Jersey.
53. Tucker G. Pharmacogenetics—expectations and reality. BMJ 2004; 329:4–6.
54. Zineh I, Gerhard T, Aquilante CL, Beitelshees AL, Beasley BN, Hartzema AG. Availability of pharmacogenomics-based prescribing information in drug package inserts for currently approved drugs. Pharmacogenomics J 2004; 4:354–358.

10

Regulatory Perspectives on Pharmacogenomics

Lawrence J. Lesko and Janet Woodcock
Food and Drug Administration, Rockville, Maryland, U.S.A.

1. INTRODUCTION

Regulatory perspectives on pharmacogenomics (PGx) have been presented in the literature over the past several years to focus attention on an emerging technology that has the potential to make targeted treatments more widely available to physicians and patients (1,2). In this chapter, PGx will be used broadly to collectively describe all of the interindividual variations in the whole genome or candidate gene single-nucleotide polymorphism (SNP) maps, haplotype markers and alterations in gene expression or inactivation that may be correlated with clinical response. We intend that PGx encompass the more narrow science of pharmacogenetics (PGt), which refers to the study of interindividual variations in DNA sequence related to drug absorption and disposition (pharmacokinetics) or drug action (pharmacodynamics) including polymorphic variation in genes that encode the functions of transporters, metabolizing enzymes, receptors, and other proteins. We will use the term PGx tests to refer to an assay to study these interindividual variations in conjunction with drug development and therapeutics.

It is well known that there is a major problem in the health care system in the United States related to providing new drugs that are effective and relatively safe in a wide diversity of patients with undifferentiated diseases whose individual dose–response relationship varies based on genetic, disease, environmental, and life-style factors. While many may argue over

the magnitude of the problem or its root causes, few can deny that a problem exists. There is increasing evidence for many drugs that genetic factors play a proportionally greater role in causing interindividual variability than other non-genomic factors. Advances in the technology to identify and measure genomic biomarkers related to the pathophysiology of diseases have led many to envision many more targeted drug/test combinations (e.g., HER 2neu coupled with trastuzumab) that are directed to specific patient subsets in the population as defined by their gene variants. From the perspective of the U.S. Food and Drug Administration (FDA), the pharmaceutical industry is facing a "pipeline problem." This perception is based on the evidence that in 2002, the FDA received only 21 submissions for new molecular entities (NMEs) intended for marketing approval (Fig. 1). This was more than a 50% decline from the NME submissions received in 1996 and the lowest number submitted over a 10-year period. Of serious concern, is the fact that this decline in NME productivity occurred despite a 2.5-fold increase in the U.S. pharmaceutical R&D spending over the same 10-year period (Fig. 2). The same observations have been made regarding the productivity of the pharmaceutical industry in Europe and in other developed countries. Furthermore, it has been estimated that to develop a single, successful new chemical entity now costs in excess of 800 million dollars including opportunity costs: the average time taken to do so is 8 to 10 years (3). The clinical component of the overall cost of new drug development is approximately

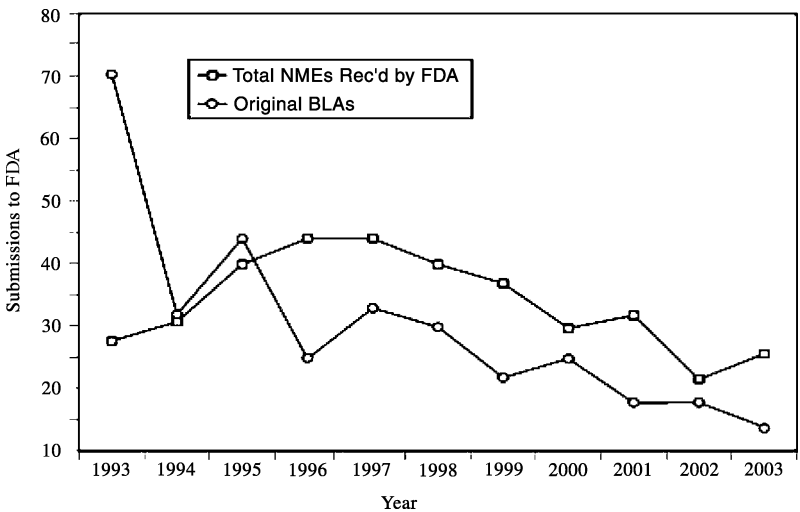


Figure 1 The number of submissions of new drug applications (NDAs) for new molecular entities (NMEs) and the number of submissions of biologics license applications (BLAs) to FDA over a 10-year period.

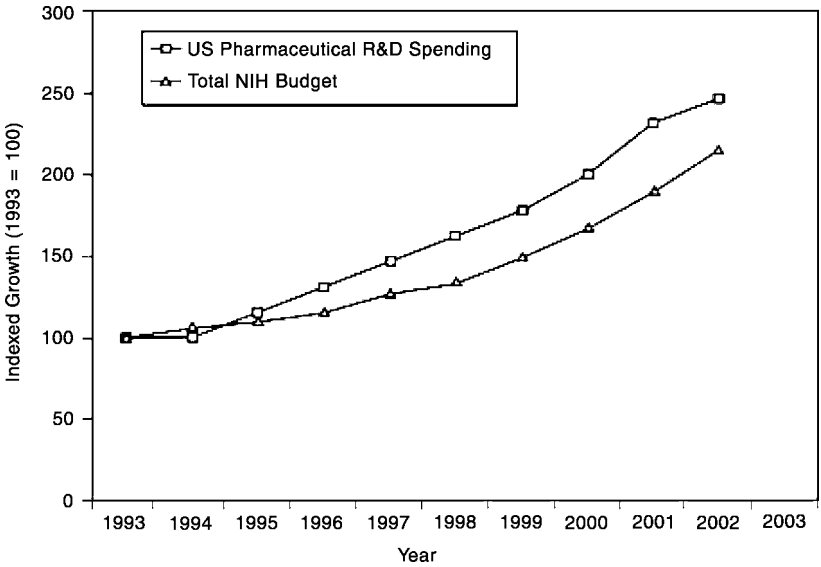


Figure 2 The 10-year trend in biomedical research spending as reflected by the National Institutes of Health (NIH) budget (FY 1993-2003) and by the pharmaceutical companies research and development investment. *Source*: Parexel's Pharmaceutical R&D Statistical Sourcebook 2002/2003.

58% or 400 million dollars. A significant amount of these dollars go towards supporting the phase-3 randomized controlled trials (RCT) that provide the most convincing evidence of a drug product's safety and efficacy. However, from a recent report, one can estimate the failure rate in phase-3 trials to be approximately 50%, although the failure rate is dependent on the therapeutic area, being higher, for example, in oncology and lower in infectious diseases (4). The system for developing NMEs is not broke but it is in need of significant improvement. Until the drug development and regulatory process itself, and the translation of new discoveries into effective medicines, catches up with the rapid pace of technological advances in science, it is doubtful that significant improvements can be made.

2. VARIABILITY IN DRUG RESPONSE

Variability in drug response is a major barrier to successful drug development. As Sir William Osler said in 1892 about the practice of medicine, "if it were not for the great variability among individuals, medicine might as well be a science and not an art." PGx can provide the scientific tools that enable us to explore the pathophysiological mechanisms for these differences in drug response at

the molecular level. We expect that there will be an increase in the public demand for more science and less art in the search for better and more effective therapies to reduce the morbidity and mortality of chronic diseases such as hypertension and cancer. In order to improve the "art" and the productivity of the drug development process, PGx can improve the predictability of preclinical safety studies and clinical safety and efficacy trials.

A key to increasing R&D success is identifying failures early in the drug development process and reducing attrition in phase-3, before the high costs of late-phase clinical trials are incurred by a sponsor. This is an important achievement because the average size of a phase-3 clinical trial has nearly tripled in the last 20 years. It simply does not make much sense to wait until a phase-3 trial fails to try and figure out why and how to design the next trial; that is expensive and time-consuming. It's typical that each phase-3 trial is preceded by a much longer preclinical and early clinical work-up of the drug, so what is needed is an increased ability to predict phase-3 success or failure, aimed at the preclinical and early clinical time period. For example, in terms of cost, a 10% improvement in predicting failure before large-scale phase-3 clinical trials begin could save approximately 100 million dollars in development costs. Other opportunities for saving 12 to 21 million dollars in direct development costs can be attained by shifting only 5% of clinical failures from phase-3 to phase-1, or by shifting 1/4 of failures from phase-2 to phase-1 (5). The major causes of attrition of drugs in late-phase clinical trials are either lack of efficacy or concerns about safety although drug development is also terminated for commercial reasons. In order to achieve increases in productivity and success, effective scientific development tools, such as PGx tests, are needed to predict product performance, whether it be success or failure, with a high degree of certainty, both early and reliably in the development process. For example, pharmacogenomic biomarkers can be used to identify potential responders. By stratifying patients by biomarker status in phase-2 clinical trials, populations with a high probability of response can be identified, thereby simplifying phase-3 trials and increasing their probability of success.

Clearly, modern innovative tools are needed to predict the performance and manufacturing quality of 21st century products. While it seems that everyone agrees with this premise, the problem is that the drug development process is no longer able to keep pace with the rate and scope of innovative basic science discoveries. For example, while imaging-based biomarkers are presently being used to develop drugs for Alzheimer's disease, there has not been a successful strategy to correlate imaging with primary clinical endpoints such of cognition and function in order to increase the success rate of new molecules intended to modify disease progression. The tools currently used in drug discovery and development—the so-called "critical path" tools—have not incorporated and linked the latest advances in biomarker technology, the basic and information sciences, such as the

new knowledge and technologies provided by the rapid development of genomic research, and the clinical sciences to impact the success of drug development and improve the quality of public health substantially. While the reasons underlying failures of drugs in development (especially those failing in late-phase clinical trials), and inefficiencies in the development process in general, are not well understood, many suspect that a lack of understanding of variability in drug response between patients is a key part of the problem. Recent, and rapidly evolving, evidence is beginning to point toward genetic and genomic factors, alone and taken together with environmental factors, as playing a major role in interindividual variability.

The principle of genomic tests is that genetic variation among individuals causes, or correlates, with responses to drugs. Such variations occur frequent enough and that the differences in dose–response are large enough to justify testing before treatment, shortly after beginning treatment, or any time during treatment. It is important to point out that genomic tests are considered as adjunct tools to complement other forms of monitoring in order to better understand the causes of beneficial and adverse responses. Frequently genotyping can be used to more precisely define dosing for an individual patient.

From a clinical perspective, being able to predict failures in therapy or severe adverse events using PGx biomarkers and validated diagnostic tests for important polymorphisms in patient's receptors and/or drug metabolism genes, or in tumor and/or organism genes has the potential to optimize drug selection and/or drug dose *a priori* or much earlier in therapy to increases the probability of more effective therapy and decrease the probability of having a serious adverse event.

3. THE FDA's APPROACH TO THE PIPELINE PROBLEM

In keeping with its mission to advance the public health by helping to speed innovations that make medicines more effective, safer, and more affordable, the FDA issued a white paper on March 16, 2004 entitled "Innovation or Stagnation? Challenge and Opportunity on the Critical Path to New Medical Products" (http://www.fda.gov/cdrh/ocd/criticalpath.html). The white paper is a serious attempt by FDA to bring attention and focus to the need for targeted scientific efforts to modernize the tools, techniques, and methods used to evaluate the safety, efficacy, and quality of drug products. The Critical Path document highlighted examples of the FDA's efforts to improve the critical path, and it discusses future opportunities as well. The public response to the critical path document was extensive and extremely positive, and many individual companies, as well as the trade associations representing the industry, provided many suggestions to FDA on ways to improve the efficiency and success of drug development. The critical path report also describes the urgent need to build coalitions

between the constituencies, such as the FDA,NIH, the private sector, and the nation's universities, to modernize the medical product development process—the critical path—to make product development more predictable and less costly, and drug products more effective with greater safety by identifying key problems for targeted solutions.

The critical path is defined in the document as the path from candidate selection to product launch and it defines the potential bottlenecks in the process of bringing a product to market. The focus of critical path is to update the product development infrastructure for drugs, biologics, and devices and the evaluative tools currently used to assess the safety and efficacy of new medical products. Examples of evaluative tools include better pathophysiological cell and/or animal disease state models for preclinical screening of new molecules, and innovative scientific approaches such as the use of Bayesian statistics and adaptive trial design, validation of pathophysiological and/or descriptive biomarkers for clinical trial patient selection (enrichment) and/or as surrogate endpoints of drug effects, and the use of modeling and computer simulation to predict drug and device failures, and improve post-market reporting of implanted device adverse events, that could assist in developing more focused premarket trials. In addition, an important example of a critical path scientific opportunity is *pharmacogenomics* and *pharmacogenetics*, or more specifically, the identification of DNA-based biomarkers or RNA expression profiles that can provide insights into the stage of a disease, disease progression, drug response, and drug dosing requirements.

4. ADVANCEMENT OF PHARMACOGENOMICS: PART OF THE FDA's MISSION

The FDA's twofold mission includes protecting and advancing public health, and speeding innovations that make medicines and foods more effective, safer, and more affordable. With regard to PGx, the FDA has attempted to assure that the regulatory pathway for moving genomic biomarkers from discovery to clinical practice is articulated clearly. The Agency intends to promote high quality in PGx studies with regard to processes to demonstrate analytical validation, clinical validation, and clinical usefulness, respectively. Beginning in earnest in June 2001, FDA took the lead with several key initiatives in PGx intending to stimulate the exploratory use of PGx technologies in drug development and foster improvements in drug product safety and efficacy. After publication of a forward-looking paper providing a regulatory perspective on the opportunities and challenges of integrating pharmacogenomic in drug development and regulatory decision-making (1), the FDA has coordinated its efforts both between its own Centers, and with the pharmaceutical and biotechnology industries to convene a series of public PGx workshops. These workshops are a

structured, modular approach to bring together stakeholders from industry and academia, and FDA scientists to openly discuss the status of PGx technology, the use of PGx in drug development and therapeutics, and what specific strategies are needed most for using PGx as a tool to facilitate more efficient and effective research along the critical path of drug development. Publications of the proceedings of these workshops are valuable references that lay out where PGx in drug development is currently and what is needed to continue to advance this critical path tool (6–8).

4.1. Guidances Related to New Drug Development

The culmination of many individual efforts within FDA and the public input derived from the synergistic FDA–Industry co-sponsored workshops led to a significant milestone in the advancement of PGx: the November 2003 publication of the Draft Guidance for Industry: Pharmacogenomic Data Submissions (http://www.fda.gov/cder/guidances/5900dft.pdf). This guidance was timely, in that there was considerable uncertainty and fear about what FDA would do with exploratory genomic data obtained during the new drug development process, and this was a stumbling block for many pharmaceutical companies. The major concern was that FDA would overreact to non-validated, exploratory genomic biomarkers, take them out of context, misinterpret them, cause delays in drug development, request additional clinical trials, and/or put clinical trials on hold. This concern led to a reluctance of the industry to introduce genomic studies into their drug development plans. The FDA wanted to breakdown these real or perceived barriers and to motivate drug developers to consider PGx and PGt strategies seriously in their drug development portfolios. The PGx data guidance proposed a new pathway for industry and others to submit non-clinical and clinical exploratory genomic data during the IND period without it undergoing formal regulatory review, and describes the submission format and regulatory review of such data by the Interdisciplinary PGx Review Group (IPRG). It introduced some new concepts related to genomic biomarkers and defined categories of biomarkers, i.e., exploratory, valid, probable valid, and known valid biomarkers. By design, the guidance shied away from presenting very specific recommendations for biomarker validation and formats for submitting genomic data in order not to encumber progress in the field. The FDA recognized that the science is still evolving.

Important components of the guidance are three decision algorithms or decision trees based on the categories of biomarkers and the stage of drug development. Generally, most genomic data submitted to FDA to date has been exploratory and not suitable for regulatory decision making. Such data, e.g., derived from gene expression microarrays, have either no clear pathophysiological correlates, and/or are not critical to entering patients into clinical trials or supporting claims about safety, efficacy, and/or dosing.

Valid biomarkers are defined as those biomarkers measured in an analytical test system with well-established performance characteristics and with an established scientific framework or body of evidence that elucidates the physiologic, pharmacologic, toxicologic, or clinical significance of the test results. Known valid biomarkers are those accepted in the broad scientific community whereas probable valid biomarkers are those that appear to have predictive value for clinical outcomes, but have not yet been widely accepted or independently replicated. The decision trees elucidate when genomic data can be submitted voluntarily, and when submissions of the data are required by FDA regulations. In addition, the guidance describes the format (full report, abbreviated report, synopsis, or voluntary submission) for submitting such data. It should be noted that the process for industry to submit VGDS to FDA was set up to be through the existing path for IND (or as a pre-IND in some cases) or NDA submissions, which assures the confidentiality of the data. The Genomic Data Submission guidance has been revised as of February 2005 and is awaiting final clearance through the Agency. Along with the finalization of the guidance, FDA has written SOPs to describe the operations of the Interdisciplinary Pharmacogenomic Review Group and the process for submitting VGDS documents. A web site http://www.fda.gov/cder/genomics/default.htm will be available in spring of 2005 to which sponsors and the general public can refer for instructions and information related to VGDS. Since the draft guidance was published, there have been at least a dozen meetings between FDA and industry regarding VGDS. The agenda for these meetings have focused discussion on questions related to the validation of preclinical toxico genomics biomarkers, retrospective analysis of predictive biomarkers of drug safety and efficacy from phase-2 and phase-3 clinical trials, and the prospective designs of clinical PGx studies including study design and data analysis issues.

The FDA hopes that voluntary submissions will benefit both the industry and the Agency and will provide a rational scientific basis for future data standards and genomic policies. Information and knowledge gained from voluntary submissions will be shared publicly across submissions in a way that protects the proprietary interests of companies. The FDA is currently in the process of finalizing the Draft Guidance on Pharmacogenomic Data Submissions and writing two other internal documents that will describe the process for sponsors submitting voluntary genomic data submissions and the roles and responsibilities of the IPRG.

The most recent guidance initiative of the FDA is to articulate the principles and processes associated with drug or biologic and genomic test co-development. Several workshops (July 2004 and April 2005) will have been held to get public input into the key components of such guidance. The key elements of this guidance will focus on analytical performance standards for a genomic test, clinical performance attributes, and approaches to documenting clinical utility. The draft guidance, when released in 2005, will

also provide a description of the regulatory pathway for submitting drug/test co-developed products and the framework for inter-Center cooperation to expedite review of such submissions.

5. PHARMACOGENOMICS IN THE PROCESS OF NEW DRUG DEVELOPMENT

Over the past 5 years, PGx studies have become an increasingly greater part of drug development although the extent of inclusion of PGx data in NME submissions has not kept pace with the tremendous and rapid advances in genomic technology and the ability to generate extensive genomic and proteomic data profiles. Drug companies reportedly collect DNA samples from subjects in approximately 80% of clinical studies so that they can have the chance to identify genomic biomarkers of drug safety, efficacy, and dosing. The promise of PGx lies in its potential to identify sources of interindividual variability in drug response (both efficacy and safety) that arise from genomic differences in disease pathophysiology and/or genomic differences in drug pharmacology. The PGx biomarkers can serve many different purposes depending on the context and questions at hand, such as: (a) entry criteria for a clinical trial, (b) indicator of disease status, (c) patient stratification, (d) drug efficacy predictor test, (e) predictor test for adverse events, (f) test for monitoring drug response, and/or (g) guide to dose selection. PGx can also facilitate product differentiation in the marketplace by improving the benefit/risk profile compared with other similar products. Of course, whether or not to recommend mandatory or optional genotyping, which would be clinically useful, would need to be determined for each drug.

The current ''gold standard'' for evidence of drug efficacy is the randomized, adequate and well-controlled clinical trial. As mentioned previously, the number of phase-3 trials that fail to meet their prespecified acceptance criteria is around 50% as reported by PhRMA. So clearly, the current practice of controlled clinical trials is inadequate. Even when a study demonstrates an overall significant ($p < 0.05$) clinical effect, the efficacy signal often arises from a subset of patients, with other patients, in effect, being non-responders.

Two recent examples of this phenomenon from the field of oncology are provided below.

5.1. Gefitinib (Iressa)

Gefitinib is a tyrosine kinase inhibitor that targets the tumor protein, epidermal growth factor receptor (EGFR), and was approved by FDA for advanced non-small cell lung cancer in May 2003. The overall response rate in the United States was about 10% ($n = 216$). A subset analysis revealed

apparent sources of variability in clinical responsiveness with greater responses observed in women and those with adenocarcinoma (17%) as compared to responses observed in men and smokers (5%). Furthermore, the response rate in Japanese subjects with NSCLC in Japan was approximately 25–30%. Impressively, individual patients of both genders demonstrated dramatic responses. A molecular mechanism underlying gefitinib efficacy was described as activating mutations in the EGFR in which specific deletions or amino acid substitutions around the ATP binding site of gefitinib increased EGFR signaling and susceptibility to inhibition by gefitinib (9,10). In a small trial, specific mutations were identified in eight of nine gefitinib responders and were not identified in any of the seven non-responders. Related to the variability in gefitinib, lung cancer cells with mutations were 10 times more responsive than normal cells to the medicine, and mutations were more common in tumor cells from Japanese patients. This may explain the higher efficacy response rate in Japan. It is not hard to imagine how a diagnostic test for EGFR positivity can be integrated into treatment decisions and as a guide as to when and in whom to use the drug.

5.2. Erlotinib (Tarceva)

Erlotinib is an EGFR tyrosine kinase inhibitor approved by FDA in December 2004 for advanced or metastatic non-small cell lung cancer as second line monotherapy. The approved dose was 150 mg/day although doses of 25 and 100 mg were also studied. From the clinical trial data, the median survival for erlotinib treatment ($n = 488$) was 6.7 months as compared to a placebo ($n = 243$) median survival of 4.7 months. The percentage of patients alive at 12 months was either 31.2% (erlotinib) or 21.5% (placebo). The median progression free symptom time was 9.9 weeks for erlotinib and 7.9 weeks for placebo. The medicine was clearly effective in the clinical trial general population compared to placebo. A subset analysis based on EGFR protein expression status found that the relative tumor response rate was 11.6% (EGFR+ patients) vs. 3.2% (EGFR− patients), while the tumor response rate in placebo treated patients was similar to the response in the EGFR− patients. The survival hazard ratio for EGFR + patients was 0.65 demonstrating a survival benefit vs. placebo, while the survival hazard ratio for EGFR − patients was 1.01, indicating no survival benefit. However, the prevalence of EGFR positivity is thought to be approximately 50–55% so that it is likely that the positive response in the general population may have been due entirely to those patients who were EGFR positive. Had the prevalence of EGFR positive status been, for example, 10–20%, the trial in the general population may have failed. Erlotinib was approved for all patients, not just EGFR+ patients, for several reasons: (a) the EFGR status was unknown in a large percentage of patients, (b) the EGFR assay was not validated appropriately, (c) the relatively small numbers of patients in the EGFR− subset, and (d) the confidence intervals

for tumor response and survival overlapped. This example demonstrates how a differential disease diagnosis, based on EGFR protein expression patterns, can be used as a basis for future enrichment trials or as a diagnostic to identify those patients most likely to respond or not.

There are several other additional cases in oncology in which the efficacy response to medicines has been associated with genomic biomarkers, e.g., the well-known selective response to trastuzumab (Herceptin) in patients whose tumor overexpresses the HER 2 protein, the response in patients with chronic myelogenous leukemia to imatinib depends on the BRA/ABL translocation, and the response to cetuximab in patients with colon cancer depends on the EGFR protein expression. Each of these examples demonstrates how a mechanistic knowledge of PGx–clinical phenotype relationships gleaned from earlier drug development projects can be used to improve the efficiency of development of second generation drug molecules, both with and without a diagnostic test.

There are also examples of the important role that genomic biomarkers can play in drug development outside of oncology such as the demonstration that the response to PEG-interferon in patients with hepatitis C depends on the virus genotype and the clinical response to certain CNS drugs and drug classes (selective serotonin reuptake inhibitors) depends on the receptor genotype.

5.3. Atomoxetine (Strattera)

A recent example illustrates the role that PGt played in the labeling of a new drug is the case of atomoxetine (Strattera®). This drug was approved by FDA in January 2003 for attention deficit/hyperactive disorder with a fixed dose of 0.5 mg/kg to be titrated up to 1.2 mg/kg. The drug is metabolized by cytochrome P450 2D6 (CYP2D6) with a clearance of 0.35 L/hr/kg in extensive metabolizers (EMs) and 0.03 L/hr/kg in poor metabolizers (PMs). The ratio (PM/EM) of the area-under-curve (AUC) for plasma atomoxetine was approximately 10. The sponsor did a sensible analysis of adverse events in clinical trials by looking at a post-facto stratification of patient subsets defined by genotype. The frequency of adverse drug reactions (ADR), primarily insomnia and irritability, was 9% in PMs and 6% in EMs. There were no major differences in serious ADRs between PMs and EMs. The label of atomoxetine mentions CYP2D6 in seven different sections including those describing pharmacokinetics, drug–drug interactions, adverse events and laboratory tests. However, the evidence did not warrant recommending that a pharmacogenetic test for CYP2D6 status be done before prescribing the drug, but it did provide descriptive information that could be used along with other observations (e.g., an adverse event) to guide clinician decisions about an individual's need for dosing adjustment. This example demonstrates the value that pharmacogenetic information in a package

insert can bring to the use of the drug, including knowledge related to genotype (e.g., CYP 2D6 alleles), phenotype (e.g., poor metabolizers), and clinical outcomes (e.g., adverse events) that can increase the quality of clinician's decisions about individualizing drug treatment.

5.4. Selective Serotonin Reuptake Inhibitors (SSRIs, Many)

An emerging example of PGx being used to differentiate a chronic disease is in the field of depression with the development of new SSRIs. These drugs, along with supportive care, represent the standard of care of major depressive disorders (MDD). MDD are CNS diseases caused by underling complex polygenic traits. Genetic heterogeneity in the disease, and polymorphism in the 5-HTT receptor, which is the selective site of action of SSRIs, leads to high degree of intersubject variability in the clinical phenotype. Many clinical trials fail to either demonstrate a dose–response relationship and/or beat placebo partly because of the subjective and variable nature of the empirical clinical endpoints, the relatively high placebo response rate (~50%) and the variable duration of treatment needed to assess clinical efficacy response (~6 weeks). Clearly non-responders to SSRIs (~30–50% of entered patents) reduce the effective study power. There are no valid or probable valid biomarkers to predict who will and who will not respond more favorably to SSRIs vs. placebo. Since the adverse event rate is relatively high (3.5%), patients who are non-responders have a different benefit/risk ratio than responders and it would be beneficial to differentiate between the responder and non-responder phenotypes. Recently genotypes were identified as having two forms of the 5-HTT gene, which regulates brain serotonin levels: either a short form (S allele) or a long form (L allele). In case control studies, clinical response scores (HAM-D and CGI-I) defined sequentially over time after dosing showed a significant difference in response between S/S and L/L groups. The L/L genotype responded more rapidly to model SSRIs such as sertraline, and a higher percentage of the L/L genotypes were classified as responders. The important implications of these findings are that clinical trials for new SSRIs may stratify patient populations on the basis of carrying either a S- or L-allele. Further, enriching trials with the L/L genotype will provide the maximum probability of a successful proof-of-concept study and greater success in phase-3 trials.

6. REVISING PHYSICIAN LABELING OF PREVIOUSLY APPROVED DRUG PRODUCTS

Translating PGx from bench to bedside (or discovery to marketability) is a multidisciplinary problem, involving private and public sector philosophical, societal, cultural, behavioral, educational, drug development, scientific

expertise, communication, and clinical practice issues. For drugs whose clinical activity is highly influenced by genetic factors, but were developed before the rapid growth of PGx, the Agency has decided it is worthwhile to question, if not change, the wording in drug labels. It recognizes that integrating PGx into therapeutics for previously approved drugs where there is an established history and practice of use represents a major challenge for regulatory agencies and other stakeholders in PGx. In addition, until recently, there were no FDA-approved tests for discerning the genotypes of CYP450 enzyme function.

Labeling regulations (21 CFR 201-57) related to tests to guide therapy are clear. If evidence is available to support the safety and effectiveness of a drug only in selected subgroups of the larger population with a disease, the labeling shall describe the evidence and identify specific tests needed for selection and monitoring of patients who need the drug.

The FDA has a long-standing interest in "individualization factors" such as those defined by intrinsic (e.g., age, gender, race, and renal dysfunction) and extrinsic factors (e.g., food, co-administered drugs, smoking, and alcohol). At best, these factors are crude differentiators of patient subsets at greater of lesser risk of toxicity or loss of efficacy. The Agency believes that an appreciation of controllable sources of variability in drug action and potential injury to patients should be achieved prior to the marketing of new pharmaceutical products. Information on these important co-variates influencing drug safety and efficacy are generally reported in various section of the product package insert and are in need of constant updating as post-marketing experience with the drug increases over time. The value of PGx in improving individualization of therapy depends on the ability to demonstrate a correlation between important clinical outcomes and individual genotype as expressed by validated biomarkers. It is unlikely that drug manufacturers of established drugs will be willing to subsidize prospective clinical trials to provide evidence of the clinical validity and utility of CYP 450 drug response predictor tests. Thus, the Agency has instead relied upon a systematic analysis of extensive literature data from case-control studies or prospective exploratory trials, and its own Med Watch reports, as well as prospective, randomized controlled trials wherever available. Understanding the strength, quality, and characteristics of the genotype–phenotype relationship is the first step towards developing a predictive test to optimize therapeutics on an individual basis.

In December 2004 a milestone in the evolution of clinical applications of PGx was reached. The FDA approved the first in vitro diagnostic test for CYP450 enzymes (Amplichip, Roche Molecular Systems, Inc., Almeda, California, USA) based on microarray technology (Affymetrix). The chip-based test provides a way for the physician to measure individual variations in drug metabolism genes for CYP 2D6 and 2C19. These CYPs metabolize approximately 25–30% of all clinically approved drugs including tricyclic

antidepressants, CNS-active drugs, beta-blockers, and proton-pump inhibitors. The significance of this test is that approximately 25–30% of all approved prescription drugs are substrates for these enzymes. Several drugs metabolized by these enzymes are the frequent cause of adverse events thought to be related to excess systemic drug exposure following the administration of the usual doses. The chip can be used to analyze 29 polymorphisms for CYP 2D6 and two polymorphisms for CYP 2C19 and thus can be used to genotype nearly 100% of the global patient population. Physicians can get the genotype results within 1 day and can use the results of this diagnostic to help individualize treatment doses for individual patients receiving drugs metabolized predominantly through these enzymatic pathways. Genotyping for gene variants of 2D6 and 2C19 usually only has to be done once per patient and the results could be used to reduce doses of many drugs metabolized by the same enzymes.

An unknown issue is what will be the pace of physician and patient acceptance. It is well known that there is a gap between PGx research and physician prescribing decisions, and the best way to achieve knowledge transfer and translation is not entirely clear. Further, the additional information that PGx provides makes writing a prescription more complex. Models for translating PGx research results into applications in patient care are sorely needed. Uptake of new tests, genomic or not, typically must overcome significant medical, social, and cultural barriers to become established as a standard of care and in addition, reimbursement issues come into play. With the CYP 2D6 and 2C19 testing, the important considerations as to the decision to use a PGx test or not include the following:

1. What extent of intersubject variability dose–PK relationship, what percent of drug metabolism via the polymorphic pathway, and what shape of the dose–response curves for efficacy and safety should justify genotyping? Even with good genotype–phenotype associations, these basic questions will be made more complex by considering that the dose–PK relationship will be influenced by many demographic, environmental, and drug interaction factors. Dose–response considerations will also be affected by intended use of the drug (many drugs have multiple uses), the patient's concomitant drug therapy and disease subset (many disease have a range of clinical phenotypes), all of which have the potential change the dose–response relationship and influence the impact the genotyping will have on benefit/risk.
2. Why would genotyping tests be ordered and when? The intended use of the test is usually to reduce adverse events by more precise dosing in at risk genotypes defined by the tests. Possible scenarios include a priori testing followed by tailored initial dosing in patients at-risk, e.g., patients with heart disease who need a

beta-blocker, or patients with atrial arrhythmias who need warfarin, testing only if unexpected toxicity occurs at a relatively low dose or if there is an overt lack of efficacy at maximum approved doses. Genotyping will provide complementary information to make better therapeutic decisions in these situations.

3. What dose is appropriate as defined by genotype? It is difficult to answer this question quantitatively because there are very little prospective clinical trial data on genotype-based dose–response relationships or prospective randomized, adequate and well-controlled clinical trials that have been designed to answer the question about optimal doses for different genotypes. In lieu of controlled trials, dosing adjustment similar to other patient subsets, i.e., based dosing adjustment on the relative exposure in various genotypes using area-under-the-plasma drug level curve (AUC). However, it must also be recognized that often the dose suggested by increases in AUC is greater than the doses actually needed in patients titrated to desired dose based on clinical endpoints, an observation that points to the fact that many non-genetic factors contribute to variability in dose–response for most drugs.

The atomoxetine example presented earlier and the entry of the Amplichip into clinical practice settings also brings to mind several other challenges that face sponsors, regulatory agencies, and clinicians in translating genotype information from research to the clinic. Accurate tests for CYP enzyme alleles are effectively diagnostics of hepatic enzyme function and not clinical response. Medicines are used to treat symptoms, not enzyme function. Therefore, interpretation of the test results is critical to translate this PGx knowledge into wisdom as it applies to drug dosing requirements for individual patients. Some type of computer-assisted analysis to test results (i.e., a bioinformatics solutions) is needed to provide the physician some guidance on how to use the data similar to what has been used commercially with AIDS resistance diagnostic tests.

Specifically, with regard to physician labeling, there are some other complex decisions to be made. Some examples include:

1. What is the best way to define poor metabolizers in a research setting? The PM is a phenotype that can be also be determined in non-genomics ways by the urinary metabolic ratio, the observed AUC or plasma clearance of the drug in different genotype subsets. Frequently there is overlap between the phenotypes used to define poor (PM), intermediate (IM), and extensive (EM) metabolizers. For example, there are more than 40 alleles of CYP2D6 with about 25% of them having greatly decreased or null activity. There is also significant variability in the frequency of null alleles of CYP2D6, and other polymorphic enzymes such as CYP2C19 in

different racial or ethnic groups. So, the open question is what alleles should be studied in drug development, and how should this information be translated into a product's package insert for use by physicians? Ideally, all common alleles should be studied in a way that includes patients of ancestral heritage in the general population that is expected to be prescribed the drug.

2. How should PGx information be reported in the label? The consideration are (a) whether or not to report only phenotype data (e.g., PMs and EMs), or specific alleles of CYP2D6 (e.g., *2, *3, and *10), and (b) how much information should be reported in light of who will interpret the significance of these data with respect to dosing, safety, and efficacy? Perhaps as commercial tests approved by FDA increase, there will also be an increase in clinical consultants (e.g., laboratory clinicians, clinical pharmacologists, and/or clinical pharmacists) who will interpret genotype information in the context of the total patient care. Perhaps point-of-care proprietary software will interpret the genotype data and classify the results as those coming from a PM, IM, or EM in the physician's office.
3. If PGx information is included in the label of a drug product in a way that gives physicians and patients an option to have a genomic test done as part of therapy, this raises other translational issues that include knowledge of the test availability, the quality of the test results, its turnaround time and costs, including reimbursement options, and as mentioned the proper interpretation of test results.

Despite a widespread recognition of the challenges above, and with all the reports of important gene effects on differential diagnosis and/or dose–response, there have been relatively little attempts to create models to translate of this information into drug development and even less into clinical practice. The FDA believes that there is value in applying long-established drug response predictive genetic biomarkers to older, marketed drugs in the post-marketing period in order to improve their risk/benefit ratio by optimizing or individualizing dosing. Examples of older drugs that could benefit from PGx are 6-mercaptopurine (6MP), azathioprine, and 6-thioguanine, or 6TG, (each substrates for thiopurine methyltransferase or TPMT), irinotecan (a substrate for uridine diphosphate glucoronosyltransferase or UGT), and warfarin (a substrate for CYP2C9). Each of these drugs has a narrow therapeutic range, wide interindividual variability in pharmacokinetics, wide range of dosing requirements and frequent and serious safety problems. The genes of each the enzymes mentioned above can exist in one of several isoforms (e.g., TPMT*2, UGT1A1*28 and CYP2C9*3) and these enzymes are mostly found in either red blood cells (in the case of

TPMT) or the liver (for UGT and CYP2C9). Certain mutations in these isoforms, or gene variants, produce different phenotypes but most important to drug dosing is the poor metabolizer phenotype that results in heightened exposure to either the parent drug or a major metabolite, or reduced exposure to an active metabolite (e.g., morphine from codeine administration).

In July 2003, the FDA Pediatric Subcommittee of the Oncology Drug Advisory Committee (ODAC) was presented a comprehensive amount of relevant research data from the scientific literature and from FDA files to consider and asked to discuss whether or not the package insert of 6MP should be updated to include information on TPMT genotypes. 6MP was approved decades ago for use in children with acute lymphoblastic leukemia (ALL) and, taken orally together with methotrexate and/or other chemotherapeutic agents, is the backbone of continuation therapy. Dose intensity of 6MP is a major determinant of both event free survival (efficacy) and neutropenia (safety). The clearance of 6MP, and hence exposure, is dependent on its conversion to 6-TG which, in turn, is metabolized via the TPMT pathway. More than 11% of the Caucasian population is heterozygous or homozygous carriers of TPMT alleles. There are three major genotypes, each with a range of TPMT activity (high, intermediate, and low), and each with a different relative risk of developing neutropenia when administered the standard dose of 6MP ($50\,mg/m^2$). The poor metabolizer genotype, with an incidence of 1:300, accumulates excess 6-TG that is nearly certain to lead to severe and potentially fatal bone-marrow toxicity. It has been recommended that the usual dose of 6MP be reduced by 80–90% for the PM genotype to reduce the risk of neutropenia. Based on the evidence presented in July 2003, the Subcommittee considered the consequences of a label revision thoroughly, and in the end recommended that the label of 6MP should be updated with current information on TPMT genotypes, but stopped short of recommending that testing for TPMT status be mandatory before prescribing the 6MP. The experts on the subcommittee considered many factors in making their recommendation. Some important considerations that were discussed include: (a) the scarcity of prospective clinical trials to support specific recommendations about dose reduction in patients who were either heterozygous or homozygous, (b) the wide interindividual variability in TPMT activity in patients who had one variant TPMT allele, and the subsequent risk of reducing effectiveness if doses are reduced unnecessarily, (c) the potential benefit and cost of TPMT genotyping as compared to phenotyping based on TPMT activity in red blood cells coupled with measurement of neutropenia (drop in neutrophils), and (d) the widespread availability and cost of TPMT testing. As a result of the committee recommendations, the FDA and sponsors have negotiated updated label language to include the evidence of the genotype–phenotype correlation and the significance to clinical outcome.

More recently (November 2004), the FDA's Clinical Pharmacology Subcommittee (CPSC) met to consider the evidence supporting the clinical significance of UGT1A1 polymorphism in irinotecan treatment of colon cancer. Irinotecan is converted to an active metabolite, SN38, which is then metabolized by UGT1A1. Approximately 10% of the general patient population are deficient in UGT activity (UGT1A1*28 genotype) and because of elevated SN38 exposure, are significantly at risk for developing neutropenia and/or diarrhea. SN38 exposure is the main driver for neutropenia although the evidence for its causing diarrhea is less clear. At the end of the discussion, the advisory committee voted to update the label of irinotecan with additional information on UGT polymorphism and the need to be cautious in dosing irinotecan. The FDA is currently negotiating with the sponsor on the final wording in the label.

These illustrative examples demonstrate that PGx can, in fact, make a contribution to drug safety by guiding towards appropriate dosing. However, translating PGx information from research to clinic for older drugs is in some ways more challenging than for newer drugs for the reasons cited above. The three categories of issues or questions raised as challenges following the atomoxetine example above for new drugs still apply to older or previously approved drugs. However, in addition, there are additional issues or questions.

1. What is the best model to educate clinicians about the advantages and limitations of adopting a PGx test for a drug that they have been using, albeit not optimally, for decades, especially, as in the case of 6MP, where neutropenia or another test (e.g., TPMT activity in red blood cells) has been used phenotypically as a rough guide to reduce the intensity of dosing, or in the case of warfarin where two gene variants (CYP 2C9*2 and CYP 2C9*3) and other demographic, disease, and drug–drug interaction factors, when taken together, account for only 35–40% of intersubject variability?
2. How should the dosing of drugs like 6MP, irinotecan, and warfarin be adjusted, based on genotype, when there is an absence of randomized, controlled clinical trials to demonstrate the efficacy of the reduced dose as some experts have recommended? This is a pertinent question in the case of 6MP where the success rate of event free survival in childhood ALL is nearly 80–85% and evidence supporting the reduction of dose in patients with intermediate TPMT activity is not substantial. Patients with high TPMT activity relative to a given dose may not receive the maximum benefit from the drug because of rapid clearance. The problem of an appropriate reduced dose for irinotecan is complex because the optimally effective dose for the drug in colon cancer

has not been well defined and many different dosing schedules are used in practice. On the other hand, prospective trials to determine effective reduced doses may take thousands of patients and many years, and would be costly. For previously approved drugs, especially those that are off patent, it is not clear who would pay for these studies.

3. When is the best time for genotyping patients administered 6MP for their TPMT activity status, irinotecan for their UGT activity, or warfarin for their CYP 2C9 status? Options include routinely genotyping prior to initiation of drug therapy, genotyping within the first week of receiving drug treatment or genotyping only in the case of overt toxicity, or lack of efficacy at the usual and/or maximum recommended doses.

However, as the Pediatric Subcommittee of ODAC pointed out in July 2003, and as the CPSC of the Advisory Committee for Pharmaceutical Sciences also recommended in November 2004, after careful consideration of the genotype–phenotype correlations for 6MP and irinotecan, genotyping is not a substitute for traditional monitoring of white blood cell counts in patients receiving 6MP or irinotecan, but as an adjunct to other clinical measures used to monitor for myelosuppression, or in the case of irinotecan, bilirubin measurements and genotyping, taken together are excellent predictions of a patient's risk for neutropenia. It is recognized that many other clinical factors (age, sex, indication, co-administered drugs, and body size) can influence maintenance doses of these drugs, but genotyping for TPMT or UGT1A1 testing, when combined with other tests and observations, can lead to higher quality decisions about drug selection and drug dosing that will further decrease the risk of severe and preventable bone marrow suppression, neutropenia, and possibly diarrhea (in the case of irinotecan). It is also known that genotyping alone cannot account for all of the observed toxicities of these drugs since there may be rare alleles that are not measured by the genotype tests. The FDA is in the process of revising the irinotecan label based on the recommendations of the CPSC and is deliberating all of these challenges in translating PGt data into useful information for practitioners and their patients.

7. CONCLUSION

The FDA has become a proactive and thoughtful advocate of PGx and believes, as a public health Agency, that it has a responsibility to play a leading role in bringing about the translation of PGx, as well as other emerging technologies, from bench to bedside to facilitate drug development and improve the benefit/risk of drug treatments in the marketplace. The FDA also realizes that it can hinder innovation and become a regulatory barrier

in the translational process if it is not careful with its guidances, policies, label updates, and procedures. The Agency hopes that pharmaceutical companies view advances in PGx as an opportunity in new drug development and one kind of investment in R&D that can help bring a fresh approach, and a partial solution, to addressing the "pipeline" problem outlined in the FDA Critical Path white paper. We believe that PGx has the potential to revolutionize the drug development process, making it more efficient through targeted enrichment trials, and bringing value to patient care, including more biomarker-based diagnostic or test products to individualize or target therapy with more precision. This may, in time, seem to have taken much longer than was anticipated but there are not an increasing number of examples of targeted therapies beyond Herceptin, and we expect that the pace of PGx will accelerate over the next 5 years. Regulatory agencies, pharmaceutical companies, the clinical community, third-party payers, and patient advocacy groups are all interested in strategies that can improve the cost, quality, and time of drug development, and reduce the risks associated with drug therapy in patients for both new and previously approved therapies. Clinical guidelines for use of PGx tests or laboratory guidelines for the interpretation of PGx tests are needed, and the FDA is partnering with professional organizations to develop such guidelines. We do not expect that big changes in these areas will happen overnight with one seminal event or be straightforward, but rather will occur in a more evolutionary or iterative manner built upon one successful application of PGx after another, that now seem to be occurring at a rapid rate as exemplified by thiopurines, irinotecan, warfarin, atomoxetine, and the other examples illustrated in this chapter. We acknowledge that there are and will be many different kinds of challenges in translating PGx from bench to bedside ranging from issues of historical clinical practices, concerns over cost effectiveness, test availability, and reimbursement by third party payers, to issues of biomarker discovery, analytical validation, qualification of clinical performance, health professional education, and adoption of PGx tests into routine clinical practice. But, as we have shown through examples with the promising results with gefitinib, erlotinib, SSRIs, and the tried and true examples of atomoxetine, 6MP, azathioprine, irinotecan, and warfarin, these challenges are being met and overcome to benefit both the science of drug development and the quality of public health. The FDA will be influential in a positive way and should play an important role in collaboration with others in translating the important discoveries of PGx from bench to bedside, and working with collaborators to interpret bedside findings in PGX to what might be needed in bench research to facilitate drug development.

But, regulatory agencies also need to be on guard. We are aware that drug development is a global enterprise, and we live in a small world, and thus international collaboration between regulatory agencies and other private

(e.g., World Health Organizations) and national government organizations (e.g., National Institutes of Health), and international government agencies (e.g., EMEA) must continue to work more to harmonize guidance and policies in a way that facilitates and not complicates the drug development process and facilitates the benefits of PGx to patients worldwide. We must also strive more to engage the various stakeholders and constituencies (including Health Maintenance Organizations) in both the private and public sectors in conversation regarding effective strategies to advance PGx through cost-effective, well-conceived clinical outcome studies that can either be randomized, adequate and well controlled, legitimate case-control observational studies, or naturalistic studies with prespecified goals to address the questions of clinical utility of genomic tests. It clearly is in everybody's interest to streamline the pre-approval drug development process (in terms of cost, time, and early attrition), reduce the likelihood of toxicity in the post-approval period and to improve the use of established drugs that have a history of frequent and severe adverse events. We hope that others view the FDA's important initiatives and transparent strategies—the Critical Path white paper, and its advocacy of PGx—as a willingness to work together to link bench discoveries to bedside benefits, and vice-versa, and we look forward to continued involvement with the scientific, medical, and health care community.

ACKNOWLEDGMENTS

We would like to acknowledge the helpful comments of our FDA colleagues, Dr. Felix Frueh, Dr. Shiew-Mei Huang, and Dr. Allen Rudman.

REFERENCES

1. Lesko LJ, Woodcock J. Pharmacogenomic-guided drug development: regulatory perspective. Pharmacogenomics J 2002; 2:20–24.
2. Lesko LJ, Woodcock J. Opinion: translation of pharmacogenomics and pharmacogenetics: a regulatory perspective. Nat Rev Drug Discov 2004; 9:763–769.
3. DiMasi J, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ 2003; 22:151–185.
4. Gilbert J. Rebuilding big PhRMA's business model. In: In vivo, The Business and Medicine Report, Windhover Information, Inc., November 2003:73–80.
5. Kenneth I Kaitin Tufts CDSS quantifies savings from boosting new drug R&D efficiency. Tufts CSDD Impact Rep 2002; 4(5):1–4.
6. Lesko LJ, Salerno RA, Spear BB, Anderson DC, Anderson T, Brazell C, Collins J, Dorner A, Essayan D, Gomez-Mancilla B, Hackett J, Huang SM, Ide S, Killinger J, Leighton J, Mansfield E, Meyer R, Ryan SG, Schmith V, Shaw P, Sistare F, Watson M, Worobec A. Pharmacogenetics and pharmacogenomics in drug development and regulatory decision making: report of the

first FDA-PWG-PhRMA-DruSafe workshop. J Clin Pharmacol 2003; 43: 342–358.
7. Salerno RA, Lesko LJ. Pharmacogenomics in drug development and regulatory decision-making: the genomic data submission (GDS) proposal. Pharmacogenomics 2004; 5(1):25–30.
8. Salerno RA, Lesko LJ. Pharmacogenomic data: FDA voluntary and required submission guidance. Pharmacogenomics 2004; 5(5):503–505.
9. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small cell lung cancer to gefitinib. N Eng J Med 2004; 350(21):2129–2139.
10. Guillermo Paez J, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib. Science 2004; 304:1497–1500.

11

Tools of the Trade: The Technologies and Challenges of Pharmacogenetics

Glenn A. Miller
Genzyme Corporation, Westborough, Massachusetts, U.S.A.

1. INTRODUCTION

As the 21st century begins the field of drug discovery is filled with opportunity and challenge. The ability of the pharmaceutical industry to take advantage of the opportunities and meet the challenges while maintaining a significant growth rate will be determined by the success with which the industry makes use of the rich sources of information at its disposal.

The field of pharmacogenetics is one of the many disciplines that will produce insights and avenues of opportunity for drug discovery. The facility with which the pharmaceutical industry uses this information will have a major impact on the safety, efficacy and time to market of a wide variety of therapeutic compounds in the future.

The use to which pharmacogenetics is put will be determined by how the information is collected, organized, and disseminated. The technologies brought to bear on the problems that pharmacogenetics present will be critical to the effective use of pharmacogenetics information in the therapeutic development process. The current state of technological development has permitted investigators to address an array of previously unapproachable genetic questions. This has, in turn, created an ever-growing library of DNA sequence information, cDNA clones, genes, mutations, and polymorphisms available for use in target discovery, lead identification, lead optimization, pre-clinical testing, and clinical trials. The high per patient

cost of these technologies do not, however, permit the wide scale use of this information resource in large trial populations. The development of new technologies providing low cost, high throughput alternatives to genetic testing will permit the application of pharmacogenomic information across a broad range of therapeutic development. This review will endeavor to provide an overview of the current state of the art in genetic technologies along with those technologies that may facilitate the wide spread use of pharmacogenetics information.

2. TECHNOLOGY SELECTION

The selection of an appropriate technology is best based on the requirements of a project or area of investigation. Technologies that are appropriate for gene discovery are not universally applicable to population studies or patient group stratification.

Figure 1 depicts the most common areas of use for genomic and genetic technology in the drug development pathway. As the process moves from initial target identification through development of lead compounds the impact of genomic technology lessens while that of genetics technology increases. The impact of the limited number of technologies applicable to clinical diagnostic laboratory use is reserved for those compounds well into the development process where a genetic test is being considered as an adjunct or requirement for therapeutic selection.

The selection of an appropriate technology can take two forms depending upon the state of knowledge concerning a gene, locus, or disease. In one form technology selection will be guided by a relative lack of knowledge concerning mutations and polymorphisms. The technologies selected in

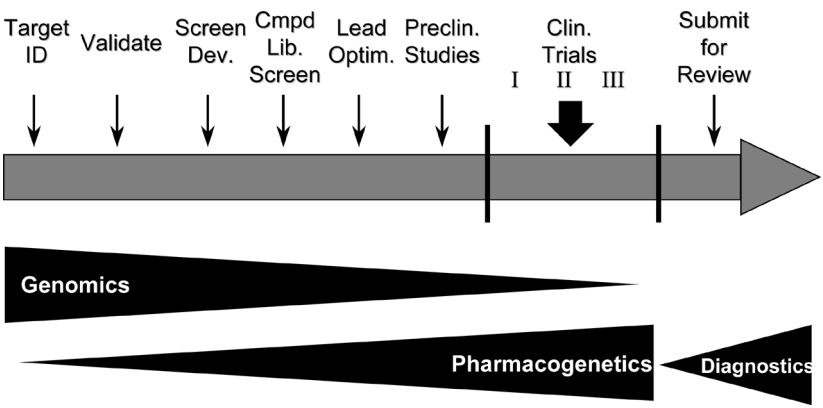


Figure 1 The steps involved in drug development and the relative use of genomics, pharmacogenetics and diagnostics during this process.

this manner will rely heavily on gene scanning methods and/or sequencing. In another form technology selection will be guided by a relative abundance of knowledge. The technologies selected in this environment will rely on high throughput, reliable, robust methods of identifying known mutations or polymorphisms. The driving force behind the selection of technology is the detailed knowledge of gene sequence and variation. Since the first edition of this volume a great deal of genomic and genetic information has entered the public domain. The ability of an investigator to obtain detailed DNA sequence information online has greatly facilitated the inclusion of genetic information in drug discovery.

2.1. Identification of Gene Variants

The completion of the final version of the sequence of the human genome provides an unparalleled resource for the identification of mutations and polymorphisms. The eventual knowledge of the location and sequence of genes important to the cause and treatment of disease will permit the rapid collection of sequence variations in a wide variety of potential patient populations. In advance of that a collection of governmental and commercial efforts have sought to identify as large a pool of sequence variants as possible and place them in the public domain. The SNP Consortium brought together a number of commercial entities in an effort to identify a large number of single-nucleotide polymorphisms (SNP). This effort was successful beyond its initial goal of 300,000 SNPs eventually releasing a total of over 1.2 million variants in a publicly available database (http://snp.cshl.org). The availability of this SNP resource is a valuable starting point in any pharmacogenetics study. A limitation to these databases is the limited amount of population frequency information in the public domain. There remains a need to both identify and characterize sequence variants in patient populations. The tools used in any variant discovery fall into two categories, those technologies that are information rich but relatively costly in capital and man-hour terms (i.e., sequencing) and those technologies that are less information rich but relatively less expensive (i.e., gene scanning).

2.2. Brute Force Variant Identification

The use of DNA sequencing to identify population wide variation is, except for the largest commercial and governmental operations, a labor intensive and costly exercise. With the completion of the Human Genome Project the task of finding appropriate primer sites for both PCR and sequencing became greatly simplified. The most practical sequencing based method of variant detection is via one of the various automated DNA sequencers currently on the market. All take advantage of substantially similar sequencing chemistry while using a variety of labeling technologies. The major difference between manufacturers, marketing claims aside, is the number

of different labels that can be loaded into one lane of a gel or capillary at one time and the speed with which the labeled sequence can be run by an appropriate detector. By detecting multiple labels one is able to perform all four sequencing reactions in a single tube and/or load more than one sequencing reaction into a sequencing gel lane. The throughput advantage of automated sequencers able to detect multiple labels is a strong point in their favor when variant detection on a large scale is planned. For most laboratories, however, variant detection via sequencing remains less a completely automated collection of data and more a task composed of visually sifting through quantities of sequence data in search of variants from a "normal." Advances in DNA analysis software have automated a portion of this task but the efficiency of this process varies by sequencing chemistry and automated analyzer. The major difficulty in sequence analysis is the determination of heterozygotes and true SNPs vs. sequencing artifact. The SNP Consortium has addressed this difficulty in two ways. The genome centers that identified the sequence variants re-sequenced randomly selected SNPs for verification and, second, the consortium contracted with Orchid Biosciences (Princeton, NJ) and Celera (Rockville, MD) to verify and develop assays for randomly selected SNPs (1). In this way it is hoped that the accuracy level of SNPs collected by the Consortium will reach a 95% standard level.

When one looks back on the SNP discovery process in the future it is likely that the majority of sequence variants will have been identified using a brute force sequencing approach. This will no doubt be as a direct result of the amount of raw sequencing horsepower being applied to the task. There will be, however, a portion of the SNPs that will have been identified through the use of a variety of gene scanning methods. The use of such methods is appropriate for those laboratories that do not have access to large scale sequencing operations.

2.3. Gene Scanning Technologies

The identification of sequence variants in the absence of a large scale sequencing capability can be accomplished through the use of gene scanning techniques. This collection of technologies is capable, at varying degrees of precision, of comparing a "normal" DNA (and in some instances RNA) sample to a test sample. The result of this comparison is either complete identity with the standard sequence or the identification of a variant at some point within the sequence under comparison. The common weakness of virtually all such scanning technologies is the requirement for a subsequent sequencing step to characterize the variant.

The technologies that purport to identify sequence variants generally fall into two categories:

- gel shift assays,
- mismatch identification.

2.3.1. Gel Shift Assays

The mobility of DNA, either single or double stranded, through a gel matrix is determined by a number of factors. By comparing the mobility of a test DNA sample to that of a normal control sample it is possible to determine the presence of a sequence alteration in the test sample. The major factors that affect DNA mobility are listed below:

- sequence length,
- base composition,
- single strand secondary structure,
- double strand melting characteristics,
- double strand mobility in the presence of mismatched strands.

A variety of techniques are used to take advantage of these factors in order to discover previously unknown mutations and polymorphisms. Two excellent reviews of these technologies have been published by Cotton (2) and Schafer and Hawkins (3). The following list outlines several of these techniques and a selection of their related methods.

- Single strand conformation polymorphism (SSCP):
  - restriction endonuclease fingerprinting (REF),
  - dideoxy fingerprinting (ddF).
- Denaturing gradient gel electrophoresis (DGGE):
  - temperature gradient gel electrophoresis (TGGE).
- Heteroduplex analysis (HA).

2.3.2. Single Strand Conformation Polymorphism

This technique is perhaps the easiest of all of the variant detection methods to establish and perform. Developed by Orita et al. (4), PCR amplified fragments are heat denatured and then rapidly cooled to prevent reassociation of the complementary strands. The single stranded DNA is then electrophoresed through a native acrylamide gel. The rate of transit through the gel is dependent upon sequence length, and single strand secondary structure. Variations in the sequences of single stranded DNA fragments will, in theory, produce a change in the secondary structure of the DNA thereby affecting its progress through the gel matrix. It is this change in mobility that is resolved by the gel matrix producing additional or displaced bands as compared to a normal control. The resolving power of the gel matrix is affected by temperature and gel additives such as glycerol, which enhance variant detection. The detection rates for SSCP can fall into the 60–95% range depending upon the sequence context in which the variants occur. By maintaining a relatively small fragment size ($\sim$200 bp) and selecting

appropriate gel conditions it is possible to routinely achieve variant detection rates of greater than 85%.

2.3.3. Restriction Endonuclease Fingerprinting

Originally described by Liu and Sommer (5), this technique takes advantage of the capabilities of SSCP and Restriction Fragment Length Polymorphism (RFLP) analysis to increase the amount of information available to an investigator in an alteration discovery process. A restriction fragment analysis of a normal control region of interest is completed to select a collection of restriction enzymes capable of generating a series of fragments resolvable on a standard acrylamide gel. A PCR amplified segment of DNA up to 1–2 kb in length is restricted using the pre-selected enzyme set. The fragments are then radioactively or fluorescently labeled, heat denatured and run as single stranded segments on a native acrylamide gel. The electrophoretic step is identical to an SSCP gel with the resulting fragment pattern providing multiple bands whose mobility would be potentially affected by an alteration present in a fragment. In addition, the loss or gain of a restriction site would be identified by the gain or loss of an expected fragment and the increase or decrease in length of another fragment. Thus, REF provides two concurrent methods of alteration analysis, SSCP with the advantage of multiple analyzable bands and an RFLP component. The major advantage of this technique is the increased number of bands that are available for analysis using the SSCP component. While the RFLP analysis component is useful it is not likely to provide consistently valuable information, as the statistical likelihood of an alteration creating or eliminating an enzyme site that is also a member of the REF set is low. The multiple banding patterns, however, creates several opportunities for shifts in mobility to be identified. The increased fragment size available for analysis is also a major advantage of REF over SSCP and favorably impacts throughput for this technique. The detection rate for this technique has been reported to be in the 90–95% range (5). One of the difficulties in this technique is the selection of the enzyme set. It is often a difficult task of balancing informativeness, exotic enzyme expense and fragment size. A subsequent publication from the Sommer group described a computer program, which aided the enzyme selection process (6). As with SSCP, the detection rate is dependent upon conformation changes and their resolvability on an acrylamide gel. The read-out of the assay provides a great deal of information for analysis, which can be difficult to interpret on a routine basis. Precise localization of a variant for subsequent sequence identification is also difficult given the number of fragments, which must be analyzed.

2.3.4. Dideoxy Fingerprinting

This technique uses a combination of SSCP and Sanger sequencing to create a larger set of bands for analysis in a native acrylamide gel (7). In this method

an amplified fragment is sequenced using a standard Sanger protocol and primers spaced at a SSCP friendly size of approximately 200–250 bp throughout the amplified fragment. One modification to the Sanger protocol is the use of a single dideoxy nucleotide to create the banding pattern. The change in banding pattern as compared to a normal control is then indicative of a sequence alteration in that region. This technique has been adapted to fluorescence as well as using primers running in both directions to increase detection rates. Through the appropriate selection of dideoxy nucleotide, fragment size and gel running conditions a significant improvement in detection rate over SSCP can be achieved. In addition, a degree of alteration localization can be determined by noting the location of the shift in banding patterns. The ability to perform a single PCR amplification followed by multiple Sanger reactions also reduces the overall cost of the assay. The fragment size used for variant detection remains in the SSCP range thus decreasing assay throughput. The original paper (7) suggests that reaching the 90–95% detection rate range may be difficult for some genes and some mutation types. In this event ddF would be useful as an adjunct to SSCP rather than a subsitute.

2.3.5. Denaturing Gradient Gel Electrophoresis

This technique makes use of the reduced double stranded melting characteristics inherent in heteroduplexed DNA as compared to a normal homoduplex control to identify novel sequence alterations (8). A key modification to DGGE was made by Myers et al. (9,10) whereby a DNA sequence of interest is amplified using a specialized PCR primer pair consisting of one standard, sequence specific primer and an opposing primer containing sequence specific information at the primer's 3′ end and a GC enriched sequence at the primer's 5′ end. The purpose of the GC enriched sequence (often called a GC clamp) is to normalize the melting profile of the sequence of interest. The GC clamp acts as a high melting temperature region of the resulting amplicon stabilizing the melting characteristics over the entirety of the sequence. Without this normalization of amplicons containing multiple melting domains would denature at different temperatures. The lower melting temperature domains would denature first obscuring the decrease in melting temperature contributed by sequence variants present in higher temperature melting domains. In performing DGGE the sample is heat denatured, allowed to slowly re-anneal and then electrophoresed through an acrylamide gel containing a gradient of denaturant (usually formamide). As the double stranded amplicon progresses through the gel matrix it encounters an increasing denaturant concentration and the amplicon begins to denature. This denaturation slows the progression of the amplicon through the gel. Any alteration in a heteroduplex DNA sequence will result in decreased mobility yielding a multiple banding pattern. By comparing the extent of mobility in the gradient matrix to a normal control it is possible to

determine the presence of a sequence change. In theory, any sequence alteration, which changes the melting characteristics, will result in additional bands being produced on the gel. The melting profile of any sequence can be predicted using an algorithm created by Lerman and Silverstein (11). If the sequence under investigation is amplified as a single melting domain using the GC clamp method it is then possible to design denaturing gradient conditions under which all sequence alterations in that single domain would be identified. The commonly accepted detection rates for DGGE are in the 90–95% range given an amplicon containing a single melting domain. The size of the routinely analyzable fragments (up to ~650 bp) is greater than the 200–250 bp for SSCP. The pouring of gradient gels is a technically difficult task requiring considerable experience and patience to achieve consistent results. The cost of the clamped primers is greater than non-clamped primers due to the addition of as many as 40 or more additional bases at the 5′ end of the primer to produce the clamp. Alterations closer than 40–50 bp from the 3′ end of the clamped primer are unlikely to be identified due to the increase thermal stability of this region. This requires the design of a second set of clamped primers opposing the initial set of primers for each region of interest. In this manner, the region nearest to one clamp, and thus refractory to analysis, is accounted for by 3′ end of the opposing clamped amplicon. This necessarily doubles the cost of the assay from the PCR step onwards.

2.3.6. Temperature Gradient Gel Electrophoresis and dHPLC

These variants of DGGE make use of carefully controlled temperature gradients as a substitute for chemical denaturation. The simplest form of this method from an equipment standpoint immerses a uniform denaturant gel in a temperature controlled buffer chamber. The buffer temperature is gradually increased during the running of the gel providing the denaturing gradient range desired.

Another modification of this type of denaturing gradient methodology uses the high resolution characteristics of high performance liquid chromatography to physically separate sequences which differ in their denaturing characteristics (12). The main advantage is the rapid nature of HPLC coupled with the fine temperature control of HPLC resulting in an ability to resolve small differences in melting characteristics. The semi-automated manner in which the samples can be loaded into the HPLC and the potential for customizing the run characteristics is also a favorable feature for this approach. The use of dHPLC does not provide any inherent improvements over traditional DGGE other than in the area of automation, as detection for all variants must still labor under the constraints of DNA melting theory. The major advantage of dHPLC over DGGE is more one of process and throughput for the end user.

2.3.7. Heteroduplex Analysis

This technique makes use of the change in mobility caused by mismatched heteroduplexed DNA strands as compared to perfectly homoduplexed DNA strands (13). In this technique a DNA sample is amplified, heat denatured and then reannealed. This results in a fraction of the single stranded DNA re-annealing with a complementary strand from an opposing allele thus forming a heteroduplex. When this mixture is subsequently electrophoresed through a non-denaturing acrylamide gel a mobility shift can be resolved indicating the presence of a sequence alteration as compared to a normal control homoduplex. The relative ease of this protocol is the major advantage to this technique. The generally accepted detection rates for a specific PCR run on an appropriate acrylamide gel ranges from the low 80% to 90–95% range. Modifications to the assay to generate heteroduplexes may increase the detection rate to a more routine 90–95%. The major difficulty with this technique is the establishment of sufficiently specific PCR amplification conditions to avoid missing important heteroduplexes or the creation of false alterations. The lack of resolution of the acrylamide matrix is another area of false negative results. In practice, the amplicon size range for routine detection is from 250 to 500 bp thus limiting the use of this technique for very large genes.

2.4. Mismatch Identification

The discovery of sequence alterations can be accomplished by the creation of mismatched base pairs, which occur when a test sample containing a sequence alteration is denatured and re-annealed with an otherwise complementary normal control sequence. The routine and specific discovery of these mismatches and their localization to within an easily sequenced region is the goal of several techniques. There are two general methods for identification of mismatched sequences:

- chemical,
- enzymatic.

2.4.1. Chemical

The major method involving chemical identification of mismatched sequences is a modification of Maxam–Gilbert sequencing termed Chemical Cleavage of Mismatch (CCM) (14). Using the C- and T-specific reactivity of hydroxylamine and osmium tetroxide, respectively, to identify mismatched bases followed by piperidine cleavage of the modified products it is possible to not only scan a genomic segment for mismatched bases but also localized the site of the mismatch. When optimized, this technique is highly efficient in alteration discovery (nearly 100%). The current practical limit of the size of a fragment suitable for analysis is in the 1–1.5 kb range. A considerable

disadvantage, however, to CCM is the use of highly toxic chemicals in a clinical environment. Hydroxylamine and osmium tetroxide both require chemical grade fume hoods for use. The relatively large number of manipulations associated with the technique and the toxic nature of the chemicals render this technique a difficult choice for a clinical lab. A recent improvement on the assay has substituted potassium permangenate for osmium tetroxide making the manipulations involving these chemicals less toxic and more attractive for the routine laboratory (15).

2.4.2. Enzymatic

This set of methods makes use of specific enzymes to recognize mismatched base pairs in heteroduplexed DNA strands. In general a test sample is amplified and heat denatured in the presence of a normal control sample. This combination of test and normal control samples is allowed to slowly reanneal forming a mixture of homo and heteroduplexes. The mixture is then exposed to any one of a number of enzymes, which recognize mismatched base pairs at various efficiencies and the sample is then analyzed for these recognized mismatches. There are several types of enzymes and assays available which purport to recognize mismatched base pairs in a heteroduplex or secondary structures formed by mismatches. A short selection of these enzymes and assays are listed below:

- ribonuclease cleavage,
- T4 endonuclease VII,
- Cleavase® (Third Wave Technologies, Madison, WI, USA).

2.4.3. Ribonuclease Cleavage

This technique, developed by Myers et al. (16) is useful for sequence alteration discovery due to its ability to cleave double stranded RNA or RNA: DNA hybrid sequence at areas of mismatch. The mismatched areas form single stranded regions in a heteroduplex, which are then amenable to cleavage by the RNAses. A significant improvement in the assay was published by Murthy et al. (17) whereby an amplified product is produced following a standard PCR of regions of interest. The test sample DNA is then mixed with a normal control RNA sample to produce a normal/test duplex, which creates the potential for mismatched heteroduplexes open to cleavage by RNAse. The cleaved products are then electrophoresed on a standard agarose gels (this is the most recent form of the assay which has been reduced to its simplest practice) and the banding pattern is compared to a normal control. Relative to SSCP and DGGE this technique offers slightly longer analyzable read lengths and the potential for multiple bands for analysis. The RNA handling requirements, however, are often a negative factor in a high throughput or clinical environment.

2.4.4. T4 Endonuclease VII

Mashal et al. (18) originally described the use of two bacteriophage resolvases to identify mismatches in heteroduplexed DNA. The current technique depends upon one of those enzymes, T4 endonuclease VII, to recognize and cleave double stranded DNA at the site of the mismatch. This technology is now incorporated in kit form by Amersham Pharmacia Biotech (Piscataway, NJ). The length of fragment amenable to this sort of scanning is approximately 1 kb. The assay is essentially the same as a restriction enzyme analysis with respect to ease of use. The size of the generated cleaved fragment is then used to roughly localize the site of the alteration for subsequent sequencing.

2.4.5. Cleavase

This enzyme cleaves secondary structure in single stranded DNA. The principle used here, as described by Brow et al. (19), is one of secondary folding structure of single stranded DNA being altered when the underlying sequence is changed by polymorphism or mutation. Cleavase acts as a structure specific endonuclease by cutting at the 3′ end of a loop structure in a single stranded DNA producing a characteristic banding pattern. When compared to a normal control sample a cleaved test sample containing a different base composition would demonstrate an altered banding pattern. This method has been developed for commercial use by Third Wave Technologies (Madison, WI).

The main advantage of any enzyme-based gene scanning system is the relative ease of use and cost effectiveness when compared to full gene sequencing. The ability to scan through large amounts of genetic real estate identifying only those individuals demonstrating variations from a control reduces the overall amount of DNA sequencing required to characterize the panoply of genetic variation in a population. A significant disadvantage, however, is the amount and types of variations that the enzymatic methods have difficulty identifying. Depending upon the sequence context and variant type the enzyme may not recognize the mismatched base pair leading to a false negative result. This may not be a significant difficulty in the context of a large variant discovery effort where missing one polymorphism is not of great import. In a clinical diagnostic effort where each patient sample must be completely and accurately scanned enzymatic based assays must be designed with great care to reduce the false negative rate to an absolute minimum. All of the described methods are now offered commercially so it remains the consumers challenge to sort through the various detection claims and identify the enzyme which works best in their hands and for the sequence to be scanned.

2.5. Variant Identification

The technologies available to the research pharmacogenetics community are varied and numerous. The task confronting the end user in the selection of the appropriate tool set is one of overall throughput, accuracy and thoroughness. How one goes about using the vast quantities of variants, which are discovered is the subject of the next section.

2.6. Testing for Previously Identified Gene Variants

The development of large and publicly available SNP and mutation databases provides ample intellectual material for pharmacogenetics. The real value of this information glut will be in using it to design appropriate therapeutics and to select individuals most likely to benefit from these new treatments. The techniques and reagents, which were included in the description of variant identification methods, are not universally applicable to the task of genotyping individuals on a clinical basis. There are several technologies, which are useful to perform this type of known variant detection. Virtually all of these technologies use probes of one sort or another to detect variants. The probes in these cases are highly specific for a particular known variant and are therefore not useful for generalized gene scanning. The universe of variant detection can be broken into two categories: array based and non-array based.

2.6.1. Array-Based Systems

The use of a fixed probe system with a test sample in solution has promised to revolutionize genetic testing for nearly a decade (20,21). The promise has not entirely become a reality as of yet but the uses of arrays have increased over the decade to accommodate the potential to screen individuals for thousands of discrete polymorphisms in a single hybridization (22). In fixed array systems the probes are attached to a surface by a number of methods. The method pioneered by Fodor et al. (20) uses a modification of the photolithographic system used to manufacture computer chips. In this method a mask is placed over the surface of the DNA chip covering those regions to be protected from the light activated oligonucleotide synthetic process. The process is completed and the mask realigned to uncover subsequent areas. In this manner a series of chemical reactions can take place in a defined order resulting in the building of an oligonucleotide probe of a specific design at a known location on the array. By carefully controlling the synthetic process it is possible to increase the number of probes on the array to the hundreds of thousands or greater. Other fixed array types attach fully pre-constructed probes to a number of different substrates by a variety of chemical means (23–27).

The process of analyzing a sample on fixed arrays generally takes place in several steps. The sample must first be amplified for the specific region of

the genome to be probed. The sample must then be fragmented in order for optimal hybridization to a static probe. The hybridization then occurs and analysis of the hybridization pattern takes place. The number of probes able to be queried on a fixed array is not equal to the number of separate elements fixed to the substrate. Due to the detection limits of the fluorescent technology, which underlies the majority of fixed array systems multiple probe elements must hybridize to the test sample in order for a sufficient signal to be measured. The hybridization characteristics of oligonucleotide probes are often difficult to normalize when dealing with large numbers of probes. This is exacerbated when a probe must be placed in a sequence context of unfavorable hybridization characteristics due to the presence of an important sequence variant. As a result there are regions of the genome and types of variants that are not easily amenable to oligonucleotide hybridization. Nanogen (San Diego, CA) has approached the hybridization problem with a methodology involving control of the hybridization and denaturation of disparate oligonucleotides by varying the electric field strength of the hybridization repelling molecules with reduced affinity for the probe (28).

One of the drawbacks of a fixed array system in an era where not all the relevant information concerning a biologic system may be present is the relatively inflexible nature of a fixed system. In order to add elements to a fixed array one needs to return to the chip manufacture process and redesign the array. Not all array-based systems, however, make use of a fixed solid support. Two types of arrays make use of bead technologies to permit rapid rearrangement and alteration of the components of an array. Luminex (Austin, TX) attaches oligonucleotide probes to polystyrene beads that have been colored with one of a number of colors separately identifiable by a fluorescence activated cell sorter. It is then possible to perform standard oligonucleotide hybridization to a collection of these probe bound beads using test sample material that has been labeled with another identifiable fluorescent tag. The beads are then washed and run through the cell sorter producing a profile of test sample bound and unbound beads. The colors of the bound beads are then correlated to the type of oligonucleotide previously attached to them and a positive identification of the probe is made. This process has the advantage of speed and thus throughput as well as flexibility of array composition. In order to change the array one only needs to mix the appropriate combination of beads. A second type of non-fixed array was developed at Tufts University using beads similarly color coded with oligonucleotide probes covalently attached (29). In this method, however, the beads are deposited onto the ends of fiber optic cable following a process of etching carefully controlled pits into the end of the fiber optic. The amplified test sample is labeled with a fluorescent tag and is hybridized to the beads. The laser excitation of the fluorescent tag and the colored beads provides evidence of a positive hybridization and identification of the probe.

This technology has been further developed and commercialized at Illumina (San Diego, CA).

2.6.2. Non-Array-Based Systems

Hybridization based array systems are not the only method of analyzing test samples for known genetic variants. One of the most effective methods of analyzing SNPs is a technique known as mini-sequencing or single-base extension. This technique was described in detail in the first edition of this text and is used by Orchid BioScience (Princeton, NJ) as well as Sequenom (San Diego, CA) to type SNPs. The method takes advantage of DNA sequencing technology and the chain terminating effect of dideoxynucleotides to determine the presence of a particular nucleotide at a known position. A primer specific for a region immediately adjacent to the SNP under study is hybridized to a test sample. A dideoxynucleotide corresponding to one of the two bases indicative of the SNP is then added to the reaction. If the SNP corresponding to the particular dideoxynucleotide is present the reaction is terminated. By labeling the dideoxynucleotide with a fluorescent tag (or a mass tag for mass spectrometry based methods) it is possible to perform these assays in an automated manner increasing throughput and decreasing manpower requirements.

Another SNP detection system that does not rely on fixed arrays has been developed by Third Wave Technologies (Madison, WI). The Invader assay (30–32) makes use of two probe molecules, which partially overlap a known SNP site. The probes compete for hybridization to the specified genetic sequence. When one probe binds specifically to the test sample it forces the overlapping probe to leave a portion of the overlapping region non-hybridized. The Cleavase endonuclease described earlier in this review then cleaves the overlapping flap releasing a small fragment, which is either directly labeled or used as an ''invading'' oligonucleotide itself in a secondary reaction, which releases labeled sequence producing a detectable signal. This isothermal reaction is compatible with several assay formats and has been developed for a number of clinical assays (32,33). The isothermal nature of the assay and the single base discriminatory nature of the probe reactions make this system an attractive choice for those laboratories seeking an alternative to PCR-based systems.

3. REDUCTION TO PRACTICE

The construction of large-scale databases and genotype/phenotype correlations within large population groups is of great value to the future of medical care. This large-scale research effort will create opportunities for the genetic testing community previously seen as mere science fiction. For the testing community to take advantage of these opportunities a shift in technologies must take place on a scale greater than currently available.

The volume of assays that might be expected over the next decade or more will be far closer to that seen in the clinical chemistry laboratories. If one considers the development of clinical chemistry testing on an industrial scale a few similarities to the challenges that pharmacogenetics will face in the future can be seen. The major development, which permitted clinical chemistry to achieve the volumes of testing that laboratories currently handle was the wide scale adoption of automation. The modern clinical chemistry laboratory receives samples for testing, bar codes each sample and places the sample tube into an automated system. This system reads the bar code, consults a database to determine the requested testing and proceeds to perform the necessary procedures often without any human intervention. There is no currently available system for widespread genetic analysis that approaches this level of automation for even the most straightforward of test protocols. As it is unlikely that a pharmacogenetics test will involve a single SNP as the definitive predictor of therapeutic response a new generation of molecular genetic technologies and automation must be developed that can perform this highly complex testing in a clinical chemistry-like atmosphere. An example of the scale that the testing industry will have to face in the not too distant future is that presented by the volume of testing for a pharmaceutical that has a rather modest market size and a different pharmaceutical with a large market size. One of the central assumptions often made in pharmacogenetics is that there will be a growing number of pharmaceutical products whose prescription will be linked to a genetic test. This will probably not be a safety related issue but one of selecting those patients who will benefit most from being treated by a drug or to determine an appropriate dosage level. If one creates a hypothetical drug whose market size is approximately $200 million the monthly new prescriptions can be in the range of 18,000. The market size for such a genetic test would be over 200,000 tests per year. If one increases the market for the drug to the $1 billion size the monthly new prescription rate can equal the number of prescriptions written in an entire year for the smaller market example. In this scenario the yearly testing market for a $1 billion drug with a genetic test linked to prescription would be over 2 million tests per year. To put this into context, the current number of tests for carrier status for cystic fibrosis in the United States is approximately 300,000–500,000 tests annually. The molecular genetics technology development community and the genetic testing community are only now developing the collective ability to perform that level of genetic testing. It will be some time before the level of automation seen in the clinical chemistry laboratory reaches the molecular genetics laboratory. The caveat often linked to the foregoing argument, however, is that there are unlikely to be drugs with companion diagnostics that reach the $1 billion market level. The problem becomes essentially identical to the billion-dollar drug issue, however, when one envisions multiple smaller drugs on the market that have a linked genetic test for therapeutic efficacy.

This is an issue that will only increase in seriousness as more information concerning genetic markers of drug response become an increasing part of routine medical practice.

4. THE FUTURE

The promise of pharmacogenetics to enhance the future of pharmaceutical development and medical care has been discussed in this volume as well as a number of others. The technological task challenging the scientific community is twofold. The first challenge is identifying the relevant genetic markers that will provide significant and medically relevant information to the clinician. This effort will require the establishment of large, well-annotated databases linking information concerning disease pathogenesis, functional genomics, proteomics, and genetic variation. The most useful data will be that correlating the large databases with sub-populations of individuals particularly susceptible to severe disease, particular sequelae or in need of or obtaining a particular benefit from a specific treatment regimen significantly different from a larger population group. As the future of pharmacogenetics unfolds it will hold opportunities for the developers of novel, highly specific and efficient technologies, which provide valuable information to the medical community. Those successful in the development effort will have solved the many-layered problem of medical relevance, technologic possibility and cost effectiveness. The technologies and concepts described in this volume go a long way down this path.

REFERENCES

1. Holden A. The SNP consortium: a case study in large pharmaceutical company research and development collaboration. J Commun Biotech 2000; 6(4): 320–332.
2. Cotton RG. Mutation detection and mutation databases. Clin Chem Lab Med 1998; 36(8):519–522.
3. Schafer A, Hawkins JR. DNA variation and the future of human genetics. Nat Biotech 1998; 16:33–39.
4. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA 1989; 86:2766–2770.
5. Liu Q, Sommer SS. Restriction endonuclease fingerprinting (REF): a sensitive method for screening mutations in long, contiguous segments of DNA. Biotechniques 1995; 18(3):470–477.
6. Scaringe WA, Liao D, Liu Q, Sommer SS. REF Select: expert system software for selecting restriction endonucleases for restriction endonuclease fingerprinting. Biotechniques 1999; 27(6):1188–1190, 1192–1194, 1196 passim.
7. Sarkar G, Yoon HS, Sommer SS. Dideoxy fingerprinting (ddE): a rapid and efficient screen for the presence of mutations. Genomics 1992; 13(2):441–443.

8. Fischer S, Lerman LS. DNA fragments differing by single base pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. Proc Natl Acad Sci USA 1983; 80:1579–1583.
9. Myers RM, Fischer SG, Lerman LS, Maniatis T. Nearly all single base subsitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res 1985; 13(9):3131–3145.
10. Myers RM, Fischer SG, Maniatis T, Lerman LS. Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. Nucleic Acids Res 1985; 13(9):3111–3129.
11. Lerman LS, Silverstein K. Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. Methods Enzymol 1987; 155:482–501.
12. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 1997; 7(10):996–1005.
13. White MB, Carvalho M, Derse D, O'Brien SJ, Dean M. Detecting single base substitutions as heteroduplex polymorphisms. Genomics 1992; 12:301–306.
14. Cotton R, Rodrigues NR, Campbell RD. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc Natl Acad Sci USA 1988; 85: 4397–4401.
15. Roberts E, Beeble VJ, Woods CG, Taylor GR. Potassium permanganate and tetraethylammonium chloride are a safe and effective substitute for osmium tetroxide in solid-phase fluorescent chemical cleavage of mismatch. NAR 1997; 25:3377–3378.
16. Myers R, Larin Z, Maniatis T. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. Science 1985; 230(4731):1242–1246.
17. Murthy K, Shen SH, Banville D. A sensitive method for detection of mutations—a PCR-based RNase protection assay. DNA Cell Biol 1995; 14(1): 87–94.
18. Mashal R, Koontz J, Sklar J, Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. Nat Genet 1995; 9(2):177–183.
19. Brow M, Oldenburg MC, Lyamichev V, Heisler LM, Lyamicheva N, Hall JG, Eagan NJ, Olive DM, Smith LM, Fors L, Dahlberg JE. Differentiation of bacterial 16S rRNA genes and intergenic regions and Mycobacterium tuberculosis katG genes by structure-specific endonuclease cleavage. J Clin Microbiol 1996; 34(12):3129–3137.
20. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science 1991; 251(4995): 767–773.
21. Erlich H, Bugawan T, Begovich AB, Scharf S, Griffith R, Saiki R, Higuchi R, Walsh PS. HLA-DR, DQ and DP typing using PCR amplification and immobilized probes. Eur J Immunogenet 1991; 18(1–2):33–55.

22. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP. Accessing genetic information with high-density DNA arrays. Science 1996; 274(5287):610–614.
23. Zammatteo N, Jeanmart L, Hamels S, Courtois S, Louette P, Hevesi L, Remacle J. Comparison between different strategies of covalent attachment of DNA to glass surfaces to build DNA microarrays. Anal Biochem 2000; 280(1):143–150.
24. Rehman FN, Audeh M, Abrams ES, Hammond PW, Kenney M, Boles TC. Immobilization of acrylamide-modified oligonucleotides by co-polymerization. Nucleic Acids Res 1999; 27(2):649–655.
25. Rogers YH, Jiang-Baucom P, Huang ZJ, Bogdanov V, Anderson S, Boyce-Jacino MT. Immobilization of oligonucleotides onto a glass support via disulfide bonds: a method for preparation of DNA microarrays. Anal Biochem 1999; 266(1):23–30.
26. Beier M, Hoheisel JD. Versatile derivatisation of solid support media for covalent bonding on DNA-microchips. Nucleic Acids Res 1999; 27(9):1970–1977.
27. Okamoto T, Suzuki T, Yamamoto N. Microarray fabrication with covalent attachment of DNA using bubble jet technology [see comments]. Nat Biotechnol 2000; 18(4):438–441.
28. Sosnowski RG, Tu E, Butler WF, O′Connell JP, Heller MJ. Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. Proc Natl Acad Sci USA 1997; 94(4):1119–1123.
29. Healey BG, Matson RS, Walt DR. Fiberoptic DNA sensor array capable of detecting point mutations. Anal Biochem 1997; 251(2):270–279.
30. Kwiatkowski RW, Lyamichev V, de Arruda M, Neri B. Clinical, genetic, and pharmacogenetic applications of the Invader assay. Mol Diagn 1999; 4(4): 353–364.
31. Cooksey RC, Holloway BP, Oldenburg MC, Listenbee S, Miller CW. Evaluation of the invader assay, a linear signal amplification method, for identification of mutations associated with resistance to rifampin and isoniazid in Mycobacterium tuberculosis. Antimicrob Agents Chemother 2000; 44(5):1296–1301.
32. Ryan D, Nuccie B, Arvan D. Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. Mol Diagn 1999; 4(2):135–144.
33. Hessner MJ, Budish MA, Friedman KD. Genotyping of factor V G1691A (Leiden) without the use of PCR by invasive cleavage of oligonucleotide probes. Clin Chem 2000; 46(8 Pt 1):1051–1056.

12

Technologies for the Analysis of Single Nucleotide Polymorphisms—An Overview

Denis M. Grant

Department of Pharmacology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

1. INTRODUCTION

From the preceding chapters, it is evident that there is a critical need to better understand the instances where pharmacogenomics will be of value for the pre-market development or the post-market prescribing of individualized medicines. In order to enable both the experimental assessment of the utility of pharmacogenomics and ultimately its reduction to routine clinical practice where appropriate, there is an ongoing need for improved assay methods for the analysis of single nucleotide polymorphisms (SNPs), the most abundant type of genetic variant in the human genome. The purpose of the present chapter is to provide a general discussion of strategic and experimental issues related to the choice of suitable genotyping methodologies for various research and clinical settings. The reader is referred to other chapters in this monograph that provide further technical details of many of the emerging SNP genotyping technologies that will be briefly summarized here.

2. SNP ANALYSIS TECHNOLOGIES: BIOCHEMISTRIES, READOUTS, AND PLATFORMS

When comparing currently available and newly emerging technologies for their suitability in SNP genotyping assays, it is important to consider both the needs and the resources of potential end users. In this regard, study design and/or sample number can determine, and be limited by the availability of, a suitable genotyping technology. Assay cost is an important consideration for the choice of a method, and can place significant practical constraints on study design. On one hand, in clinical diagnostic settings where relatively small numbers of patient samples might require analysis for a validated genetic variant over a given time period, assay accuracy may be more important than cost per assay. On the other hand, both low assay cost and capability for high assay throughput (speed and volume of sample processing) will be important for early-stage novel drug target discovery studies that will require the analysis of large numbers of genome-wide or candidate-gene SNP markers in large numbers of clinical DNA specimens. The ideal genotyping technology is thus likely to be one with sufficient flexibility to adapt to a variety of different study designs or SNP throughputs—in terms of both the number of SNPs that require analysis per DNA sample and the number of patient DNA samples that require analysis. Such flexibility requires the ability to adapt the assay technology to different throughputs, which in turn may require a variety of detection methods and instrumentation platforms to cover the spectrum of cost, throughput, and convenience in a robust and optimally accurate fashion.

In this regard, it is useful to distinguish between the analytical biochemistries that form the basis of different SNP genotyping assays, and the variety of novel platforms and methods of detection, or readout, of the genotyping results. Although genotyping readout and genotyping platform (which includes devices for automation, miniaturization, and multiplexing) will strongly influence the throughput and cost components of a given assay technology, the fundamental nature of the underlying biochemistry primarily determines the potential accuracy and robustness of the method. With this in mind, the next section focuses on a discussion of some assay biochemistries, the particular features of which will determine the degree to which they may be ported to various detection, automation, and multiplexing platforms.

3. ASSAY BIOCHEMISTRIES

As mentioned above, the central core component of a genotyping technology is contained within the assay biochemistry. In this regard, a general distinction may be made between those methods that rely primarily upon differential hybridization stringency for their specificity, and those that derive specificity primarily from the ability to detect a product of an enzymatic reaction. A number of

methods also use combinations of hybridization and enzyme reactions, with varying degrees of contribution to specificity from each of the two components.

3.1. Hybridization-Based Approaches

The specificity of hybridization-based approaches in SNP genotyping relies on the fact that the melting temperatures of DNA–DNA or DNA–RNA hybrids that perfectly match are higher than those that do not. Thus experimental conditions may be found where differential rates or equilibrium levels of hybridization may be distinguished, even for single-base mismatches such as would occur in the presence of a SNP. Original hybridization-based approaches that employed the hybridization of radioactively labeled oligonucleotides on Southern blots of human genomic DNA suffered from all of the practical disadvantages of classical restriction fragment length polymorphism (RFLP) analysis (below), including use of large quantities of starting genomic DNA and the use of tedious and hazardous hybridization methods. The use of PCR reaction products and substitution of non-radioactive labels improved the assays, but the requirement for gel electrophoresis made high-throughput problematic. More recent methods that utilize the monitoring of the kinetics of probe hybridization to target DNA with changes in hybridization temperature, termed dynamic allele-specific hybridization (DASH) can now allow for greater specificity and higher throughput in a solely hybridization-based discrimination method (1,2).

3.2. Enzyme-Based Approaches

Because of the high inherent catalytic specificity of enzymes for their substrates, enzyme-based approaches to SNP genotyping generally possess a higher degree of assay fidelity than those primarily dependent on hybridization for their specificity. Historically, the standard method for genotyping of SNPs and of other types of genetic variants, RFLP analysis, is an enzymatic approach that relies upon the exquisite selectivity of bacterial restriction endonucleases for short stretches of defined DNA sequence that act as recognition sites for DNA strand cleavage. Thus in instances where a SNP changes a restriction enzyme recognition sequence, differential digestion of normal and variant sequences can be observed. First iterations of this approach required the digestion of large quantities of genomic DNA, electrophoretic separation of digested fragments, transfer of fragments to nitrocellulose or nylon membranes and detection of fragment size differences by hybridization with a radiolabeled probe complementary to the fragments. Advancements in the method, most notably prior PCR amplification of defined DNA segments containing SNP sites and PCR-mediated introduction of novel restriction sites, have improved the conservation of starting material, the ability to detect fragments visually in gels without the need for labeled probe hybridization and the applicability to a broader range of

SNP variant sites in the genome. However, the major drawback to such methods is still the requirement for electrophoretic separation of digested products, which severely limits the throughput and automatability of the method and increases reagent and labor costs.

Dideoxy DNA sequencing represents another example of an enzymatic approach to genotyping, which uses the specificity of DNA polymerase to incorporate appropriate nucleotide bases opposite a primed single-stranded DNA template, followed by size separation of terminated polymerase-extended reaction products by gel electrophoresis to detect the identity of nucleotide variants at defined sites. DNA sequencing remains the gold standard to which other genotyping methods are compared, and is also the method of choice for discovering new SNPs among DNA samples from populations or for confirming and localizing those discovered using scanning methods such as single-strand conformation polymorphism (SSCP) analysis or denaturing HPLC. With respect to the genotyping of pre-existing SNPs, major disadvantages of DNA sequencing include the ongoing requirement for electrophoretic separation of extended DNA fragments (which restricts the platform and hampers throughput), potential technical difficulties in detecting heterozygosity at a particular SNP locus when genotyping from uncloned PCR products, and the higher cost of reagents associated with the requirement to extend reactions by many more nucleotide bases than the polymorphic site under investigation.

An alternative method of DNA sequencing, called Pyrosequencing (3) uses a series of enzymatic reactions to enable the continuous readout of short stretches of DNA sequence without the requirement for electrophoretic fragment separation. The method may be useful for low-throughput research and clinical diagnostic applications, but is unlikely to be applied to high-throughput applications.

Another variation of DNA sequencing is called single-base primer extension, or minisequencing (4) (see also Chapter 20). Primer extension in its original form involves the annealing of an oligonucleotide primer to a single-stranded PCR product (analogous to the primer annealing step in DNA sequencing reactions) at a location which lies immediately adjacent to, but not including, the polymorphic SNP site, followed by the addition of DNA polymerase and subsequent enzymatic extension of the primer in the presence of only chain-terminating dideoxynucleotides, which may be labeled in a variety of ways to facilitate subsequent detection of the identity of the single incorporated nucleotide (see below). As with DNA sequencing, an important distinguishing feature of primer extension is that specificity arises not from primer hybridization but rather from the catalytic activity of the DNA polymerase. An additional advantage of primer extension over classical sequencing is that electrophoretic separation is not required to detect the extended product, but rather any one of a number of simple methods to detect the label attached to the incorporated dideoxynucleotide.

Finally, the simplicity of the biochemistry allows reactions to be performed either in solution or on primers attached to solid supports, providing flexibility with regard to platforms and capacity for reaction multiplexing (see below).

3.3. Combined Hybridization/Enzymatic Approaches

One of the most commonly used small-scale genotyping methods is allele-specific PCR amplification, which combines allele-selective PCR primer hybridization with a subsequent PCR reaction (5). Conditions are optimized so that hybridization and subsequent amplification occur only when the PCR priming oligonucleotide is perfectly matched with the target site (usually with the polymorphic site at the 3′ end of the oligonucleotide). The result of the test is therefore determined electrophoretically as either the presence or absence of a PCR product. Again, a major drawback of this method for high-throughput applications is the requirement for gel electrophoresis and visualization of product, a process that is not amenable to automation. It is also important to note that in this assay accuracy is still dependent solely on differential hybridization specificity, which must be carefully optimized for each polymorphic site to be analyzed and is therefore difficult to multiplex.

A number of newer assays have been designed to introduce an enzymatic step in order to improve the specificity, and therefore the overall accuracy, of hybridization-based assays. Oligonucleotide ligation assays (OLA) (6) are an example of this approach, where addition of a DNA ligase to a hybridization reaction results in the attachment of an oligonucleotide to an immobilized capture DNA only when the fragments are perfectly matched. The Invader assay (7) utilizes a sequence specific cleavage enzyme called a cleavase to release a quenched fluorescent dye label when hybridization of a perfectly complementary oligonucleotide produces a unique structure that is recognized by the enzyme. Although a potential significant advantage of this assay is the elimination of the requirement for prior PCR amplification of target sequence, the method currently requires the use of a relatively large quantity of input genomic DNA, which can often be a scarce and valuable commodity. The TaqMan assay (8) uses the exonuclease activity of Taq DNA polymerase to liberate a fluorescent signal from a quenched probe oligonucleotide when the sequence is perfectly complementary to a target, but not when the sequence differs.

4. DETECTION METHODS

As mentioned above, most classical genotyping methods have typically relied upon the electrophoretic separation of DNA fragments in order to visually detect differences in size or fragmentation patterns between allelic variants. Although the advent of capillary electrophoresis in DNA sequencing has enabled the throughput of these detection methods to be improved considerably, such systems require a significant capital outlay and there are still limitations to the scale of genotyping studies that can be performed. For

this reason, a number of detection methods that do not require electrophoretic fragment separation have been developed.

Mass spectrometry may be used to simultaneously identify oligonucleotide fragments that differ by as little as a single nucleotide (9). This method has been applied to a variation of the primer extension genotyping assay, whereby fragments differing in size by one or a few nucleotides are generated in primer extension reactions, and detected by their mass differences. Other detection methods used in conjunction with the primer extension biochemistry include the use of an enzyme-linked immunochemical assay that produces different color reactions corresponding to the allelic variants in micotiter plate wells, and fluorescently labeled dideoxynucleotides that can be detected on microarrays of immobilized extended oligonucleotides (10). Fluorescence polarization (FP) (11) has been used as an effective detection marker for primer extension, Invader and TaqMan assays, and is based on the observation that the degree of emitted FP of a labeled molecule is proportional to its molecular mass. Thus by monitoring the FP of a fluorescent dye during an enzyme reaction, changes in the molecular mass can be detected without sample purification or electrophoresis.

Other approaches using fluorescence to monitor the course of enzyme reactions in SNP genotyping make use of the ability of an adjacent molecule to quench the fluorescence of a fluorophore. Reactions that result in the removal of the quencher or its movement away from the fluorophore will result in termination of the quenching and emission of a fluorescent signal. A variation of this theme is the recent development of molecular beacons (12), which are single oligonucleotide-based molecules containing both a fluorophore and a quencher, such that hybridization of the molecule produces a conformational change that releases the quenching and produces a fluorescent signal.

5. PLATFORMS

In order to produce integrated genotyping systems that meet the practical needs of a broad range of experimental uses, a variety of system embodiments have been developed. These are designed so as to optimize assay convenience, reliability, sample usage, cost-effectiveness, and throughput, as required for the application. For many future genomic applications, the requirement for cost-effective, large-scale genotyping of large numbers of SNPs will be required. Three of the most popular means by which to achieve cost-effective high-throughput genotyping include miniaturization, process automation, and reaction multiplexing. Automation, which involves robotic sample handling and reagent delivery to assays, will not be discussed further here. Miniaturization and multiplexing are of considerable value in dramatically reducing reagent usage, assay time, and overall genotyping cost. One of the most effective ways of achieving this is through the use of a variety of novel microarray technologies. Microarrays may in fact be

considered as miniaturized multiplexing devices, enabling the simultaneous reaction and detection of multiple samples.

Classical microarrays are two-dimensional immobilized spotted arrays of DNA or oligonucleotides on glass slides, with the locations of specific sequences being pre-defined. Traditional expression microarrays use immobilized DNA to hybridize with cDNA, which has been synthesized from isolated tissue mRNA, in order to obtain semi-quantitative measures of transcript levels in the tissues. Microarrays may also be used very effectively for primer extension based SNP genotyping, whereby oligonucleotide primers specific for the interrogation of particular SNPs are immobilized in defined locations on arrays, incubated with pooled target PCR products and extended with fluorescently labeled dideoxynucleotides. The presence and identity of the extended fluorescent labels can then be detected by automated fluorescence imaging of the microarrays (10). Alternatively, a variation on this approach is the use of the so-called universal arrays (13), where the primer extension oligonucleotide is synthesized with a further generic ''tag'' sequence attached, that is complementary to a sequence immobilized on a microarray. In this embodiment, multiplexed primer extension reactions can be performed in solution, followed by incubation of the extended products with the microarray in order to capture the extended products for detection on the microarray.

Another microarraying technique involves the use of labeled microspheres as the attachment matrix instead of glass slides. In this case, the array is a three-dimensional solution array rather than a two-dimensional glass array, and the identity and fluorescence of the attached products can be measured by flow cytometry of the microspheres (14). A recent innovation for the labeling of microspheres involves the use of fluorescent semiconductor nanocrystals, or quantum dots, which create a spectral bar code for microsphere identification (15). The method allows for a high level of genotyping reaction multiplexing, due to the ability to create a large number of unique identifying signatures.

6. CONCLUDING REMARKS

From the above discussion, it is clear that there are a large number of methods now available for the genotyping of SNPs in pharmacogenomic studies. No single combination of biochemistry, detection method, and platform will be able to meet the needs of every genotyping ''customer.'' Thus the potential user of genotyping assays must therefore carefully assess the needs of one's own genotyping laboratory, and determine the relative importance of throughput, cost-effectiveness, speed of analysis, accuracy, and availability of existing and projected equipment and infrastructure capabilities in making choices of one or more appropriate embodiments to meet those needs.

REFERENCES

1. Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, Blennow K, Brookes AJ. Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. Genome Res 2001; 11:152–162.
2. Jobs M, Howell WM, Stromqvist L, Mayr T, Brookes AJ. DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays. Genome Res 2003; 13:916–924.
3. Fakhrai-Rad H, Pourmand N, Ronaghi M. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. Hum Mutat 2002; 19:479–485.
4. Suomalainen A, Syvanen AC. Analysis of nucleotide sequence variations by solid-phase minisequencing. Methods Mol Biol 2003; 226:361–366.
5. See D, Kanazin V, Talbert H, Blake T. Electrophoretic detection of single-nucleotide polymorphisms. Biotechniques 2000; 28:710–714, 716.
6. Iannone MA, Taylor JD, Chen J, Li MS, Rivers P, Slentz-Kesler KA, Weiner MP. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. Cytometry 2000; 39:131–140.
7. Olivier M, Chuang LM, Chang MS, Chen YT, Pei D, Ranade K, de Witte A, Allen J, Tran N, Curb D, Pratt R, Neefs H, de Arruda Indig M, Law S, Neri B, Wang L, Cox DR. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. Nucleic Acids Res 2002; 30:e53.
8. Latif S, Bauer-Sardina I, Ranade K, Livak KJ, Kwok PY. Fluorescence polarization in homogeneous nucleic acid analysis II: 5′-nuclease assay. Genome Res 2001; 11:436–440.
9. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. Hum Mutat 2001; 17:296–304.
10. Liljedahl U, Karlsson J, Melhus H, Kurland L, Lindersson M, Kahan T, Nystrom F, Lind L, Syvanen AC. A microarray minisequencing system for pharmacogenetic profiling of antihypertensive drug response. Pharmacogenetics 2003; 13:7–17.
11. Kwok PY. SNP genotyping with fluorescence polarization detection. Hum Mutat 2002; 19:315–323.
12. Mhlanga MM, Malmberg L. Using molecular beacons to detect single-nucleotide polymorphisms with real-time PCR. Methods 2001; 25:463–471.
13. Allen M, Divne AM. Universal tag arrays in forensic SNP analysis. Methods Mol Biol 2004; 297:141–154.
14. Chen J, Iannone MA, Li MS, Taylor JD, Rivers P, Nelsen AJ, Slentz-Kesler KA, Roses A, Weiner MP. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. Genome Res 2000; 10:549–557.
15. Xu H, Sha MY, Wong EY, Uphoff J, Xu Y, Treadway JA, Truong A, O'Brien E, Asquith S, Stubbins M, Spurr NK, Lai EH, Mahoney W. Multiplexed SNP genotyping using the Qbead system: a quantum dot-encoded microsphere-based assay. Nucleic Acids Res 2003; 31:e43.

13

Molecular Diagnostics

Tracy L. Stockley and Peter N. Ray

Division of Molecular Genetics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

1. INTRODUCTION

Pharmacogenomics is a rapidly evolving area driven by the new genetic information and new molecular technologies arising from the mapping and sequencing of the human genome. The field of pharmacogenomics has a parallel with the more traditional area of molecular diagnostics, in which molecular techniques are used for identification of a disease mutation in a particular patient population with a genetic disorder. Both fields have been impacted greatly by new genetic information and technologies, and have certain similar challenges to overcome. This chapter will focus on describing technologies and general issues in molecular diagnostics, and similarities to pharmacogenomic molecular applications.

Molecular diagnostics is an integration of molecular genetic knowledge and technology and conventional laboratory medicine for patient diagnostics. Due to the cost of performing testing, molecular diagnosis has historically been restricted to testing of a limited number of individuals at high risk, typically people who are suspected of being affected or carriers of a particular monogenic inherited disorder. This scenario is beginning to change as diagnostics expands into higher-volume testing of larger patient groups and at-risk populations (1–3). For example, widespread population carrier screening for cystic fibrosis carriers using molecular diagnostic methods has recently been implemented in the United States, with the goal of

identifying carrier couples prior to the birth of an affected child (4,5). The increased volume of testing produced by these new applications has required the development of higher-throughput and lower cost molecular diagnostic assays to detect nucleotide changes. Large-scale molecular diagnostic applications parallel molecular pharmacogenomic testing requirements for high throughput and cost-effective molecular diagnostic tests suitable for screening a large number of individuals for multiple genetic changes.

An important issue in a discussion of typical molecular diagnostics and pharmacogenomics is the difference in the interpretation of testing for mutations in a single gene disorder vs. testing for mutations in one or more genes involved in multifactorial disorders. The majority of current molecular diagnostic assays are used to test for mutations in monogenic disorders. For example, cystic fibrosis is a relatively common single gene disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene (6). Many diagnostic laboratories routinely test for 25–30 of the most common CFTR mutations, which account for the majority of cystic fibrosis mutations in most populations (7). The finding of a mutation is diagnostic in that a prediction of disease can be made on the basis of finding two mutations in the CFTR gene carried by an individual.

In contrast, multifactorial diseases such as hypertension and obesity are more complex as disease onset is likely to be dependent on changes in several genetic regions, each with a different influence on disease progression, plus environmental factors (8). Individuals with a family history of hypertension may carry mutations in several genetic regions that lead to an increased susceptibility for hypertension. However, due to environmental differences only a portion of the individuals who carry these mutations will develop hypertension. The interpretation of molecular testing in multifactorial diseases is complex, as it requires interpretation of the results of testing several genetic regions and involves gene–environment interactions that are difficult to predict. Many individual drug responses are likely to be multifactorial traits, and these same issues will complicate prediction of drug response outcome based on molecular testing results.

2. TYPES OF GENETIC VARIATIONS

Genetic diseases can be divided into two main groups based on whether they are caused by relatively few common mutations or by many unique mutations. Tay-Sachs disease, a fatal neurodegenerative disorder caused by a deficiency of hexosaminidase A, is common in the Ashkenazi Jewish population. Three hexosaminidase mutations account for ~96% of the disease in the Ashkenazi Jewish population (9). In contrast most patients with Fabry disease, a metabolic disorder caused by deficiency of the enzyme galactosidase A, have unique mutations in the galactosidase A gene (10). This

distinction is significant, as the detection of a small number of known mutations in a gene requires quite different technologies than the analysis of a complete gene to search for unknown mutations.

Other types of genetic variation known to cause genetic disorders include deletions and duplications that can range in size from a single base to large regions encompassing whole exons or entire genes. Deletions and duplications are of particular significance in molecular diagnosis, as they require special assays for detection, such as quantitative PCR to determine gene copy number (11). Genetic mutations can also alter function at levels other than the DNA coding sequence, such as mutations at conserved splice site sequences, which alter RNA splicing.

The methods of detection of these different types of genetic changes vary depending on the type of mutation to be detected. Methods of mutation detection include specific assays designed to identify a certain nucleotide change at a particular base pair in a sequence, or the use of direct sequencing to detect many unique mutations. Direct sequencing may be preceded by the scanning methods to highlight particular exons of a gene that may contain a mutation. If a genetic disease is caused by a few recurrent mutations that account for the majority of disease, it is often more appropriate to test for these few mutations by a specific assay rather than by sequencing. In contrast, diseases caused by a large number of mutations in a single gene are more appropriately tested by direct sequencing methods (for review, see Ref. 12). Therefore, it is critical that a molecular diagnostic test be appropriately matched to the types of mutations to be detected for a particular gene.

3. METHODS TO DETECT KNOWN MUTATIONS

There is a vast array of techniques that are currently being used to look for mutations for clinical molecular diagnostic use. The following overview of methods describes some of the most common techniques to detect recurrent mutations used in many diagnostic laboratories. As detection of unique mutations is not of relevance to pharmacogenetic applications, this topic will not be discussed in this chapter.

3.1. PCR-Restriction Enzyme Assays

Certain genetic changes alter a restriction enzyme recognition site, either by creating a new site or destroying an existing site. In these cases, a molecular diagnostic assay can be designed in which a region containing the potential mutation site is amplified via PCR and digested with the appropriate restriction enzyme to determine if the restriction digest pattern is altered due to the presence of the mutation. In some cases, designing a PCR primer containing a mismatch that anneals near a mutation site can artificially produce a

restriction enzyme site. When combined with the mutant genetic sequence, the mismatched PCR primer sequence and the mutation alters a restriction enzyme site and so produces altered restriction enzyme patterns. Quality control for this type of assay requires that appropriate positive and negative control samples must be present in each assay in order to confirm that the enzyme is properly active.

An example of a PCR-restriction enzyme assay to detect variation for a pharmacogenomic application is the original detection of mutations in the gene for the cytochrome isozyme P250D6 (*CYP2D6*), which cause the "poor metabolizer" of debrisoquine phenotype (13). Since this paper was published, many assays using different methods have been designed to detect mutant alleles in the P250D6 gene, highlighting the complexity in choosing a method for molecular analysis of mutations in a particular gene.

The main advantage for molecular diagnostics using a PCR-restriction enzyme assay is that it is simple to develop and perform, likely accounting for its continued use in many diagnostic laboratories. The major disadvantage is that the assay is time consuming, as each mutation must be individually analyzed. Although this assay is effective for testing a small number of samples in specific cases, it is not suited for testing samples for multiple mutations or for screening large numbers of samples.

3.2. Allele-Specific Oligonucleotide Assay

The basis of the allele-specific oligonucleotide (ASO) assay is that DNA duplexes which contain a mismatch are destabilized and have a lower melting temperature than correctly paired duplexes. To test for mutations using ASO, two probes, one containing the normal sequence and one containing the mutant sequence, are produced and hybridized to the patient's DNA. For each normal and mutant probe, conditions can be found where the probe will hybridize to only its perfectly matched duplex. If the patient sample contains only normal sequence, only the normal probe will hybridize. In a heterozygous sample, both the mutant and normal probes will hybridize, and in a homozygous mutant sample only the mutant probe will hybridize.

An advantage of the ASO method is that it can be used to simultaneously test samples for several different mutations by the use of multiple probes bound to a solid matrix. In practice, the success of this method relies on precisely establishing conditions for optimal oligonucleotide hybridization in order to ensure specific probe hybridization, and so multiplex ASO assays can be difficult to develop. Molecular diagnostic kits for use in genetic disorders based on ASO methods are available (14).

There have been improvements to the ASO assay, specifically by development of a multiplex allele-specific diagnostic assay (MASDA) in which the ASO technique is adapted to a solid support and multiple regions are probed simultaneously. This has been achieved by altering probe

hybridization conditions, so that hybridization of multiple probes at a single temperature is feasible. Using these improvements, it has been possible to analyze >500 samples simultaneously for >100 known mutations in multiple genes (15).

3.3. Allele-Specific Amplification Assay

The Allele-Specific Amplification Assay (ASA) assay is based on the fact that Taq polymerase will not initiate amplification from a primer that has a mismatch at the 3′ ends. Two primers are designed so that the 3′ base of the primer corresponds to the site of the genetic mutation to be tested, with either the normal or the mutant sequence at the 3′ base positions. An unknown sample can then be tested for the presence of the mutation by using both the normal and the mutant primers in PCR with a common reverse primer. If the sample contains only normal sequence, a PCR product will only be produced when the normal primer is used, and similarly when the sample contains mutant sequence a product will only result from use of the mutant primer. Like the PCR-restriction enzyme method discussed, the ASA approach has also been applied to the detection of mutations in the *CYP2D6* gene (16).

In the original ASA protocols, the mutant and normal PCR primers were separated into two reactions, so that lack of amplification could occur in one PCR reaction depending on the sequence present in a test sample. This is not ideal for a diagnostic test due to the possible misinterpretation of a false negative result, and the ASA protocol is usually modified to be a multiplex reaction that includes a positive internal control in each PCR reaction. For example, an ASA assay has been developed, which detects 12 common CFTR mutations simultaneously (17). However in this assay, two reactions must still be run in parallel for every sample to be analyzed, since the mutant and normal products produced are the same size and so must be physically separated in order to be distinguished.

The ASA can be improved by the use of fluorescent-dye labeled primers, which avoids the need for two separate reactions by using flourochromes to distinguish normal and mutant sequences. We have developed molecular diagnostic ASA assays to detect mutations causing Tay-Sachs and Canavan disease using fluorescent-dye labeled PCR primers (18,19). The mutant and normal primers are labeled with different color dyes, so that the PCR products resulting from either the normal or the mutant allele-specific primer will be a different dye color, allowing discrimination of normal and mutant sequence. The use of fluorescent dyes thus simplifies the assay, and allows one sample to be tested for multiple mutations in a single reaction.

The advantage of the ASA method is that multiplex reactions to detect several mutations simultaneously can be developed. Multiplex reactions

reduce the labor and costs and so are ideal for detection of a larger number of mutations. The main disadvantage of the ASA method is that achieving specific product amplification can be problematic.

3.4. Oligonucleotide Ligation Assay

The Oligonucleotide Ligation Assay (OLA) is similar to allele-specific amplification in that specific interrogation of a mutation site is achieved by two oligonucleotides that contain the normal or mutant base at the 3′ end of the primer. However, in the OLA assay, the normal or mutant primer anneals directly downstream and adjacent to a common primer. The two primers are directly adjacent to one another, and thermostable ligase is able to join the annealed primers. In the case of a normal DNA sequence, only the normal and common primers will anneal and so be ligated, while a mutant DNA sequence will produce ligation of only the mutant and common primers. This method has also been applied to the detection of *CYP2D6* alleles (20).

A recent improvement in the OLA assay is the use of sequence-coded separation (SCS), in which non-nucleic mobility altering compounds are attached to the specific primers. The mobility altering compounds are designed so that the products from each primer are a different size, and so will allow discrimination between primer pairs for different mutations. This technology is used in a diagnostic assay for cystic fibrosis in which 32 mutations in the CFTR gene can be detected simultaneously (21). This novel method of size separation may expand the utility of the OLA assay, as it will allow multiplex assays for a large number of mutations. The main advantage of the OLA technique is the ability to multiplex several mutations; while the disadvantage is that without SCS the assay is of somewhat limited application. Many diagnostic laboratories currently use the commercially available OLA assay for mutations causing cystic fibrosis.

3.5. Primer Extension/Minisequencing Assay

The primer extension assay is similar to the dideoxy method commonly used in sequence analysis. In the primer extension assay, a region containing the mutation to be assayed is amplified in a first PCR reaction. A specific primer, which is designed to anneal directly upstream of the base which is the site of a known mutation, is then used in a second reaction. Radioactively labeled dideoxy nucleotides corresponding to either the normal or the mutant base at the potential mutation site are added in separate tubes. During the reaction, the labeled nucleotide added to the 3′ end of the primer will depend on the sequence at the potential mutation site. If the normal sequence is present at the potential mutation site, only the reaction containing the normal labeled dideoxy nucleotide will produce a labeled primer. Conversely, if the mutant sequence is present at the mutation site, only

the reaction containing the mutant nucleotide will produce a labeled primer. Individuals who are heterozygous for the mutation will produce labeled primers in both dideoxy tubes, due to the presence of both sequences at the potential mutation site.

Primer extension assays are very sensitive for mutation detection and may be advantageous for large-scale testing with some modifications to the basic protocol. Primer extension assays have been designed to use fluorescent dye labeled nucleotides to eliminate the need for radioactivity and may also be adaptable to use on solid supports (22). For molecular diagnostics purposes, there are also commercially available kits and protocols for diagnostic applications (23).

4. TECHNICAL ADVANCES IN MOLECULAR DIAGNOSTIC TECHNIQUES

In recent years, there have been major technical advances in genetic analysis methods due primarily to use of new technologies developed for genomic applications. This is leading to an increased emphasis on faster, more efficient methods of detecting genetic mutations facilitated by the use of novel methods and equipment, with increased automation to reduce labor intensive steps in genetic analysis.

Major technological advances will be essential as routine molecular diagnosis of genetic disorders adjusts to increasing volumes and large-scale population screening. Most current molecular diagnostics assays are not highly automated, and are generally labor-intensive and expensive. For large-scale molecular diagnostic testing, more automation with reduced personnel involvement is essential in order to reduce the cost of testing. Other requirements of large-scale testing will be improved software programs capable of dealing with large amount of data.

4.1. Capillary Electrophoresis

A significant advance in common molecular diagnostic applications is capillary electrophoresis, in which traditional slab electrophoresis gels are replaced by capillaries. The electrophoresis of samples is carried out in a thin capillary tube filled with a matrix, and the movement of DNA molecules through the tube is detected and recorded. Capillary electrophoresis is amenable to automation as samples can be automatically loaded from reaction plates. Capillary electrophoresis can be readily applied to many genetic tests that traditionally would be analyzed on a slab gel, including methods for recurrent mutation detection, fragment analysis and direct sequencing. Capillary electrophoresis with equipment containing 96–384 capillaries has greatly reduced the labor required for certain molecular diagnostics tests, particularly direct sequencing for mutation detection (24).

4.2. DNA Chip Technology

The DNA chips are high-density arrays, in which many nucleic acid sequences are anchored on glass supports similar in size to microscope slides. The DNA chips can be designed in certain formats depending on the application of the chip. The current applications of DNA chip technology of relevance to diagnostics include use in determining sequence of an unknown fragment by hybridization, detection of SNPs or gene expression profiling in various tissues (25). The major benefit of the use of DNA chips for these applications is that information on thousands of genetic regions can be obtained from a single chip experiment. For example, chips that are used to determine expression profiles from various genes can analyze thousands of RNA fragments on a single chip. This has been used for cancer applications, in which gene expression profiles from tumor tissues are compared to normal tissues to determine, which genes are differentially regulated in the cancer tissues (26). Currently, DNA chips for use in gene expression studies work well. However, sequencing and SNP chips are still not accurate enough for routine use in identifying genetic changes in a clinical molecular diagnostic laboratory. In addition, there are ethical considerations regarding provision of molecular diagnostic testing of a large number of genes or mutations on DNA chip, which need to be addressed (27).

4.3. Denaturing High Pressure Liquid Chromatography Analysis

New applications for Denaturing High Pressure Liquid Chromatography (HPLC) may also have important future applications for testing for genetic variation. The HPLC has been adapted to DNA fragment analysis for separation of fragments under partially denaturing conditions. Fragments to be analyzed for the presence of a mutation are amplified by PCR and then run on a denaturing HPLC column. The DNA heteroduplexes containing a mismatched base due to the presence of a mutation have a different mobility through the HPLC column from normal matched homoduplexes, due to the altered melting temperature of the heteroduplex. The mobility difference between the normal and the mutant sample allows for screening of fragments for genetic changes (28).

The main application of DHPLC technology for diagnostics is likely to be its use as a screening tool for genetic variation prior to sequencing, which is similar in principle to traditional methods of scanning such as single-strand conformational polymorphism (SSCP) and denaturing gradient gel electrophoresis (DGGE) assays. However, DHPLC also has the potential to be a method to screen samples for previously known mutations (29). The advantages of DHPLC are its automated nature, the simplicity of preparing fragments for analysis and the short run times required for fragment analysis.

4.4. MALDI-TOF Mass Spectrometry

Mass spectrometry is also being applied to genetic analysis by the use of matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. The principle of this application is that sequence differences can be determined by analyzing the inherent mass differences of the four-nucleotide bases. For molecular diagnostic applications, the most obvious application of MALDI-TOF is the use of modified primer-extension assays, in which the base extended is at the nucleotide site of a known sequence change (30,31). MALDI-TOF is promising for high throughput applications, as more than one genetic region or mutation can be simultaneously analyzed and the assays. An example of the use of MALDI-TOF in diagnostics is in detection of mutations in CFTR causing cystic fibrosis (32).

4.5. Real-Time PCR

Additional developments in PCR technologies are also impacting on molecular diagnostics. One of the most important has been the development of PCR machines that have the capability of detecting product formation during the PCR reaction, known as real-time PCR. Assays can be designed in which the binding of a sequence-specific probe to its homologous PCR product results in an increase in fluorescence during the PCR reaction, allowing for real-time detection of the PCR product. This has been achieved using various probe designs that maintain a fluorescent reporter dye in close proximity to a quencher dye. Upon hybridization to its specific sequence the quencher is separated from the reporter thus generating a fluorescent signal from the reporter dye (33,34). This technology has been applied to detection of mutations causing genetic disease as well as for pharmacogenomic research (35,36). The advantage of real-time PCR techniques is that they do not require any post-PCR analysis such as gel electrophoresis, since the amplification and detection of the specific product are completed within the PCR reaction.

4.6. Robotics

A common objective of all of the new technologies discussed is to reduce the labor and expense required for large-scale testing. High-throughput detection machines require high sample input rates, which are not feasible without the use of robotics. Robotics will be required to automate the isolation of DNA, to prepare reactions for PCR, and to load detection machines after PCR (37). For example, the labor associated with manually preparing DNA samples for analysis on a 384-sample capillary electrophoresis machine would negate the benefit of using the advanced equipment. Robotics instruments are now available that can perform various functions, including

nucleic acid extraction and preparation of PCR reactions. Robotics is likely to become more flexible in future, with robots specifically designed to interact with specialized equipment for specific applications. An important aspect of increased robotic use will be the development of software programs able to perform data analysis on the high volume of data generated. The increasing use of robotics will allow genetic diagnostic laboratories to increase the number of tests it is able to perform, without continually increasing laboratory staff, due to the ability to perform more tasks in less time.

The methods listed above are some of the major areas of interest in the development of biotechnology-based molecular diagnostic procedures. Some private companies are currently offering genetic testing based on new biotechnology-based large-scale molecular diagnostic methods. An overview of some companies involved in developing diagnostic strategies for large-scale genetic testing for pharmacogenomics is provided in Persidis (38).

5. CONCLUSIONS

The need to develop high-throughput, sensitive and cost-effective molecular testing for genetic variation remains a significant challenge in molecular diagnostics, which is shared with pharmacogenomic applications. Since many individuals will require testing, and many genetic changes may need to be tested, the diagnostic methods will need to be robust, cost-effective and specific. Some of the new advances in genetic analysis, such as real-time PCR and MALDI-TOF, hold promise for meeting these demands. In addition, as the throughput of molecular diagnostic testing improves, testing is likely to still be limited by uncertain significance of new sequence changes detected. This is also an issue in common with pharmacogenetics, as the application of routine pharmacogenomics still requires a more thorough understanding of the genetic causes underlying drug response variations. The genetic variations may be simple changes, or they may be complex alterations that are also influenced by environmental factors. Significant research on the effect of genetic variation on enzyme activity and the resultant effect on drug response will be required. New technologies for DNA analysis and mutation detection, and improved understanding of the interpretation of mutations, will make it possible to meet the demands of both future molecular diagnostic and pharmacogenomic applications.

REFERENCES

1. Grody WW. Molecular genetic risk screening. Annu Rev Med 2003; 54: 473–490.
2. Khoury MJ, McCabe LL, McCabe ERB. Population screening in the age of genomic medicine. N Engl J Med 2003; 348:50–58.

3. ACOG Committee on Genetics. Prenatal and preconceptional carrier screening for genetic diseases in individuals of Eastern European Jewish descent. Obstet Gynecol 2004; 104:425–428.
4. NIH Consensus Development Conference Statement. Genetic testing for cystic fibrosis. Arch Intern Med 159:1529–1539.
5. Watson MS, Cutting GR, Desnick RJ, Driscoll DA, Klinger K, Mennuti M, Palomaki GE, Popovich BW, Pratt VM, Rohlfs EM, Strom CM, Richards CS, Witt DR, Grody WW. Cystic fibrosis population screening: 2004 revision of American College of Medical Genetics mutation panel. Genet Med 2004; 6:387–391.
6. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989; 245: 1066–1073.
7. Richards CS, Bradley LA, Amos J, Allitto B, Grody WW, Maddalena A, McGinnis MJ, Prior TW, Popovich BW, Watson MS, Palomaki GE. Standards and guidelines for CFTR mutation testing. Genet Med 2002; 4:379–391.
8. Bell J. Medical implications of understanding complex disease traits. Curr Opin Biotech 1998; 9:573–577.
9. Paw BH, Tieu PT, Kaback MM, Lim J, Neufeld EF. Frequency of three Hex A mutant alleles among Jewish and non-Jewish carriers identified in a Tay-Sachs screening program. Am J Hum Genet 1990; 47:698–705.
10. Blaydon D, Hill J, Winchester B. Fabry disease: 20 novel GLA mutations in 35 families. Hum Mutat 2001; 18:459.
11. Armour JA, Barton DE, Cockburn DJ, Taylor GR. The detection of large deletions or duplications in genomics DNA. Hum Mutat 2002; 20:325–337.
12. Cotton RGH. Mutation Detection. New York/Oxford: Oxford University Press, 1997.
13. Douglas AM, Atchison BA, Somogyi AA, Drunmer OH. Interpretation of a simple PCR analysis of the CYP2D6(A) and CYP2D6(B) null alleles associated with the debrisoquine/sparteine genetic polymorphism. Pharmacogen 1994; 4:154–158.
14. Innogenetics assay kits, http://www.innogenetics.be/site/diagnostics.html.
15. Shuber AP, Michalowsky LA, Nass GS, Skoletsky J, Hire LM, Kotsopoulos SK, Phipps MF, Barberio DM, Klinger KW. High throughput parallel analysis of hundreds of patient samples for more than 100 mutations in multiple disease genes. Hum Mol Genet 1997; 6:337–347.
16. Heim M, Meyer U. Genotyping of poor metabolisers of debrisoquine by allele-specific PCR. Lancet 1990; 336:529–532.
17. Robertson NH, Weston SL, Kelly SJ, Duxbury NJ, Pearce SR, Elsmore P, Webb MBT, Newton CR, Little S. Development and validation of a screening test for 12 common mutations of the cystic fibrosis CFTR gene. Eur Respir J 1998; 12:477–482.
18. Feigenbaum A, Moore R, Clarke J, Hewson S, Chitayat D, Ray PN, Stockley TL. Canavan disease: carrier frequency determination in the Ashkenazi Jewish population and development of a novel molecular diagnostic assay. Am J Med Genet 2004; 124A:142–147.

19. Stockley TL, Ray PN. Multiplexed fluorescence analysis for mutations causing Tay-Sachs disease. In: Potter NT, ed. Methods in Molecular Biology, 217: Neurogenetics: Methods and Protocols. New Jersey: Humana Press, 2001.
20. Hansen TS, Petersen NE, Iitia A, Blaabjerg O, Peterson PH. Robust non-radioactive oligonucleotide ligation assay to detect a common point mutation in the CYP2D6 gene causing abnormal drug metabolism. Clin Chem 1995; 41:413–418.
21. Brinson EC, Adriano T, Bloch W, Brown CL, Chang CC, Chen J, Eggerding FA, Grossman PD, Iovannisci DA, Madonik AM, Sherman DG, Tam RW, Winn-Deen ES, Woo SL, Fung S. Introduction to PCR/OLA/SCS, a multiplex DNA test, and its application to cystic fibrosis. Gene Test 1997; 1:61–68.
22. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. Mini-sequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Gen Res 1997; 7:606–614.
23. Applied Biosystems, ABI Prism SNapShot Multiplex kit, http://www.appliedbiosystems.com.
24. Righetti PG, Gelfi C, D'Acunto MR. Recent progress in DNA analysis by capillary electorphoresis. Electrophoresis 2002; 23:1361–1374.
25. Geschwind DH. DNA microarrays: translation of the genome from laboratory to clinic. Lancet Neurol 2003; 2:275–282.
26. Liang P, Pardee AB. Analysing differential gene expression in cancer. Nat Rev Cancer 2003; 3:869–876.
27. Grody WW. Ethical issues raised by genetic testing with oligonucletide microarrays. Mol Biotechnol 2003; 23:127–138.
28. Kuklin A, Munson K, Gjerde D, Haefele R, Taylor P. Detection of single-nucleotide polymorphisms with the WAVE DNA fragment analysis system. Genet Test 1997; 1:201–206.
29. Cobb CJ, Scott G, Swingle RJ, Wilson S, Ellis J, MacEwen CG, McLean WHI. Rapid mutation detection by the Transgemoic wave analyzer DHPLC identifies MYOC mutations in patients with ocular hypertension and/or open angle glaucoma. Br J Ophthalmol 2002; 86:191–195.
30. Roskey MT, Juhasz P, Smirnov IP, Takach EJ, Martin SA, Haff LA. DNA sequencing by delayed extraction-matrix-assisted laser desorption/ionization time of flight mass spectrometry. Proc Natl Acad Sci USA 1996; 93:4724–4729.
31. Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res 1997; 7:378–388.
32. Buyse IM, McCarthy SE, Lurix P, Pace RP, Vo D, Bartlett GA, Schmitt ES, Ward PA, Oermann C, Eng CM, Roa BR. Use of MALDI-TOF mass spectrometry in a 51-mutation test for cystic fibrosis: evidence that 3199del6 is a disease-causing mutation. Genet Med 6:426–430.
33. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. Gen Res 1996; 6:986–994.
34. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. Nat Biotech 1996; 14:303–308.
35. Cusco I, Barcelo MJ, Biaget M, Tizzano EF. Implementation of SMA carrier testing in genetic laboratories: comparison of two methods for quantifying the SMN1 gene. Hum Mutat 2002; 20:452–459.

36. Endrizzi K, Fischer J, Klein K, Schwab M, Nussler A, Neuhaus P, Eichelbaum M, Zanger UM. Discriminative quantification of cytochrom P4502D6 and 2D7/8 psuedogene expression by TaqMan real-time reverse transcriptase polymerase chain reaction. Anal Biochem 2002; 15:121–131.
37. Wilkinson DS. The role of technology in the clinical laboratory of the future. Clin Lab Manage Rev 1997; 11:322–330.
38. Persidis A. Pharmacogenomics and diagnostics. Nat Biotech 1998; 16:791–792.

14

Metabonomics

Lois D. Lehman-McKeeman
Discovery Toxicology, Bristol Myers Squibb Co., Princeton, New Jersey, U.S.A.

Metabonomics is a branch of ''omics'' technologies focused on the analysis and measurement of endogenous metabolites. The workhorse of metabonomic applications has been nuclear magnetic resonance (NMR) spectroscopy, but applications and approaches incorporating separation techniques such as high performance liquid chromatography (HPLC), gas chromatography (GC), and mass spectrometry (MS) are being used with increasing frequency. Metabonomic approaches have been applied broadly in botanical sciences and biomedical studies, including both diagnostic medicine and basic research. To date, some of the most significant efforts and advances in metabonomics have been made relative to plant science (for reviews see Refs. 1 and 2; Feihn et al. 2000). This review will focus on the technological aspects of metabonomic platforms, data collection and analysis, and will emphasize the use of metabonomics in understanding the pathophysiological changes associated with toxicological responses.

1. TERMINOLOGY: METABONOMICS OR METABOLOMICS

In reviewing the literature related to this field, two terms, namely metabonomics and metabolomics have been used by investigators. Originally, the field was described as metabonomics and defined as the ''quantitative measurement of time-related multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification'' (3). The term takes it roots from the Greek words ''meta'' (change) and ''nomos'' (rules or

laws) referring to chemometric models used to classify changes in metabolism (4). In contrast, the term metabolomics, although used widely is less well defined. Fiehn (5) described it as the ''comprehensive and quantitative analyses of all metabolites,'' which although more concise, does not appear to differ from the original definition of metabonomics. Some investigators have also used metabolomics to denote the measurement of metabolites in cells or cell systems. To confound and confuse the situation further, metabonomics and metabolomics have been described as subsets of each other (5,6). In this review, the term metabonomics is used, and is considered to represent a systems approach to the analysis of endogenous metabolites in biofluids or tissues using analytical methods and pattern recognition technology. From a practical perspective, metabonomics encompasses the application of NMR spectroscopy, HPLC, GC, and/or MS analyses coupled with pattern recognition tools and multivariate statistical methods to evaluate endogenous metabolites in biofluids and tissues. Regardless of terminology, the technology represents a potentially powerful method for determining the systemic response to toxicity or disease.

2. METABONOMIC PLATFORMS

Just as transcriptomic and proteomic applications have used a variety of platforms and approaches, metabonomics also has utilized a variety of platforms, and the field continues to evolve. Historically, NMR spectroscopy has been the workhorse of the approach. Briefly, NMR spectroscopy is a characterization technique in which a sample is immersed in a strong, static magnetic field and exposed to an orthogonal low-amplitude, high frequency radio frequency. The most common application is ^{1}H (proton) NMR, and typically, chemical shifts, associated with various function groups are used to establish compound structure. For metabonomic applications, proton NMR spectroscopy is typically used, and it is a non-destructive and non-invasive method that offers the advantages of minimal sample preparation, as it requires only sample dilution, pH adjustment, and the addition of an internal standard. It is also broadly applicable to a variety of biofluids (including urine, plasma, saliva, etc.) and tissues. Sample throughput is high, with analysis time under 30 min. The analytical method is quantitative and provides detailed structural information. For analysis of biofluids, time course evaluations can be readily conducted, providing for assessment of baseline parameters and to follow the progression of toxic change from onset to resolution. Although thousands of metabolites are theoretically present, the most common metabolites identified by NMR analyses, especially in urine, are typically those associated with major endogenous metabolic pathways. In particular, major metabolites are those involved in the Kreb's cycle and include intermediates such as citrate, succinate, oxaloacetate, and α-ketoglutarate. Additional metabolites that are frequently

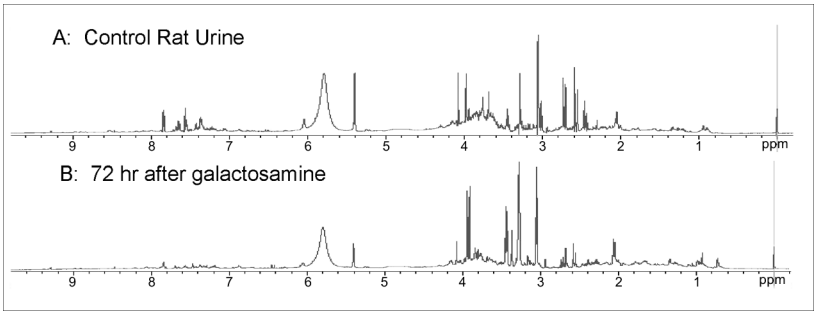


Figure 1 ^{1}H NMR spectra of urine obtained from a rat before and after oral administration of galactosamine (500 mg/kg). Urine was collected over a 24-hour period and prepared for NMR analyses (600 mHz) following addition of an internal standard (3-trimethylsilyl propionic acid).

detected include acetate, lactate, hippurate, glucose, creatine, creatinine, trimethylamine oxide, and taurine (7). The time-dependent changes in metabonomic data are illustrated in Figure 1, which provides an example of NMR spectra collected from urine of a rat prior to (control) and 72 hours following administration of the hepatotoxic agent, galactosamine.

The potential disadvantages of NMR spectroscopy include its recognition as a generally insensitive technique, often requiring at least 100 ng of mass for detection. The use of additional analytical methods, such as MS, can help to overcome this shortcoming. In all cases, the analytical tools are generally expensive, and for NMR, this includes both the instrumentation as well as the laboratory requirements to accommodate the instrument and its sensitive magnet (at least 600 MHz).

Recent research efforts have included the use of NMR coupled with MS, thereby enabling more complete identification of metabolites. The sensitivity of MS methods is a clear advantage over NMR and the ability to separate metabolites by HPLC or GC coupled with MS and/or NMR methods. The GC-based separation methods provide for analysis of volatile compounds or compounds made to be volatile by derivitization, whereas HPLC methods, although biased by solvent systems and columns, can accommodate a larger range of molecules. In both cases, the separation methods afford investigators the opportunity to increase resolution and separation of metabolites, and this enhancement in overall selectivity is particularly useful when novel biomarkers are sought.

Another useful and unique component of the NMR platform is the potential to analyze whole tissue by magic angle spinning (MAS). In this application, samples are spun rapidly at 54.7°, the so-called magic angle, relative to the applied magnetic field. In doing so, line-broadening effects that would normally obfuscate proton spectra of a solid sample are reduced.

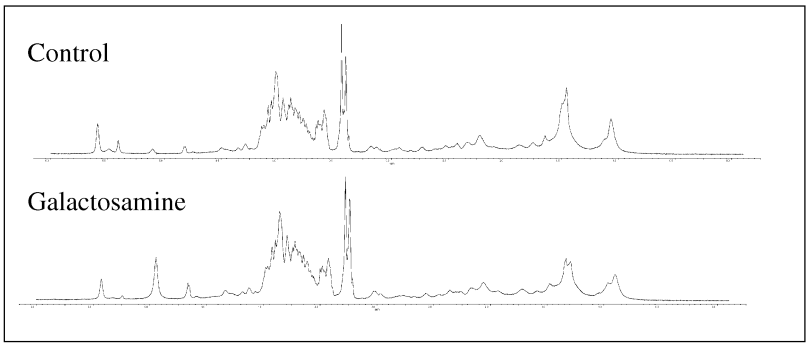


Figure 2 ^{1}H NMR-MAS spectra of liver from a control and galactosamine-treated rat. The liver sample was collected 8 hours after administration of galactosamine (500 mg/kg). Approximately 20 mg of liver tissue was spun at 5 kHz at the magic angle and the ^{1}H-NMR spectra were measured at 600 mHz.

The MAS does not represent a rapid throughput procedure, but because there is no sample extraction or other manipulation prior to analysis, it represents an unbiased method in which changes in tissues can be studied, and as appropriate, compared to alterations observed in biofluids. In this manner, MAS is synergistic with and complimentary to analysis of biofluids by metabonomic applications. Figure 2 shows the typical NMR-MAS spectra from the liver of a control and galactosamine-treated rat. This method readily detects glucose and glycogen, choline and related metabolites and a variety of lipids, and fatty acids. An example of the utility of NMR-MAS data in concert with biofluid analyses is described later in this review.

3. METABONOMIC DATA ANALYSES

As with any omics technology, the ability to generate large and complex data sets requires sophisticated methods to reduce and analyze data. In reality, the application of methods for data analysis is as important, if not more important, than the platform used to generate the data. Chemometrics is designated as the application of statistical methods to chemistry (Ref. 8; Lavine 2004) and with respect to metabonomics, includes approaches used to process NMR spectra and analyze peak alignments and normalize data. Chemometrics is distinguished from bioinformatics, which involves the storage, retrieval, and analysis of biological information stored in computer databases. To date, chemometric methods for analysis of NMR data are more fully developed than for other separation techniques. This is largely

because the NMR approaches have been used for many more years, enabling more comprehensive assessment of data analysis methods (9–11).

Principal component analysis (PCA) is commonly used to identify those analytes that are most different from the control samples and provides for a visual characterization of the data set. Following data reduction, PCA is used to find linear combinations (eigenvectors) of the original resolved peaks most different from controls, and these vectors are used to create visually characterize data sets. The PCA Eigenvectors have several desirable properties, including: (a) the combinations are not correlated and (b) they can be rank-ordered (from most to least).

An example of PCA analysis derived from the NMR spectra of rat urine from young, healthy, and aging male and female rats in which chronic progressive nephropathy has developed is shown in Figure 3. Chronic progressive nephropathy is an age-related phenomenon observed in rats, particularly males. It is characterized clinically by advancing proteinuria, and its histopathological features include degeneration of the renal tubular epithelium, glomerular lesions, interstitial inflammation, and fibrosis. The PCA plot provides a general illustration that urinary profiles obtained from male and female, young and old rats are metabolically different as they occupy

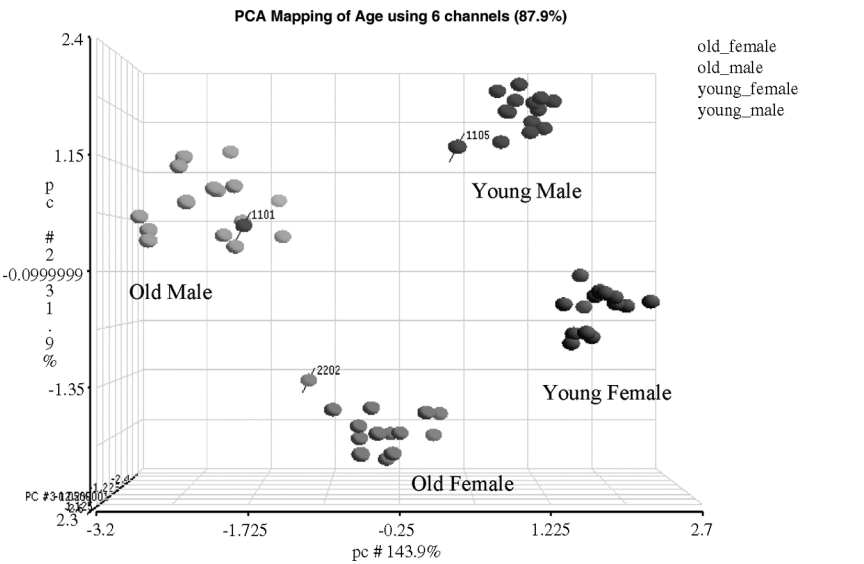


Figure 3 The PCA showing the metabolic differences in urinary metabolite profiles between young (3 months) and old (15 months) male and female Sprague–Dawley rats. Older rats develop chronic progressive nephropathy that affects renal histopathology and function, and these biochemical differences have been captured by metabonomic analyses.

very different three-dimensional space. Moreover, as shown in the plot, there is a young male rat that appears in the same space as the aging male rats, suggesting that this animal may have a renal defect.

As noted earlier, there is a common set of metabolites typically identified by NMR analyses, especially in urine, associated with major endogenous metabolic pathways. However, there is clear evidence that the pattern of changes noted, even with these commonly occurring metabolites, is altered in a toxicity- and tissue-specific manner, a feature that is extremely powerful with respect to the predictive utility of metabonomic data. As such, pattern recognition methods are as important as individual metabolite identification. For pattern recognition and predictive model development, additional multivariate statistical models are required. There are many examples of the application of pattern recognition techniques for characterizing and interpreting NMR spectral data (12–15), with most of these methods applied to the evaluation of toxicological responses or organ-specific toxicity. To illustrate this utility, Table 1 is a summary of the patterns of changes noted by a variety of kidney toxicities, including chronic progressive nephropathy. These data illustrate how metabonomic information and patterns of changes can be used to assess toxicity, and how it is essential to consider data sets from a multi-variate perspective rather than a univariate mindset. At the same time, there is the clear potential to identify unique and/or specific endogenous metabolites that are associated with a specific type of change, enabling the identification of biomarkers of toxic changes. Although this discussion represents a superficial overview of data analyses tools, it serves to emphasize that the successful application of metabonomic data to toxicology issues will always require the coordinated, multidisciplinary effort of toxicologists, analytical chemists, statisticians, chemometricians, and bioinformaticists.

Table 1 Metabolite Changes Observed in Rat Urine in Response to Renal Toxicity

Site of toxicity	Increased	Decreased
Kidney–Cortex (S1)	Glycine, glutamate, and glucose	Hippurate
Kidney–Cortex (S2,3)	Alanine, glutamine, valine, glucose, lactate, and 3-hydroxybutyrate	Creatinine
Kidney–Medulla	Acetate, succinate, dimethylamine, and *N*,*N*-dimethylglycine	2-Oxo-glutarate and trimethylamine-*N*-oxide
Age-related nephropathy	Glycine, glucose, and amino acids	Hippurate, creatinine, and 2-oxoglutarate

4. APPLYING METABONOMICS IN BIOLOGICAL SCIENCES: TOXICOLOGICAL RESEARCH

Metabonomics has received considerable attention in the toxicological community, but based on published literature, metabonomic applications do not appear to be as widely used as transcriptomics (toxicogenomics). However, this is not a reflection on the utility of the approach, but more likely a consequence of the need to have a complex infrastructure, particularly with respect to data analysis and interpretation in order to carry out the technology.

To date, a variety of work has been published which describes the metabonomic evaluation of urine of animals in response to toxicological insults, with some, albeit fewer investigators reporting analysis of plasma and tissues. In general, metabonomic analyses in toxicology have focused mainly on the identification of changes associated with liver or kidney toxicity (13,16). In this regard, changes detected in metabonomic patterns do not always directly correlate with histopathological changes and frequently precede those detected through clinical chemistry analyses. However, after extensive data analysis and application of pattern recognition tools, it is clear that metabonomic data can be used to accurately classify compounds as causing hepatic and/or kidney injury (13). More recently, investigations have also shown that metabonomic data can also be used to assess the development of vascular lesions and vasculitis in rats (17,18).

One potential issue with metabonomic analyses concerns its overall sensitivity to environmental and physiological changes. For example, in Figure 3, age-dependent changes in renal function in rats, especially male rats, alters the typical urinary metabolite profile observed in untreated rats. In the example given, young and old rats were defined as 3 and 15 months of age. However, Robertson et al. (16) showed that changes in renal function during the conduct of a typical subchronic toxicity study (13 weeks) could influence data interpretation. Similarly, variation in metabonomic patterns occurs during the normal estrus cycle (19), and any change in the general health status of the animals, most notably alterations in the resident flora within the gastrointestinal tract, can alter typical metabonomic profiles (20). On one hand this can be a disadvantage, but if carefully controlled, monitored and characterized, this sensitivity is easily dealt with. Conversely, as illustrated in the PCA plot shown in Figure 3, the sensitivity may be useful for identifying animals that are biochemically or physiologically outliers relative to a normal group distribution.

With respect to preclinical studies, one interesting observation determined in metabonomic studies, is that vehicles used for compound administration are not biologically inert, and may influence the analysis of metabonomic data. In fact, some commonly used vehicles for compound administration are not compatible with metabonomics applications, most notably Labrofil® and polyethylene glycol (21). Finally, because NMR is

a non-discriminating tool, the profiles of compounds administered to animals to evaluate metabonomic changes can confound the interpretation of the NMR data. In this regard, the use of additional methods, including chromatography and MS detection can reduce or eliminate this potential interference.

Metabonomic evaluations have also been used to identify potential biomarkers of toxicity. One such example is the detection of phenylacetylglycine (PAG) as a potential useful biomarker for compound-induced phospholipidosis (22). As with any biomarker, careful validation is required to assess the overall utility of the biomarker (across species, sensitivity, and specificity) and efforts continue in this regard. However, a biomarker need not be a single metabolite. For example, the combination of changes in the urinary levels of four metabolites (trimethylamine-*N*-oxide, *N*,*N*-dimethylglycine, dimethylamine, and succinate) has been shown to accurately predict renal papillary necrosis (Table 1; Ref. 23). Prior to the application of metabonomics, the diagnosis of both phospholipidosis and renal papillary necrosis required histopathological evaluation. Accordingly, the ability to assess these toxicities in a non-invasive manner incorporating metabonomic analysis of urine represents an important scientific advancement.

A relevant and critical question regarding the utility of metabonomics concerns its overall comparison to other omic platforms, including transcriptomics and proteomics. This is an important consideration that has to date, not been fully evaluated. For perspective and consideration, Table 2 summarizes changes observed in the mouse following administration of a hepatotoxic dosage of acetaminophen (APAP). This is one of only a few examples of such a ''trans-platform'' comparison. The APAP is metabolized to a reactive intermediate that covalently modified macromolecules to cause severe centrilobular hepatic necrosis. The early events of necrotic cell death include mitochondrial defects, and the metabonomic data point to a shift in overall energy metabolism as a major change in the liver. In this particular example, the transcriptomic and proteomic results were only partial analyses, in that the investigators focused on changes in mitochondrial mRNAs and proteins. However, the results demonstrated an early change in specific genes and mitochondrial proteins that correlated with time-dependent changes in the metabonomic profiles. For example, down-regulation of the lipoprotein lipase gene, which catalyzes the hydrolysis of triglycerides, correlated with the increase in triglycerides in the liver. Similarly, the increase in mRNA for D-β-hydroxybutyrate dehydrogenase was reflected in the plasma metabonomic results in which D-β-hydroxybutyrate increased (24–26). These results suggest that the various platforms are likely to be complimentary to each other. Moreover, the results demonstrate how metabonomics provides the tools to evaluate the downstream or phenotypic effect of changes in gene and protein expression that may ultimately be more

Table 2 Comparison of Transcriptomic, Proteomic, and Metabonomic Changes Observed in Mice After Administration of Acetaminophen

Analysis	Increased	Decreased
Transcriptomics	Choline kinase	Lipoprotein lipase
	Oxysterol binding protein	VLDL receptor
	Uteroglobin	Fatty acid binding protein
	Insulin-like growth factor binding protein	B-*N*-acetyl Hexosaminodase
	ATP synthesis protein 8	HMG CoA-reductase
		Cholesterol 7α-hydroxylase
		ATP synthesis λ chain
Proteomics	Acetyl CoA acetyltransferase	3-Ketoacyl CoA thiolase
	D-β-hydroxybutyrate dehydrogenase	Aldehyde dehydrogenase
	Mito stress 70 protein p1	ATP synthesis α chain
		Heats shock protein (HSP) 10
		HSP 60
		Mitochondrial thioredox-dependent peroxide reductase
		Mitochondrial matrix protein p1
Metabonomics		
Liver-MAS	Triglycerides	Glucose, glycogen
		Alanine
		Phospholipids
Liver-lipid extract	Triglycerides	Phospholipids
Liver-aqueous extract	Choline, phosphocholine	Glucose
	Glycerophosphocholine	
	Lactate	
	Amino acids	
Plasma	Glucose	
	Pyruvate	
	Acetate	
	Lactate	
	3-D-hydroxybutyrate	

Source: Compiled from Refs. 26–28.

stable and, because plasma or urine samples are readily obtained, is likely to be more easily and more routinely evaluated.

As described herein, many metabonomic studies conducted to date have focused on assessing the patterns of change associated with toxicity and reporting the kinds metabolites that have been altered by chemical

treatment or physiological alteration. These reports are extremely useful for establishing databases that can be used for predicting toxic liabilities. However, for all of its power, efforts to use metabonomics to identify mechanisms of toxicity are also important. For example, recent work from Mortishire-Smith et al. (27) provided evidence of altered fatty acid metabolism as a mechanism of a novel drug-related hepatotoxicity and showed the potential for metabonomics to address mechanistically based hypotheses. This type of work, linking metabonomic results to better definition of mechanisms of toxicity will ultimately enhance this field by enabling more mechanistically founded predictive models to be evaluated and by providing the opportunity to translate findings in laboratory animals to the clinical situation.

5. APPLYING METABONOMICS TO BIOLOGICAL SCIENCES: CLINICAL APPLICATIONS

The NMR methods have been used in clinical medicine for many years, and metabonomic evaluation of human samples has been conducted for at least the past 10 years (28). Classical examples include the application of NMR to the evaluation of inborn errors of metabolism (29). More recent work has applied metabonomics to the evaluation of the clinical severity of coronary artery disease and to establish a relationship between serum metabolic profiles and hypertension (30,31). Because metabonomics is highly sensitive to environmental or dietary influences (discussed under toxicological applications), concern has been raised that the natural variation in the human population would preclude the application of metabonomics to clinical problems. However, such concerns have been dealt with directly, and recently, Lenz et al. (32) demonstrated that urine and plasma could be collected from human subjects and used successfully for metabonomic analyses. Furthermore, in addition to the disease states described above, metabonomics has been shown to a potentially useful tool for describing alterations associated with dietary and nutritional practices (33).

6. CONCLUSIONS

As the use of metabonomics advances, there are several challenges facing scientists using this tool that must be addressed in order to make it more mainstream and more relevant to predicting toxicity, and useful for hazard identification, human risk assessment and clinical medicine. First, advancing the use of metabonomics to identify mechanisms of toxicity is essential, and such efforts should help to increase the overall usefulness, validity, and relevance of toxicity prediction and biomarker development. Second, the use of metabonomic evaluations in the course of chronic toxicity rather than the heretofore emphasis on acute studies will help to establish its place in following the

progression of toxicity or disease. Third, the application of metabonomics to biomarker identification from animal studies that can be translated to clinical use is an important challenge. To this end, rats have been the major animal model used in most metabonomic evaluations, but extension to other preclinical species, including non-human primates, is important. Ultimately, clinical evaluation of the technology is an important challenge and goal, and its use in clinical medicine is burgeoning. The technology will find even greater utility in clinical medicine when additional basic research has identified important biomarkers of toxicity or disease, or, as mechanistic information about toxicity or diseases is uncovered. Finally, although metabonomic data can be used independently to evaluate disease state or chemically induced insult, the integration of metabonomic data with other tools, particularly transcriptomics and proteomics, is necessary and critical and may also help to validate results obtained from these approaches.

REFERENCES

1. Roberts JK, Xia JH. High resolution NMR methods for study of higher plants. Methods Cell Biol 1995; 49:245–258.
2. Weckwerth W. Metabolomics in systems biology. Annu Rev Plant Biol 2003; 54:669–689.
3. Nicholson JK, Lindon JC, Holmes E. Metabonomics: understanding the metabolic responses of liver systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 1999; 29:1181–1189.
4. Lindon JC, Holmes E, Nicholson JK. Toxicological applications of magnetic resonance. Prog Nucl Magn Reson Spectrosc 2004; 45:109–143.
5. Fiehn O. Combining genomics, metabolome analysis and biochemical modeling to understand metabolic networks. Comput Funct Genomics 2001; 2: 155–169.
6. Lindon JC, Holmes E, Nicholson JK. So what's the deal with metabonomics. Anal Chem 2003; 75:384A–391A.
7. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov 2002; 1:153–161.
8. Deming SN. Chemometrics: an overview. Clin Chem 1986; 32:1702–1706.
9. Gartland KP, Beddell CR, Lindon JC, Nicholson JK. Application of pattern recognition methods to the analysis and classification of toxicological data derived from proton nuclear magnetic resonance spectroscopy of urine. Mol Pharmacol 1991; 39:629–642.
10. Farrant RD, Lindon JC, Rahr E, Sweatman BC. An automatic data reduction and transfer method to aid pattern recognition analysis and classification of NMR spectra. J Pharm Biomed Anal 1992; 10:141–144.
11. El-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. NMR Biomed 1997; 10:99–124.

12. Holmes E, Nicholoson JK, Bonner FW, Sweatman BC, Beddell CR, Lindon JC, Rahr E. Mapping the biochemical trajectory of nephrotoxicity by pattern recognition of NMR urinanalysis. NMR Biomed 1992; 5:368–372.
13. Holmes E, Nicholls AW, Lindon JC, Connor SC, Connelly JC, Haselden JN, Damment SJP, Spraul M, Neidig P, Nicholson JK. Chemometric models for toxicity classification based on NMR spectra of biofluids. Chem Res Toxicol 2000; 13:471–478.
14. Holmes E, Antti H. Chemometric contribution to the evolution of metabonomics: mathematical solutions to characterizing and interpreting complex biological NMR spectra. Analyst 2002; 27:1549–1557.
15. Stoyanova R, Nicholson JK, Lindon JC, Brown TR. Sample classification based on Bayesian spectral decomposition of metabonomic NMR data sets. Anal Chem 2004; 76:3666–3674.
16. Robertson DG, Reily MD, Sigler RE, Wells DF, Paterson DA, Braden TK. Metabonomics: evaluation of nuclear magnetic resonance (NMR) and pattern recognition technology for rapid in vivo screening of liver and kidney toxicants. Toxicol Sci 2000; 57:326–337.
17. Slim RM, Robertson DG, Albassam M, Reily MD, Robosky L, Dethloff LA. Effect of dexamethasone on the metabonomics profile associated with phosphodiesterase inhibitor-induced vascular lesions in rats. Toxicol Appl Pharmacol 2002; 183:108–116.
18. Robertson DG, Reily MD, Albassam M, Dethloff LA. Metabonomic assessment of vasculitis in rats. Cardiovasc Toxicol 2001; 1:7–19.
19. Bollard ME, Holmes E, Lindon JC, Mitchell SC, Branstetter D, Zhang W, Nicholson JK. Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution ^{1}H-NMR spectroscopy of urine and pattern recognition. Anal Biochem 2001; 295:194–202.
20. Nicholls AQ, Mortishire-Smith RJ, Nicholson JK. NMR spectroscopic-based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. Chem Res Toxicol 2003; 16:1395–1404.
21. Beckwith-Hall BM, Holmes E, Lindon JC, Gounarides J, Vickers A, Shapiro M, Nicholson JK. NMR-based metabonomic studies on the biochemical effects of commonly used drug carrier vehicles in the rat. Chem Res Toxicol 2002; 15:1136–1141.
22. Espina JR, Shockcor JP, Herron WJ, Car BD, Contel NR, Ciaccio PJ, Lindon JC, Holmes E, Nicholson JK. Detection of in vivo biomarkers of phospholipidosis using NMR-based metabonomic approaches. Magn Reson Chem 2001; 36:559–565.
23. Holmes E, Bonner F, Nicholson JK. Comparative studies on the nephrotoxicity of 2-bromoethanamine hydrobromide in the Fischer 344 rat and the multimammate desert mouse (*Mastomys natalensis*). Arch Toxicol 1995; 70: 89–95.
24. Coen M, Lenz EM, Nicholson JK, Wilson ID, Pognan F, Lindon JC. An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. Chem Res Toxicol 2003; 16:295–303.
25. Coen M, Ruepp SU, Lindon JC, Nicholson JK, Pognan F, Lenz EM, Wilson ID. Integrated application of transcriptomics and metabonomics yields new

insight into the toxicity due to paracetamol in the mouse. J Pharm Biomed Anal 2004; 35:93–105.
26. Ruepp SU, Tonge RP, Shaw J, Wallis N, Pognan F. Genomics and proteomic analysis of acetaminophen toxicity in mouse liver. Toxicol Sci 2002; 65:135–150.
27. Mortishire-Smith RJ, Skiles GL, Lawrence JW, Spence S, Nicholls AW, Johnson BA, Nicholson JK. Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of drug-induced toxicity. Chem Res Toxicol 2004; 17:165–173.
28. Holmes E, Foxall PJ, Nicholson JK, Neild GH, Brown SM, Beddell CR, Sweatman BC, Rahr E, Lindon JC, Spraul M, Neidig P. Automatic data reduction and pattern recognition methods for analysis of ^{1}H nuclear magnetic resonance spectra of human urine from normal and pathological states. Anal Biochem 1994; 220:284–296.
29. Lindon JC, Nicholson JK, Everett JR. NMR spectroscopy of biofluids. Ann Rep NMR Spectrosc 1999; 38:1–88.
30. Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, Clarke S, Schofield PM, McKilligin E, Mosedaile DE, Grainger DJ. Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using ^{1}H-NMR-based metabonomics. Nat Med 2002; 8:1439–1444.
31. Brindle JT, Nicholson JK, Schofield PM, Grainger DJ, Holmes E. Application of chemometrics to ^{1}H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. Analyst 2003; 128:32–36.
32. Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF. A ^{1}H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal 2003; 33:1103–1115.
33. German JB, Roberts MA, Watkins SM. Personal metabolomics as a next generation nutritional assessment. J Nutr 2003; 133:4260–5266.
34. Fiehn O. Metabolomics—the link between genotypes and phenotypes. Plant Mol Biol 2002; 48:155–171.
35. Lavine B, Workman JJ Jr. Chemometrics. Anal Chem 2004; 76:3365–3371.

15

Multiplex Minisequencing on Microarrays: Application to Pharmacogenetics of Antihypertensive Drug Response

Ulrika Liljedahl and Ann-Christine Syvänen

Department of Medical Sciences, Uppsala University, Uppsala, Sweden

1. INTRODUCTION

The DNA sequences of different individuals are identical to 99.9% (1,2). The remaining variable 0.1% consists mainly of single-nucleotide polymorphisms (SNPs), which explain a large fraction of the inter-individual differences in phenotype between populations and individuals. Depending on the location of an SNP, the phenotypic consequences of it may differ. The SNPs within the protein coding regions of a gene may alter an amino acid and thus change the structure and function of the protein. A nucleotide change can also result in a synonymous change that has no effect on the amino acid in the protein. The SNPs in regulatory, non-coding regions of a gene may affect the binding of transcription factors and thereby the expression level of genes (3). The majority of the SNPs are located in the non-coding parts of the genome, where they have no currently known functional effect.

The SNPs are useful as genetic markers owing to their frequent and even distribution in the genome. They have low mutation rates, and since they are biallelic they are suitable for high throughput automated genotyping assays. The SNPs are used as markers in forensic identification, in tissue

typing, for population genetic studies and evolutionary studies. SNPs (point mutations) causing monogenic disorders have been routinely analyzed for diagnostics and identification of disease carriers for more than a decade. In pharmacogenetics, SNPs in genes for drug metabolizing enzymes are analyzed to assess an individual's response to drug treatment (4). The most well-characterized drug metabolizing enzymes are the cytochrome P450 enzymes (CYP450), which are responsible for metabolizing numerous drugs in clinical use today. The CYP450 enzymes harbor genetic variants that are important determinants of drug response (5). As also other molecules than drug-metabolizing enzymes, such as drug receptors or transporters, are becoming targets for pharmacogenetic analysis (6,7) this field is a rapid growing application area of SNP typing today. For example, a SNP leading to the change of glycine to arginine at amino acid position 17 in the β_2-adrenergic receptor has been associated with the response to β-agonist therapy in asthma (8). Genotyping of SNPs in thiopurine methyltransferase (TPMT) is required prior to treatment of acute lymphoblastic leukaemia with thiopurines in children (9).

Essential hypertension has a multifactorial aetiology with interactions of many genetic and environmental factors through the intermediate systems regulating blood pressure control (10,11). Hypertension has a prevalence of 25–35% in the adult population in developed countries (12). Twin studies indicate that the heritability of blood pressure is 30–70% (13). About half of the individuals with hypertension are treated, and only 50% of them have adequate blood pressure control although multiple drugs are used (14,15). The problem of poor drug response may be due to the heterogeneity of hypertension as a disease entity, as well as to a large inter-individual variation in drug response. There are no biomarkers available today to determine in advance how an individual patient will respond to treatment by a specific antihypertensive drug. Identifying genetic variants that could be used as predictors for the outcome of antihypertensive treatment would help identifying the best drug for the individual patient, thus reducing the cost for the treatment and its duration (16,17). We developed a microarray-based minisequencing system for genotyping of SNPs in candidate genes for hypertension.

2. MICROARRAY-BASED GENOTYPING

The major reaction principles that form the basis of the genotyping technologies available today are primer extension, ligation, and hybridization. Many of the methods for SNP genotyping rely on PCR amplification of the sequence of interest prior to allele determination. Primer design and assay optimization for multiplexed and reproducible genotyping of SNPs in large sample sets have become major bottle necks in most methods used today.

The microarray format is attractive for analyzing large sets of SNPs because it allows simultaneous and highly parallel genotyping of multiple sequence variants. The cost for reagents is also reduced owing to the miniaturized format of the microarrays. The microarray format was first designed for expression profiling, where typically very large numbers of mRNA species are analyzed in a relatively small number of samples (18,19). The standard microscope slide format used for expression profiling, where one sample is analyzed per slide, is not practical for SNP genotyping studies, where a large number of samples are to be analyzed for each set of SNPs. To circumvent this problem, we have devised an "array-of-arrays" conformation that allows parallel analysis of up to 80 samples for each set of SNPs on a single microscope slide (7,20,21) (Fig. 1). Each microarray is divided into multiple separate reaction wells by a silicon rubber grid that is placed on the microscope slide. A similar "array-of-arrays" concept is utilized in a 384-well-microtiter-plate format instead of using a microscope slide (GenomeLab SNPstream system, Beckman Coulter, (22)). The "array-of-arrays" format was originally devised for genotyping by primer extension (20), but the format can equally well be used with all other reaction principles for SNP-typing.

In "minisequencing" single nucleotide primer extension (SNE) or single base extension (SBE), a DNA polymerase is used to extend a detection primer, which anneals immediately adjacent to the site of the SNP, with a labeled nucleotide analog (23,24). In the microarray format of

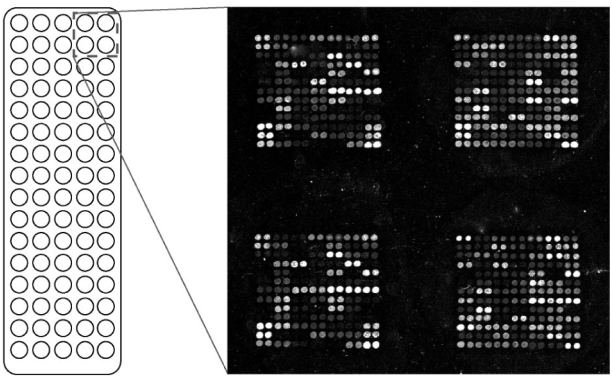


Figure 1 In the "array-of-arrays" conformation a standard microscope slide is divided into 80 subarrays with a diameter identical to that of a 384-well-microtiter-plate reaction well, schematically illustrated to the left in the image. Up to 200 oligonucleotide spots can be printed per subarray at a center-to-center distance of 200 μm. With the possibility to analyze 80 samples in parallel, up to 16,000 genotypes per slide can be generated. Fluorescence images of four reaction wells, where the minisequencing reaction have been performed are shown to the right.

minisequencing, also denoted arrayed primer extension (APEX), the SNP-specific detection primers are attached covalently to the surface of activated microscope slides through their 5′-end, allowing their 3′-ends to be extended (7,25–27) (Fig. 2). Different types of slides and a variety of chemical attachment methods are available. We use the CodeLink™ activated slides from GE Healthcare since they performed best in our system according to a comparison with five other slide types (28).

To achieve the high parallelism of genotyping in the microarray format, the DNA regions comprising the SNPs are amplified in several multiplex PCR reactions, which are combined prior to the minisequencing reactions on the arrays. Reproducible multiplexing of more than 10 fragments in multiple samples has proven to be a demanding task. The success of a multiplex PCR depends on the primer design, which aims at avoiding unspecific interactions between different primers, and achieving uniform amplification efficiency for all fragments.

The multiplex PCR products are allowed to hybridize to the immobilized primers on the array surface to serve as template in the minisequencing reaction. After removing the unhybridized templates by washing the microarray slide, the primers are extended with fluorescently labeled ddNTPs that are complementary to the nucleotide at the SNP site. The fluorescence signals of the extended primers are detected on the slide using a four-color laser scanner (e.g., ScanArray 5000, Perkin-Elmer LifeSciences). The signal intensities are measured with the analysis software of the scanner (QuantArray®). The mean value of the signals from duplicate spots is

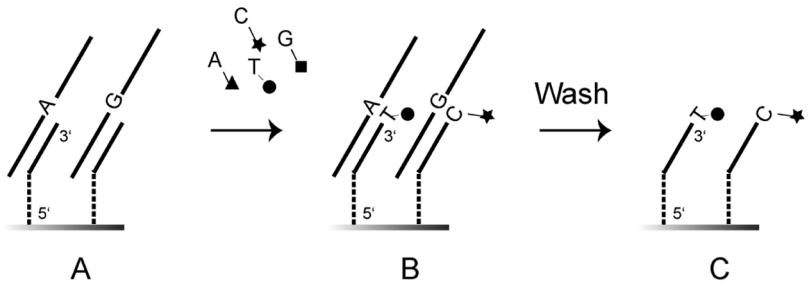


Figure 2 Minisequencing reaction on with extension primers immobilized on the microarray surface. The multiplex PCR products of the regions containing the SNPs are allowed to hybridize to the oligonucleotides immobilized on the microarray (**A**). The primers are extended with fluorescently labelled ddNTPs complementary to the nucleotide at the SNP position. Four different fluorophores, one for each nucleotide is used allowing for a simultaneous detection of the four nucleotides in one single reaction (**B**). After washing the slide with sodium hydroxide and salt buffers, only the extended primers covalently attached to the surface remains and the fluorescence is measured (**C**) and the genotypes assigned.

corrected for the average background in the reaction well. Genotypes are assigned by calculating the ratio between the signal intensity from one of the alleles divided by the sum of the signals from both alleles. This is easily done manually or using Excel® when analyzing only a few samples and SNPs. However, our "array-of-arrays" format generates >60,000 fluorescence signals per array. Analysis of such a large amount of genotyping data is most conveniently done by cluster analysis of the fluorescence signals. In our system, scatter plots with the sum of the fluorescence signals for the two alleles on the *y*-axis and the signal intensity fraction plotted on the *x*-axis are generated for each SNP. This should theoretically result in three distinct clusters of signal intensity fractions close to 0, 0.5, and 1, respectively, corresponding to the three SNP genotypes. For SNP genotype assignment based on cluster analysis, we use the SNPSnapper software (http://www.bioinfo.helsinki.fi/SNPSnapper/).

3. PHARMACOGENETICS OF ANTIHYPERTENSIVE DRUG RESPONSE

We established our microarray-based minisequencing system for a panel of 74 SNPs located in candidate genes of pathways related to blood-pressure regulation (7). These SNPs were genotyped in 97 hypertensive individuals participating in the multicenter Swedish Irbesartan Left Ventricular Hypertrophy Investigation versus Atenolol (SILVHIA) trial (29). Initially, we performed a stepwise multiple regression analysis of variance (ANOVA) in which SNPs with $p < 0.1$ were retained in the model. By this exploratory analysis, we identified combinations of four to five SNP genotypes that explain approximately 50% of the variation in the reduction of systolic and diastolic blood pressure in the two treatment groups (Fig. 3). For example, the individuals with the favorable genotype combination CYP11B2 T267C TT, EDNRB G40A AA, APOA-IV A1449G AA, and ENOS G498A GG showed an average reduction in systolic blood pressure of 45 mm Hg after irbesartan treatment, compared to 11 mm Hg in individuals with other genotype combinations.

We also performed a more detailed analysis to identify association between individual SNP genotypes in genes in the renin-angiotensin-aldosterone system (RAAS) and lipid metabolism and blood pressure reduction. Two SNPs, AGT T1198C causing a methionine to threonine substitution at amino acid residue 235 of angiotensinogen (AGT), and AGT A1218G, located six nucleotides up-stream the transcription start site of the gene, were related to the reduction in SBP during atenolol treatment (Table 1) (30). In a similar fashion, carriership of the C-allele of the C711T SNP in the apolipoprotein B (APOB) gene was found to be associated with the reduction in systolic blood pressure ($p = 0.004$) during irbesartan treatment, while the individuals homozygous for the T-allele showed no reduction in

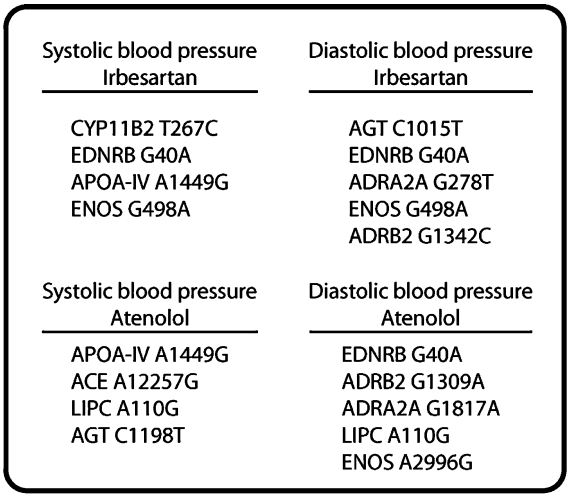


Figure 3 Combinations of SNP genotypes in association to blood pressure reduction during antihypertensive treatment. ACE, angiotensin converting enzyme; ADRA2A, adrenergic α_2 receptor; ADRB2, adrenergic β_2 receptor; AGT, angiotensinogen; APOA-IV, apolipoprotein A-IV; CYP11B2, aldosterone synthase; EDNRB, endothelin receptor type B; ENOS, endothelial nitric oxide synthase; LIPC, lipase hepatic.

systolic blood pressure (Fig. 4). The same pattern of response related to genotype was seen for diastolic blood pressure, although this finding did not reach statistical significance. In the atenolol treatment group, the SNP C16730T in the low-density lipoprotein receptor gene (LDLR) showed association to the reduction in systolic blood pressure. Presence of the C-allele was related to blood pressure reduction ($p = 0.006$), while the individuals

Table 1 Single Nucleotide Polymorphisms Associated to Reduction in Systolic Blood Pressure Upon Antihypertensive Treatment

Gene	SNP[a]	Treatment[b]	*p*-Value[c]
Angiotensinogen	AGT T1198C	β-blocker	0.03
Angiotensinogen	AGT A1218G	β-blocker	0.008
Apolipoprotein A-IV	APOA-IV A1449G	AT_1 blocker	0.054
Apolipoprotein A-V	APOA-V C34155T	AT_1 blocker	0.049
Apolipoprotein B	APOB C711T	AT_1 blocker	0.004
Low-density lipoprotein receptor	LDLR C16730T	β-blocker	0.006

[a]Gene acronym, nucleotide variation and position in gene sequence (see further Ref. 7).
[b]The active substance in the β-blocker was atenolol and in the AT_1 blocker irbesartan.
[c]*p*-Value from analysis of variance relating the genotype to the change in phenotype upon treatment (30,45).

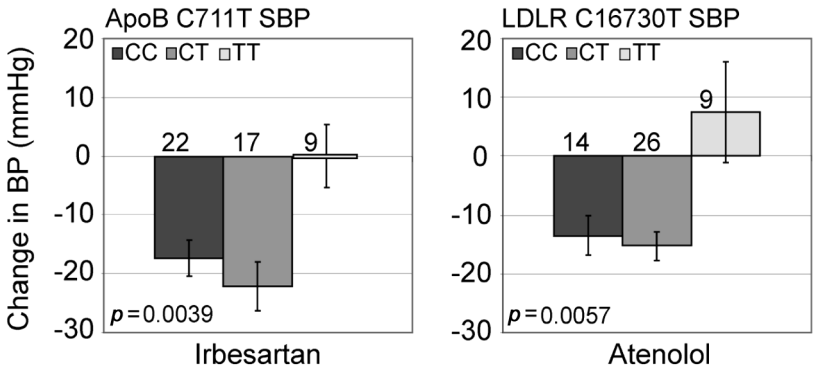


Figure 4 The average reduction in systolic blood pressure for the individuals with the favorable C-allele of the SNP ApoB C771T was 19 mmHg, compared to 0 mmHg for the individuals lacking this allele. In the atenolol treatment group, the individuals carrying the favorable C-allele of the SNP LDLR C16730T showed an average reduction of 14 mmHg in systolic blood pressure compared to an increase of 7.5 mmHg for the individuals homozygous for the T-allele.

homozygous for the T-allele ($n = 9$) actually showed an increase in systolic blood pressure (Fig. 4).

4. PROSPECTS OF MICROARRAY-BASED GENOTYPING METHODS

Most of the currently used genotyping methods depend on amplification of the genomic region of interest by the polymerase chain reaction (PCR) (31–33) to provide sufficient sensitivity and specificity to detect a SNP among the 3×10^9 base pairs of DNA that constitute the human genome. However, today PCR is the major bottleneck for high-throughput genotyping of previously known SNPs at different locations of the genome due to the difficulty of performing multiplex amplification (34). In applications where complete individual genes or exons are resequenced to detect previously unknown mutations, the problem of designing multiplex PCR is avoided to some extent (35,36).

The primer extension reaction allows specific genotyping of most SNPs at similar reaction conditions using only a single primer per SNP, which is an important feature in multiplexed assays in a microarray format. In a side-by-side comparison with ASO hybridization in the same microarray format, the minisequencing reaction provided tenfold higher power of discrimination between heterozygous and homozygous genotypes than hybridization with ASO probes (25). In an alternative format of the minisequencing system, multiplex cyclic primer extension reactions are performed in solution with primers tailed with 5′-tag sequences (21,37). The products of

the minisequencing reaction are then captured to complementary tag sequences immobilized on the microarray by hybridization. The tag-array assays are flexible in their design and production, and since the array is generic it can be used for many different sets of SNPs. The ''array-of-arrays'' format is particularly well suited for genotyping by the flexible tag–array approach (21,38). Additionally, the cyclic extension reaction also serves to increase the signal strength. The accuracy of the primer extension reactions in solution allows multiplex quantification of variant alleles present as a small minority (2–5%) of a sample (21), which is comparable with the system using immobilized extension primers (39).

Tag microarrays have also been adapted to ligase-assisted genotyping (40). Circularizable oligonucleotide ligation probes, padlock probes (41), have been used for SNP genotyping using tag oligonucleotides on microarrays (42). Padlock probes have specific target recognition sequences in their 5′ and 3′ ends and a connecting sequence between the target specific regions (41). When hybridized to its target the two ends of the probe are brought adjacent to each other, and the junction is ligated when there is a perfect match. A promising approach for increasing the specificity of genotyping SNPs directly in genomic DNA is to combine a primer extension reaction with a ligation step. Two examples of such methods are the BeadArray technology from Illumina (43), and the use of molecular inversion probes in the system from ParAllele BioScience, which uses the Affymetrix microarray platform (44). In both these systems the ligation products are amplified using a single set of universal PCR primers, which allows for equal amplification of many fragments in a highly multiplex fashion, without the need for optimization of the amplification reaction.

The new, highly multiplexed genotyping systems that are under development are promising tools for large scale, and even genome wide identification of genetic variants that predispose to multifactorial traits. They offer the promise of identifying new drug target molecules, and will most likely also contribute to the identification of combinations of SNP alleles in multiple genes that are predictive of individual drug response. Once such predictive SNP panels have been identified, microarray methods with lower complexity, such as the one we describe in this chapter, will find utility in routine clinical testing of individuals prior to drug treatment in hypertension and other complex traits.

REFERENCES

1. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 409:928–933.
2. Marth G, Schuler G, Yeh R, Davenport R, Agarwala R, Church D, Wheelan S, Baker J, Ward M, Kholodov M, Phan L, Czabarka E, Murvai J, Cutler D, Wooding S, Rogers A, Chakravarti A, Harpending HC, Kwok PY, Sherry ST.

Sequence variations in the public human genome data reflect a bottlenecked population history. Proc Natl Acad Sci USA 2003; 100:376–381.
3. Hudson TJ. Wanted: regulatory SNPs. Nat Genet 2003; 33:439–440.
4. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999; 286:487–491.
5. Ingelman-Sundberg M. Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. Naunyn Schmiedebergs Arch Pharmacol 2004; 369:89–104.
6. Johnson JA, Lima JJ. Drug receptor/effector polymorphisms and pharmacogenetics: current status and challenges. Pharmacogenetics 2003; 13:525–534.
7. Liljedahl U, Karlsson J, Melhus H, Kurland L, Lindersson M, Kahan T, Nystrom F, Lind L, Syvanen AC. A microarray minisequencing system for pharmacogenetic profiling of antihypertensive drug response. Pharmacogenetics 2003; 13:7–17.
8. Liggett SB. The pharmacogenetics of beta2-adrenergic receptors: relevance to asthma. J Allergy Clin Immunol 2000; 105:S487–S492.
9. McLeod HL, Krynetski EY, Relling MV, Evans WE. Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. Leukemia 2000; 14:567–572.
10. Guyton AC, Hall JE, Lohmeier TE, Jackson TE, Kastner PR. Blood pressure regulation: basic concepts. Fed Proc 1981; 40:2252–2256.
11. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. Cell 2001; 104:545–556.
12. Staessen JA, Wang J, Bianchi G, Birkenhager WH. Essential hypertension. Lancet 2003; 361:1629–1641.
13. Pausova Z, Gossard F, Gaudet D, Tremblay J, Kotchen TA, Cowley AW, Hamet P. Heritability estimates of obesity measures in siblings with and without hypertension. Hypertension 2001; 38:41–47.
14. WHO. 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. J Hypertens 2003; 21:1983–1992.
15. Mein CA, Caulfield MJ, Dobson RJ, Munroe PB. Genetics of essential hypertension. Hum Mol Genet 2004; 13(suppl 1):R169–R175.
16. Turner ST, Schwartz GL, Chapman AB, Hall WD, Boerwinkle E. Antihypertensive pharmacogenetics: getting the right drug into the right patient. J Hypertens 2001; 19:1–11.
17. Turner ST, Boerwinkle E. Genetics of blood pressure, hypertensive complications, and antihypertensive drug responses. Pharmacogenomics 2003; 4:53–65.
18. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995; 270:467–470.
19. Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 2003; 100:3889–3894.
20. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res 2000; 10:1031–1042.

21. Lindroos K, Sigurdsson S, Johansson K, Ronnblom L, Syvanen AC. Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. Nucleic Acids Res 2002; 30:e70.
22. Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, Modica F, Pohl M, Varde S, Zhao R, Zhao X, Boyce-Jacino MT. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques 2002; 30:S70–S77.
23. Syvanen AC, Aalto-Setala K, Harju L, Kontula K, Soderlund H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 1990; 8:684–692.
24. Syvanen AC. From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat 1999; 13:1–10.
25. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res 1997; 7:606–614.
26. Kurg A, Tonisson N, Georgiou I, Shumaker J, Tollett J, Metspalu A. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. Genet Test 2000; 4:1–7.
27. Raitio M, Lindroos K, Laukkanen M, Pastinen T, Sistonen P, Sajantila A, Syvanen AC. Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. Genome Res 2001; 11:471–482.
28. Lindroos K, Liljedahl U, Raitio M, Syvanen AC. Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. Nucleic Acids Res 2001; 29:e69.
29. Malmqvist K, Kahan T, Edner M, Held C, Hagg A, Lind L, Muller-Brunotte R, Nystrom F, Ohman KP, Osbakken MD, Ostergern J. Regression of left ventricular hypertrophy in human hypertension with irbesartan. J Hypertens 2001; 19:1167–1176.
30. Kurland L, Liljedahl U, Karlsson J, Kahan T, Malmqvist K, Melhus H, Syvanen AC, Lind L. Angiotensinogen gene polymorphisms: relationship to blood pressure response to antihypertensive treatment. Results from the Swedish Irbesartan Left Ventricular Hypertrophy Investigation vs. Atenolol (SILVHIA) trial. Am J Hypertens 2004; 17:8–13.
31. Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature 1986; 324:163–166.
32. Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 1987; 155:335–350.
33. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 1988; 239:487–491.
34. Shuber AP, Grondin VJ, Klinger KW. A simplified procedure for developing multiplex PCRs. Genome Res 1995; 5:488–493.
35. Hacia JG, Sun B, Hunt N, Edgemon K, Mosbrook D, Robbins C, Fodor SP, Tagle DA, Collins FS. Strategies for mutational analysis of the large multiexon

ATM gene using high-density oligonucleotide arrays. Genome Res 1998; 8:1245–1258.
36. Tonisson N, Zernant J, Kurg A, Pavel H, Slavin G, Roomere H, Meiel A, Hainaut P, Metspalu A. Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. Proc Natl Acad Sci USA 2002; 99: 5503–5508.
37. Fan JB, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. Genome Res 2000; 10:853–860.
38. Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen AC. Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. Nucleic Acids Res 2003; 31:e129.
39. Liljedahl U, Fredriksson M, Dahlgren A, Syvanen A-C. Detecting imbalanced expression of SNP alleles by minisequencing on microarrays. BMC Biotechnol 2004; 4:24.
40. Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, Barany F. Universal DNA microarray method for multiplex detection of low abundance point mutations. J Mol Biol 1999; 292:251–262.
41. Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock probes: circularizing oligonucleotides for localized DNA detection. Science 1994; 265:2085–2088.
42. Banér J, Isaksson A, Waldenstrom E, Jarvius J, Landegren U, Nilsson M. Parallel gene analysis with allele-specific padlock probes and tag microarrays. Nucleic Acids Res 2003; 31:e103.
43. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques 2002:56–61.
44. Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol 2003; 21:673–678.
45. Liljedahl U, Lind L, Kurland L, Berglund L, Kahan T, Syvanen A-C. Single nucleotide polymorphisms in the apolipoprotein B and low density lipoprotein receptor genes affect response to antihypertensive treatment. BMC Cardiovasc Disord 2004; 4:16.

16

MALDI-TOF MS: Applications in Genomics

Martin Beaulieu, Dirk van den Boom, Paul Oeth, Mathias Ehrich, Guy del Mistro, Christiane Honisch, and Charles Cantor

Sequenom Inc., San Diego, California, U.S.A.

1. INTRODUCTION

For the past two decades, dideoxy DNA sequencing has been the method of choice for DNA sequence analysis (1). This invention has revolutionized the field of biological sciences and greatly facilitates applications like pharmacogenomics. A fluorescence-based version of the method allowed the completion of the human genome and the genomes of many other organisms. The availability of these reference sequences offers new opportunities to investigate genetic diversity within and between species. For example, it is now possible to extract information of human medical value by studying inter-individual variations. Genome-wide individual re-sequencing efforts have spurred the demand for novel high-throughput DNA analysis tools. In recent years, the discovery of single nucleotide polymorphisms (SNPs) as the prevailing type of sequence variation, have allowed the conduction of elaborate genome-wide genetic associations analyses at a fraction of the prior cost. It is believed that SNP genotyping will help reveal disease susceptibility genes and predict adverse reactions to drugs.

Mass spectrometry has been used to determine the molecular weight of small labile molecules precisely. The analysis of macromolecules became possible following the development of novel mass spectrometry technologies, especially, Matrix-assisted laser desorption ionization time of flight mass

spectrometry (MALDI-TOF MS) (2). This process combined with delayed extraction time of flight (TOF), allows precise analysis of macromolecules such as DNA, RNA, and polypeptides (3,4). It has been reported that DNA/RNA molecules of up to 100 bases in length can be detected within a few microseconds. This technology breakthrough has allowed MALDI-TOF MS to become the analytical platform of choice for accurate, large-scale DNA analysis without the use of expensive fluorophores or potentially hazardous radioactive isotopes.

This chapter will review recent developments in genomic applications using MALDI-TOF MS. We will discuss genotyping and how multiplexing can be optimized to further improve throughput and reduce cost of genotyping SNPs. We will describe how MALDI-TOF MS can be used to perform quantitative analyses allowing DNA pool based association studies and gene expression analyses. We will discuss comparative sequence analysis applications using base-specific fragmentation of nucleic acids such as mutation detection, pathogen typing, methylation analysis, and molecular haplotyping.

1.1. The MALDI-TOF MS Process

In MALDI MS, the sample is mixed with an excess of a small organic compound called the matrix. Typically, for nucleic acids, the optimal matrix is an organic acid capable of forming homogeneous crystals with the analyte on a solid target support. The target can be introduced into the vacuum of the mass spectrometer's ion source. The co-crystals are irradiated with a UV laser beam forming a particle cloud called "the plume." Within the plume it is believed that the matrix ionizes the analyte molecules by proton transfer. The resulting ions are accelerated by a short high-voltage pulse and drift into a field-free flight tube (1–2 m long) and reach the detector. The analyte mass is calculated by measuring the TOF. The TOF increase proportionally with the mass-to-charge ratio, m/z. Since single-charged and non-fragmented ions are predominantly generated, the signals in a spectrum can be interpreted directly as masses in Daltons. There is no need to deconvolute complex spectra as required in other types of mass spectrometry. The entire process takes less than a millisecond (Fig. 1).

As with other analytical platforms, there are some drawbacks associated with the use of MALDI-TOF MS for DNA analysis. Nucleic acids are susceptible to adduct formation and fragmentation. Their negatively charged backbone makes them behave as polyanions in aqueous solutions. In the presence of cations they tend to form adducts. Most enzymes used in biochemical assays require cations as co-factors, and the predominant ones are sodium (23 Da), magnesium (24 Da), and potassium (39 Da). For example, a sodium ion would add 22 Da to the mass of the analyte, since it would replace a proton (1 Da). Therefore, if adducts are not removed prior to analysis, a distribution of several signals may be obtained, resulting in lower

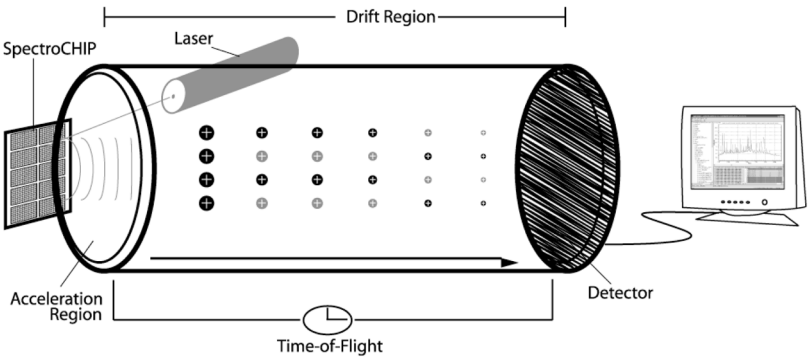


Figure 1 Schematic representation of a MALDI-TOF mass spectrometer.

sensitivity and accuracy of mass determination (Fig. 2). To circumvent this issue, a solid phase (5) and a homogeneous cation exchange resin method (6,7) have been developed. These methods can be automated, and that will be the topic of further discussion in Section 2 of this chapter.

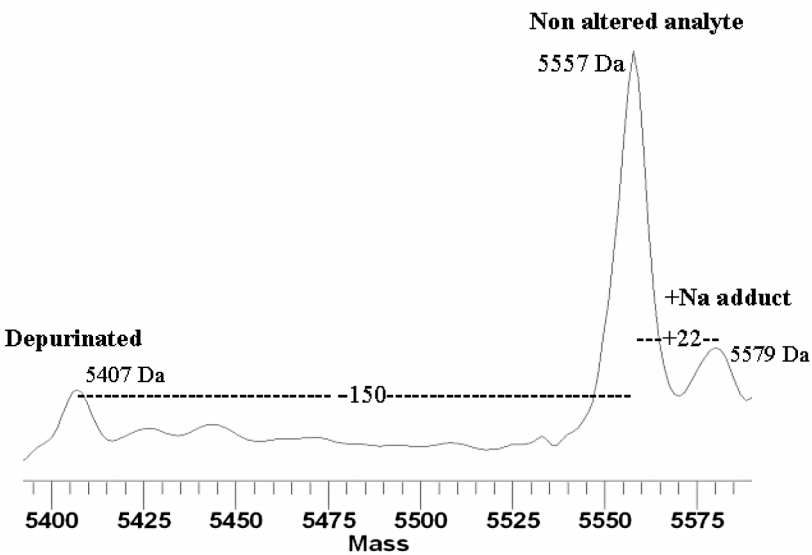


Figure 2 Oligonucleotide artifacts detectable on MALDI-TOF MS. The spectrum shows a typical example of an oligonucleotide of poor quality. The presence of a sodium adduct peak is depicted (Na adduct) and evidence of depurination potentially at G nucleobases (Depurinated). The oligonucleotide sequence used was: 5′ OH-TAGGCCTGGGAAGCAGCA-OH 3′.

During the MALDI process, nucleic acids can be fragmented through depurination. This normally occurs at A or G nucleobases following protonization at position N7 and at C bases following protonization at position N3. The weakened *N*-glycosidic linkage gets rearranged leading to backbone cleavage at the 3′carbon–oxygen bond (Fig. 2) (8). In comparison, RNA is more stable under MALDI conditions than DNA. It is believed that the presence of the 2′hydroxy group in the ribose sugar moiety convey some resistance to depurination (9). Short oligonucleotides (<30 mers) are usually stable under soft laser energy conditions. Most MALDI-TOF MS instruments currently manufactured can be optimized to reduce DNA fragmentation.

In biochemical assays, additives such as detergents, DMSO, urea, BSA, and glycerol are commonly used to improve reaction performances and enzyme stability. However, these additives also act as crystallization disturbing agents preventing the formation of optimal crystals for the MALDI process. Analytical sensitivity and mass accuracy can be affected. The challenge is to develop bioassays that can perform optimally without crystallization disturbing additives. Often, it is necessary to use elaborate purification processes prior to analysis.

The MALDI-TOF MS has several desirable features to accommodate fast and accurate nucleic acid analysis. In MS, the measurement is based on the intrinsic property of the molecule. The exact molecular mass of a nucleic acid molecule can be identified without labels. Typically nucleotide fragments in a mass range between 1000 Da (~3 bases) and 10,000 Da (~30 bases) can be resolved down to ~15 Da separation. Two fragments can differ by only a single nucleotide substitution and still be distinguishable in the mass spectrum. This resolution permits the analysis of several macromolecules at once allowing multiplexing without differential labeling. Multiplexing assay reactions generally decreases the cost per data point while increasing analytical throughput.

The topic of nucleic acid analysis by MALDI-TOF MS has recently been reviewed elsewhere (10,11). The reader can refer to those reviews for an exhaustive description of the evolution of mass spectrometry in nucleic acids analysis. This review will emphasize the new trends of using MALDI-TOF MS for specific applications in molecular biology and pharmacogenomics.

2. SNP GENOTYPING BY PRIMER EXTENSION

SNP genotyping can be achieved accurately using MALDI-TOF MS (5,12–16). A high-throughput automated SNP genotyping platform, MassARRAY™, was developed by Sequenom Inc., allowing the analysis of thousands of SNPs on a Daily basis (5,16–18). The MassEXTEND® assay, used in conjunction with the MassARRAY system, combines oligo base extension with MALDI-TOF MS. The assay consists of a post-PCR

primer extension reaction that is carried out in the presence of one or more dideoxynucleotides (ddNTPs) generating allele-specific termination extension products. Primers are designed to anneal adjacent to the SNPs of interest. Depending on the SNP identity, extension products of different lengths and masses are amplified. In the case of heterozygosity, two mass distinguishable products are generated. The DNA fragments are then purified and analyzed by MALDI-TOF MS (Fig. 3). This assay has a very high specificity because of the two successive amplification steps, each exhibiting locus-specific primer annealing. The PCR primers are designed to anneal to sequences that do not overlap with the primer used for the subsequent primer extension step. When combined with the high analytical accuracy of MALDI-TOF MS, high-throughput genotyping can be performed.

2.1. Solid-Phase Assays

Most solid phases approaches necessitate the labeling of the PCR products with an affinity tag such as biotin. The biotinylated amplification products are captured using strepavidin-coated magnetic beads in the presence of a magnet. Denaturation and wash steps are performed to generate a single-strand template without remaining unused nucleotides and other contaminants. The primer extension reaction is allowed in the presence of specific combinations of deoxy/dideoxy nucleotides. The resulting primer-extension products are washed to eliminate enzyme reagents and salts, and then denatured to remove them from the immobilized template. The supernatant is analyzed by MALDI-TOF MS (5). The advantage of a solid-phase assay is that washing on the magnetic beads can purify the primer-extension products. The disadvantages are: (1) added cost of streptavidin beads, (2) throughput compromised by an extra purification step, and (3) limited amounts of the diagnostic products that can be processed in a vial, reducing multiplexing capabilities.

2.2. Homogeneous Assays

The homogeneous MassEXTEND reaction (hME) is a simplified version of the assay in which the solid-phase purification step is replaced by a simple enzymatic treatment coupled with a cation exchange process (18,19). The reaction and the sample conditioning are performed in the same well. Following PCR amplification, a shrimp alkaline phosphatase enzyme (SAP) is added and incubated at 37°C for 15 min. The SAP is used to deactivate the remaining nucleotides from the PCR reaction. Following the primer extension step, the samples are diluted with water and cation exchange beads (NH_4^+ form) are added (9). After a brief centrifugation, the supernatant is used for MALDI-TOF MS analysis. The simplified protocol offers similar performances at lower cost and workload (Fig. 3).

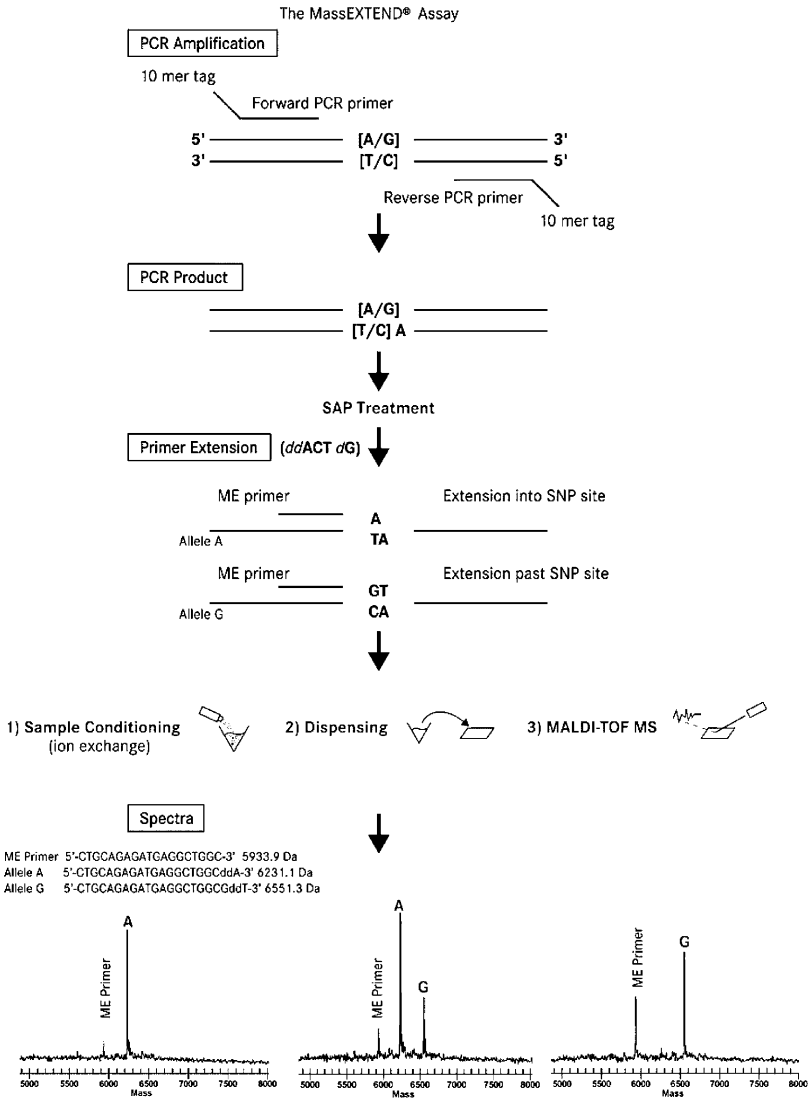


Figure 3 The MassEXTEND® reaction. Following PCR amplification of the locus of interest, a primer-extension reaction is performed using an extend primer (ME) that is designed to anneal next to the SNP. The key feature of this scheme is the use of a terminator mixture that yields extension products that differ in length in an allele-specific manner, thus creating mass separations between alleles equal to the mass of a nucleotide. In this example, a normal dG base is used along with ddA, ddC and ddT. For the A allele (A), a ddA is incorporated, extending the primer to 20 mer. For the G allele (G), the SNP calls for incorporation of the normal dG residue prior to incorporation of a ddT, extending the 19-mer primer to 21-mer.

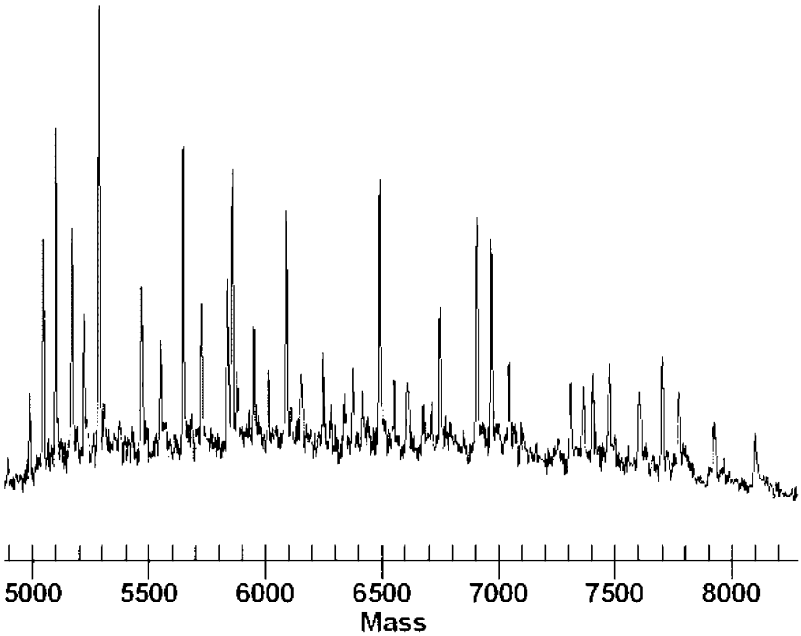


Figure 4 Example of a spectrum showing MassEXTEND® genotyping products at a 15-plex multiplexing level.

2.3. Multiplexing the MassEXTEND® Assay

Performing multiple reactions in a single well is a way to increase analysis throughput and reduce the cost per genotype. Several assays for different SNPs can be combined. The typical mass range of primer-extension products is between 5000 and 9000 Da, corresponding to 17–30 N in length. This provides a wide window for multiplexing since extension primers targeting up to 20 different SNPs can be combined so that primers do not overlap in the mass spectrum (Fig. 4).

2.4. Assay Design

When multiplexing primer-extension-based genotyping assays such as the MassEXTEND® reaction, two main considerations must be made before mixing several primers together: (1) primers for the PCR amplification reaction must be designed to avoid cross-locus amplification. (2) Primers for the extension reaction must be designed so the resulting combination of peak masses in the spectrum is resolvable. MassEXTEND® multiplex assays can be designed with a computer aided assay design algorithm such as the MassARRAY® Assay Designer. For PCR primer design, the following

guidelines are recommended: (1) target primer length of 20 mer, (2) target melting temperature (T_m) of 60°C [based on the 4 + 2 rule (20)], and (3) target G–C content of ~50%, and (4) target amplicon length of ~100 bp. Potential primers must be searched against genomic databases so that no assay will fail due to competing kinetics or cross-hybridization reactions with other loci. The SNPs with known adjacent polymorphisms within six bases should be avoided. Similarly, SNPs lying within low complexity regions should be discarded.

For extend primer design there are only two choices of primer sequence, adjacent to the SNP site on either side (i.e., on either strand). It is rare for SNP targets to fail in assay design on both sides of the SNP. For designing hME extend primers consider the following: (1) optimal length of 17–26 mer, with a minimum T_m of 45°C (by nearest-neighbor calculation, Ref. 21), (2) primers may not contain uncertain bases in the target sequence, and (3) disregard primers with low Tm, hairpin formation, false priming, and primer–dimer potentials. Avoid problematic sequence repeats such as GGGG and mass conflicting by-products such as depurination products and possible extend-pausing products. For example, an extension primer prematurely terminated with a dA base would have exactly the same mass as if normally terminated with ddG base, and therefore, these products would be indistinguishable in the mass spectrum.

In multiplexing, several oligonucleotides are involved. The analyte peaks in the mass spectrum for one assay must be sufficiently resolved from products of any other assay it is combined with. This includes pausing peaks. In addition, analyte peaks must fall within a specified mass window (preferably 5000–8500 Da). Using these considerations, 15 assays can be combined (Fig. 4).

Another way to simplify combination of assays is to use the pinpoint strategy (13,15,22). In this scheme all reactions are performed using all four ddNTPs. Thus, extension through the SNP is not allowed. Although this scheme offers greater simplicity and higher plexing potentials, currently available MALDI instruments do not have the necessary resolution. The difference in mass between ddATP and ddTTP is only 9 Da. Pinpoint designed assays, if used with automated generic conditions, are prone to peak misinterpretation and genotyping errors. Figure 5 shows a spectrum from an attempt to resolve oligomer analytes with 11 Da mass differences. Allowing one allele to be extended through the SNP, by using one to three deoxynucleotides, as in the MassEXTEND assay, is a conservative strategy allowing high genotype accuracy (Fig. 3).

2.5. Automated Genotype Calling

The automated process of producing genotype calls from MALDI-TOF mass spectra can be broken down into three stages: (1) collection of good

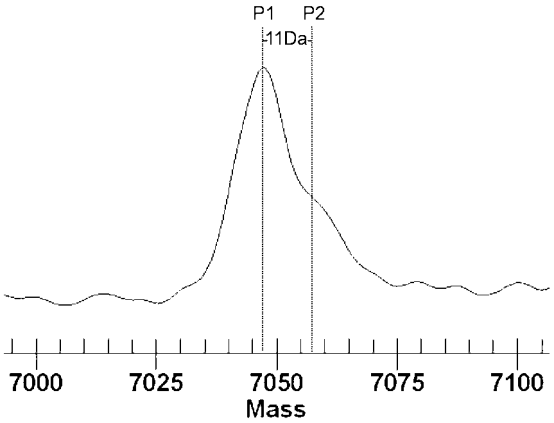


Figure 5 Axial MALDI-TOF MS spectrum showing an attempt to resolve two oligomers (P1 and P2) with only 11 Da difference in mass. Using this type of instrument, the peaks of two oligomers with a single nucleotide difference involving A/T substitution would not be resolvable, therefore pinpoint genotyping is not practical.

quality spectra; (2) analysis of the signals (peak fitting); (3) interpretation of the signals into genotype calls.

There are many factors that can affect MALDI-TOF spectra quality. Even for well-calibrated systems, with perfect chemistry using ideal assays, variations in matrix/analyte crystal quality can lead to mass shifts of several Daltons. Therefore, it is important to calibrate individual spectra. Genotyping calls are made on an averaged spectrum that is generated using a self-weighted average at every individual mass measurement. This averaging brings out the strongest signals across all rasters rather than focusing on any individual raster or analyte signal.

Since only approximate baseline and peak fitting is appropriate for crude signal detection, the ratio of signal intensity to the variance of intensity in the local baseline may be referred to as the signal-to-noise ratio (SNR). Recalibration begins with identifying all the local signal intensity maxima that have an SNR greater than a cutoff (e.g., Ref. 4). Here a noise threshold (0.2) can be added to the measured variance to ensure that noise spikes in very low intensity spectra are not detected as signals. The MALDI-TOF MS spectra are usually recorded as an (averaged) intensity measurement against a linear time (TOF) scale and three quadratic factors (ABC) that define the time-to-mass conversion equation ($\text{mass} = AT^2 + BT + C$). The ABC values can be obtained from calibration with control analytes of known masses (calibrants). It is recommended to calibrate the instrument prior to each analytical run, a process that can be automated. The signals can then be aligned with expected genotyping peaks.

Given a calibrated spectrum, signal analysis can be performed by fitting Gaussian functions to regions of the spectrum, where genotyping masses are expected. Before peak fitting can be attempted a noise baseline must be determined. On spectra of multiplexed reactions, the baseline can be found by using a running average over a mass window of 400–500 Da for the regions of the spectrum, where signals are not expected.

For signals fitted to Gaussian functions, various peak properties must be determined. Peaks height, line width, area, shape, and offset to the expected signal mass; are the most relevant properties to consider before making a genotype call. Calling genotypes is a matter of identifying, which analyte peaks are truly present. This process can be automated using sophisticated probability calculations that consider the peak properties described above.

2.6. PCR Optimization for Multiplexing

In order to perform successful high-level multiplexed reactions, the biochemical experimental conditions must be optimized so that different amplicons are amplified uniformly. Commonly used PCR additives such as DMSO, BSA, detergents, urea, and glycerol should be avoided due to their disturbing effects on matrix crystallization during the MALDI process. The PCR amplification conditions can be optimized using a statistical design of experiments approach such as the Taguchi methods (23). Orthogonal arrays were used to estimate interactions and effects of different ratios of reagents such as Taq enzyme, nucleotides, magnesium, primers, monoionic salts, and template DNA. Results showed that an increased concentration of PCR buffer to 2×, was a positive factor in overall amplification performances over 84 assays, 12-plexed (Table 1, column A). However, following a primer extension reaction and MALDI-TOF MS analysis, the new PCR conditions showed detrimental effects on the genotyping performance (data not shown). It was found that the increased monoionic salt concentration had adverse effects on Thermo Sequenase enzyme (TS) activity during the

Table 1 Comparison of PCR Conditions Changes for High-Level Multiplexing MassEXTEND

	Previous[a]	*A*	*B*
HotStarTaq PCR buffer	1×	2×	1.25×
dNTPs (μM)	200	500	500
$MgCl_2$ (mM)	2.5	3.5	3.5
PCR Primers (nM)	50	100	100
HotStarTaq (U/rxt)	0.1	0.15	0.15

[a]Previous conditions as described in Ref. 16. The conditions *A* and *B* show only reagents that incurred a change. The cycling conditions were the same as previously described (16).

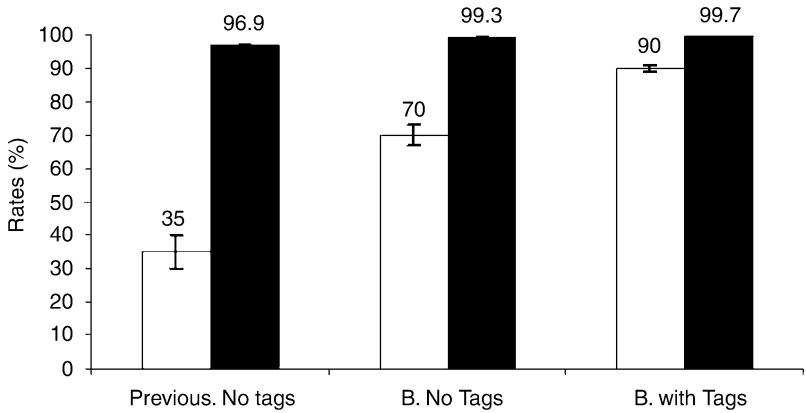


Figure 6 Performance of 12-plexed MassEXTEND® reactions under different conditions.

primer extension step of the MassEXTEND® reaction (24). Our group found that lowering the PCR buffer concentration to 1.25× allows for the best amplification performance for both Taq and TS enzymes (Fig. 6). It is also advantageous to increase the concentration of nucleotides and the Taq enzyme. Adjusting magnesium chloride concentration is important for multiplexing PCR reactions since it is an essential co-factor for the Taq enzyme. It is important to adjust the final concentration of "free" Mg^{2+} ions ([$MgCl_2$]-[total dNTPs]). When altering nucleotide concentration, magnesium concentration must be adjusted accordingly. We found that the optimal free Mg^{2+} concentration is ~1.5 mM. One of the most significant factor contributing to increased PCR amplification efficiencies in multiplexing, was the addition of a 10-mer tag on the 5′end of the PCR primers (5′-ACGTTGGATG-3′). The effect of the tag is additive and contributed to a further ~30% improvement in genotyping success at a 12-plexed multiplexing level (Fig. 6).

2.7. Primer Extension Optimization for Multiplexing

Because the increased amounts of analytes in high-level multiplexes generate spectra of higher complexity, the SNR are often lower. Variations in analyte peak intensity may stem from inconsistent oligonucleotide quality and unpredictable differences in desorption/ionization behavior in MALDI. For these reasons, it is beneficial to normalize primer cocktails prior to processing assays. This is done best by measuring samples of the primer mixtures by MALDI-TOF MS. Each peak within the mass spectra must have comparable heights. Generally, if any primer produces a peak with less than 50% of the height of the highest peak, simply add more of the weak primers

to the cocktail until the resulting spectrum shows even peak intensities (Fig. 7). Omitting to take this precaution can result in genotyping failure and is the main cause of genotyping errors. The most common error that can occur is loss of heterozygosity. For example, a heterozygous SNP extended at a marginal level in combination with a low input of its extend primer to the reaction, will almost certainly result in an homozygous genotype call. Lower mass molecules tend to fly better in MALDI-TOF MS and therefore the sensitivity of the high mass alleles is relatively lower. Marginal extension efficiencies coupled with low input extend primer may result in just below detection limits for the high mass allele producing erroneous homozygous calls.

Multiplexing is challenging to the MALDI-TOF MS process. It is speculated that in multiplex assays, containing over 30 analyte molecules, a competition for available ions in the MALDI process occurs. We found that this effect can be at least partly alleviated by increasing the concentration of extend primers in the reaction. For low-plexed assays, it is recommended to use 0.5 μM of each primer whereas, for high-plexed assays, 1 μM is appropriate.

In the case of the TS enzyme, we found that the optimal concentration to use correlates with the plex level. For low-level plexing (<5) 0.6 Units are sufficient (Amersham Biotech Inc.). For 8–12-plex reactions, 1.25 Units is

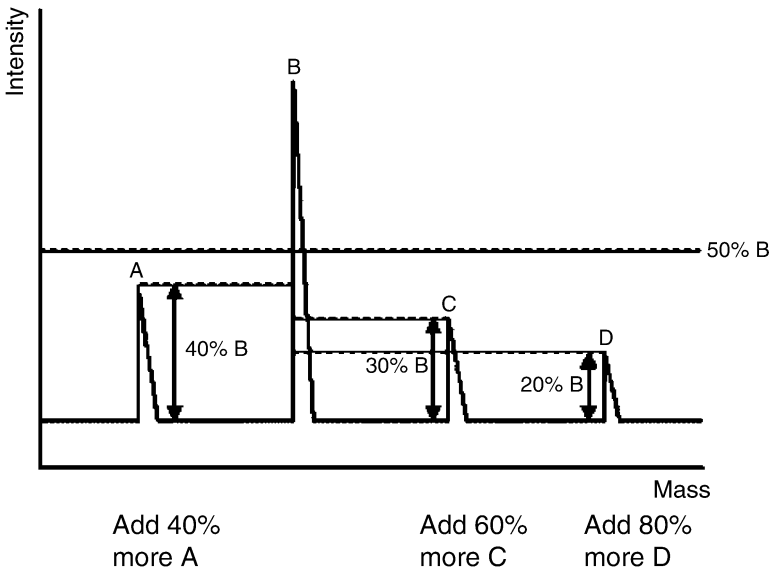


Figure 7 This diagram is a simplified representation of a MALDI-TOF MS spectrum. Relative concentrations of MassEXTEND® primers in assay cocktails can be adjusted to even out primer peak heights. The fractions to add are indicated.

sufficient and 1.5 Units is recommended when 15 plexes are attempted. The number of extension cycles during the primer extension step has a significant impact on genotyping performance in multiplexing. While 45 cycles are sufficient for single-plex reactions, 55 cycles is the minimum for low-level plexing and 75 cycles are recommended for plexing levels of over eight. It has been postulated that the extra cycles allow weaker assays to continue to extend once the strong assays have fully extended their primers.

2.8. Performance

A study conducted by the Whitehead Institute Center for Genome Research, using the MassEXTEND reaction coupled with an early version of a genotype caller algorithm, found an estimated accuracy rate of ~99.6% (25). When using optimized biochemistry combined with the latest caller algorithm as described earlier, it is possible to increase the accuracy rate to 99.9%. The MALDI-TOF MS is sensitive and precise. Nearly perfect call accuracy can be achieved with proper sample handling, spectra interpretation and assay optimization. Performance variability between multiplexes can be attributed to unpredictable behavior of individual assays when combined. Out of 84 assays used in a recent study, 10 exhibited significantly weaker extension rates as multiplexes compared to the uniplex format. Those assays generally provided lower calling rates and were prone to generate errors. Errors can also stem from biased amplification of some SNPs, an effect observed in single assays but enhanced in multiplexing. Some heterozygous SNPs can be called homozygous by lack of a detectable second allele peak. Because most of the error events do not happen randomly, problematic assays may be detectable by clustering SNRs (26) and by Hardy–Weinberg equilibrium analyses (25). These statistical tools can be useful in detecting problematic assays. Individual genotypes can be corrected or rejected. Selected assays can be filtered out from analysis and further experiments. As with other genotyping technologies, the largest contribution to assay genotyping failure is random PCR dropout.

2.9. Throughput

The MALDI-TOF MS provides high resolution, high accuracy, and wide mass range allowing the design of highly multiplexed genotyping assays. Twelve-fold multiplexing can be routinely carried out using generic experimental conditions, as described here and 15–20-fold is possible with specific assay optimization. Routine 12-fold multiplexing with ~30 minutes acquisition/real-time analysis of 384 elements, translates into an analytical speed of >150 genotypes per minutes (9000 genotypes per hour). The MALDI-TOF MS throughput coupled with a cost of below 10 per genotype, makes whole-genome scans feasible and affordable.

3. QUANTITATIVE NUCLEIC ACIDS ANALYSIS USING MALDI-TOF MS

Many important applications in genomics, especially those in the field of pharmacogenomics, require quantitative measurements in order to obtain useful Data (i.e., comparative gene expression analysis to follow the response of individuals to a particular therapeutic compound/regiment). Quantitative measurements using MALDI-TOF MS require at least two analyte peaks. This is due to the fact that peak characteristics (peak height, SNR, and peak area) from individual spectra have no direct relation to the spectra of a separate sample. Because of this, comparing a single peak's measurable characteristics from separate samples for quantitative purposes is arbitrary in the absence of a reference. The establishment of this limitation has enabled the proper use of MALDI-TOF MS in conjunction with primer extension assays for quantitative purposes as long as two or more analytes are present (27). Table 2 lists many of the applications currently being conducted, and many more are sure to emerge. As Table 2 shows, two primary approaches to quantitation are used. The first involves the measurement of two unknown nucleic acid analytes in the same reaction with known

Table 2 Possible Quantitative Applications Using MALDI-TOF MS

Quantitative method	Assay	Application
Ratio of two analytes of unknown concentration	Allele-specific quantitation	Disease association studies, allele-specific expression, and drug resistance mutations
	Allele determination in polyploid genomes	Agricultural genetics
	Gene copy number	Monosomy, trisomy, and transgenic animals
	Loss of heterozygosity	Cancer diagnostics
	Loss of imprinting	Cancer diagnostics
	Viral typing	Vaccine QC treatment monitoring
Ratio of unknown analyte(s) to internal standard of known concentration	Standard gene expression, allele-specific gene expression	Quantitative gene expression
	Quantitative PCR	Multiple applications
	Gene transfer estimations	Gene therapy
	Gene duplication/ multiplication	Genetic diagnostics
	Viral/bacterial titering	Pathogen quantitation/vaccine QC treatment monitoring

mass but unknown concentration relative to each other. Allele frequency estimation in pooled populations of nucleic acids for the purpose of disease association studies uses such an approach, as does differential allele-specific expression. For such applications, the ratio of the two measured alleles is compared between samples to quantitatively determine differences. Relative quantification using this method is achieved by normalizing the baseline of each spectrum, identifying peaks at specified masses (corresponding to expected alleles) and establishing a Gaussian fit to each peak (in the same manner described for genotyping). However, rather than assigning a qualitative call, as is the case with genotyping, the area of each peak is integrated and stored as a numeric value. Within each spectrum, allele frequencies of the primer-extension products are then estimated as the ratio of the area value of one allelic peak to the total area of all of the expected allelic peaks. Since one is dealing with populations of molecules in this model the sum of the frequencies for the alleles under investigation is one (100%). The second approach requires the addition of an internal standard oligonucleotide of known sequence, mass and concentration to all reactions. Unknown analyte concentrations are determined by titrating the internal standard over a range of concentrations such that the point at which the unknown analyte(s) and the internal standard oligonucleotide are at a 1:1 ratio, is observed (based on peak areas of primer-extension products). Exact molar concentrations can then be assigned to the unknown analyte via regression analysis of the plotted peak area ratios for each sample. A specific concentration can then be compared between samples. This approach is successfully used for the measurement of gene expression levels.

3.1. Allele Frequency Estimation

It is well established that the ratio of analyte products for nucleic acids and proteins is proportional to the ratio of peak areas for any given spectrum on a MALDI-TOF MS (28–30). These studies have established that allele frequencies in complex mixtures of DNA have a limit of detection (LOD) to 2% and a limit of quantitation (LOQ) of 5% for minor allele frequencies using MALDI-TOF MS. Extensive validation by our group has led to the use of these findings as defining parameters for large-scale association studies and genome-wide scans. Figure 8 shows a scatter plot of estimated allele frequencies vs. genotyped allele frequencies for 24-primer extension assays distributed randomly throughout the human genome. Ninety-six individual genomic DNAs were first genotyped for all assays to establish the exact allele frequencies in this population. The DNAs were then pooled at an equimolar ratio to test the accuracy of estimating allele frequencies relative to genotyped frequencies. Each assay was conducted in quadruplicate and SD of 2% or less was achieved for the majority of assays as shown in the scatter plots. The coefficient of determination (R^2) in Figure 1 exceeds

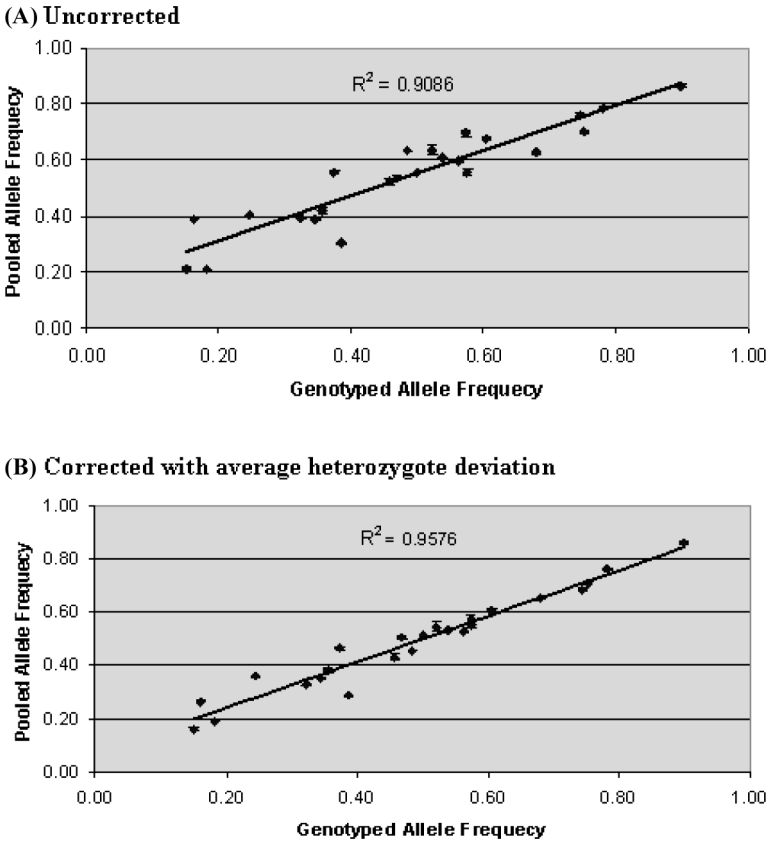


Figure 8 Allele frequency determination using MALDI-TOF MS. (**A**) Scatter plot of genotyped population allele frequencies (*x*-axis) versus allele frequencies calculated using pooled populations DNAs (*y*-axis). Twenty-four unique assays are depicted. The DNA population consisted of 96 individual DNAs at equimolar concentrations (260 pg per individual at 25 ng/uL). Frequencies were calculated using TYPER RT software (SEQUENOM). The calculated allele frequency with standard deviation for each assay represents the average of four replicate reactions each dispensed in replicates of four onto silicon chip arrays loaded with matrix (SpectroCHIP, SEQUENOM). For genotyped frequencies, each of the 96 individual DNAs was genotyped for each of the 24 assays using the MassARRAY™ system (SEQUENOM). Best-fit line and coefficient of determination (R^2) were calculated using Excel 2000 (Microsoft). (**B**) Scatter plot of genotyped population allele frequencies (*x*-axis) versus allele frequencies calculated using pooled population DNAs (*y*-axis) as described in A. The pooled allele frequencies have been corrected for of the 24 assays using the average heterozygote deviation from a 1:1 peak area for each allele as described in the text. Note the improvement in the coefficient of determination (R^2) after correction with individual heterozygote allele ratios relative to uncorrected in A.

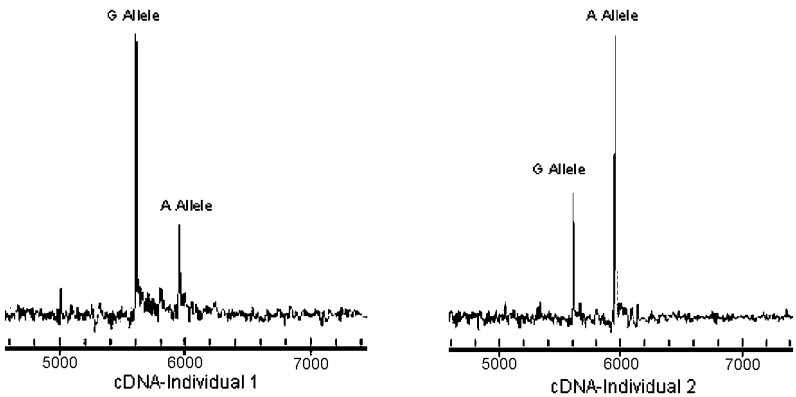
0.90 indicating excellent, but not perfect, correlation between the allele frequencies calculated from the pooled template vs. the genotyped ones. These differences arise from multiple issues such as uneven amplification of alleles during PCR and primer extension reactions that occur for all technologies relying on these processes (31). Other MALDI-TOF MS specific issues such as uneven analyte/matrix cocrystallization and preferential analyte desorption/ionization between alleles can also contribute to this phenomenon. These effects, however, appear to have a minimal contribution (<2% inaccuracy) based on oligonucleotide titrations simulating primer-extension products from 0% to 100% frequency for two alleles (32). The summed effect of all inaccuracies can be measured using individual heterozygous samples for each assay under investigation. Heterozygotes have an allele ratio of 1:1 (one allele on each chromosome) and should therefore exhibit a 1:1 ratio of allele frequencies (0.5:0.5). Any deviation from this expected ratio could be quantified in the same manner as described for pooled templates. The average frequency for each allele from multiple heterozygotes in a population for a particular assay therefore represents the summed effect of any inaccuracies for that assay. The calculated deviation can then be applied to the pooled result as a correction factor (27). Figure 8B shows the 24-primer extension assays from Figure 8A with the corrected allele frequencies for the pooled DNA results. The coefficient of determination (R^2) improves to better than 0.95 as a result of this correction. Multiple studies have compared the quantitative abilities of various platforms used for estimating allele frequencies in populations of nucleic acid molecules. MALDI-TOF MS-based measurement of primer extension reactions have been shown to be as accurate, sensitive and reproducible as all available technologies according to these studies (31,33,34).

The ability of a MALDI-TOF MS-based platform to determine allele frequencies in pooled nucleic acid populations for the purpose of conducting complex disease association studies has been investigated by several groups including us (35–37), and is indeed a viable solution to streamlining large studies which would otherwise require many thousands to millions of genotypes per association study, which is cost and time ineffective. Such studies are of great importance in pharmacogenomics for the identification of complex diseases associated with genetic loci for which therapeutics may be developed. Sequenom and other groups such as the NIH have used this pooling strategy in conjunction with MALDI-TOF to successfully associate specific genes with disease phenotypes (38,39).

Many other applications have been carried out using semi-quantitative primer extension based MALDI-TOF MS analysis with unknown analyte concentrations. One example involves the quantification of differential allele-specific expression. Medelian inheritance of allele-specific expression has been shown in humans (40). Allele-specific expression is of great importance for the identification of genetic components, which lead to population

responses due to environmental and/or administered compounds or conditions and result in disease. Individuals heterozygous for a SNP in an exon of a gene are candidates for this type of differential expression. The ratio of alleles in genomic DNA from all heterozygous individuals should be 1:1. However, in allele-specific expression it is observed that this ratio deviates in cDNA generated from specific transcripts from different individuals. This difference can be assayed with primer-extension products using cDNA as template in PCR and comparing the ratios of the two alleles in the MALDI-TOF MS spectrum. The data must be normalized with the genomic DNA allelic ratio as described above to account for any assay-specific biases. Figure 9A shows spectra from for the SNP rs1801174 (629T/C) in the TP73 gene from two individuals who demonstrate this phenomenon. The ratios of A–G have been corrected using the assay-specific bias observed

(A)



(B)

Individual	G Allele Frequency	A Allele Frequency	Ratio of G to A Allele
1	0.79	0.21	3.8
2	0.32	0.68	0.5

Figure 9 Representative spectra for two individuals showing allele-specific expression as assayed for the SNP rs1801174 (629T/C) in cDNA from the *TP73* gene. In normal heterozygous individuals the frequency of the G:A alleles is 50:50(1:1). (**A**) Here, the two alleles are highly skewed relative to each other indicating differential allele-specific expression. Such data must be normalized for any inherent assay-specific biases by measuring the frequency of each allele in genomic DNA, which will not show any differences by definition. If any bias is observed it must be used to normalize the data obtained using the cDNA template. (**B**) The data in the table below the spectra shows the relative frequencies of the G and A alleles and their ratios after normalization to the genomic DNA template results.

from the genomic DNA (41). Allele-specific expression analyses can also be conducted in the presence of the internal standard oligonucleotide (as described in Sec. 3.2) so as to determine the number of transcripts represented by each allele as opposed to the ratio between the two as shown in Figure 9B.

Allele-specific differences between regulatory polymorphisms associated with the ability of RNA polymerase II to bind and assemble its transcription complex at the start site of transcription for several eukaryotic promoters has also been measured using MALDI-TOF MS coupled with primer extension (42). This technique provides a powerful tool for identifying important regulatory SNPs and haplotypes *in vivo*.

3.2. Gene Expression Analysis

The quantitative measurement of mRNA transcripts using primer-extension products and MALDI-TOF MS, in the absence of allele-specific expression, requires a slightly different approach then that of allele frequency analysis as discussed above. The majority of transcripts are assayed for one allelic form and are therefore not useful for the comparison of two or more alleles. Ding and Cantor (43) published a solution for this problem. It works by introducing a synthetic template that varies from the wild-type reverse transcribed mRNA (cDNA) by a single base over a 60 to 90 bp oligonucleotide. This oligonucleotide template acts as a competitor in the PCR amplification step and the resulting product contains a differentiating allele. This new allele can be used to distinguish the wild-type cDNA template from synthetic template. Since the synthetic template is of known concentration it can be titrated and used as an internal standard to measure the amount of a particular cDNA. This combination of competitive PCR coupled with primer-extension works efficiently. Detection of cDNA amount is independent of PCR cycle number since the two templates rely on the same primers and only differ by one base. Primer-extension reactions, which distinguish the cDNA from the internal standard, also rely on the same primer. Titrating the internal standard over a range of concentrations and conducting regression analysis of the plotted peak area ratios yields the point at which the cDNA and the internal standard alleles are at a 1:1 ratio. This point represents the concentration of cDNA present from a particular tissue or sample.

Multiplexing is possible with this system allowing for the analysis of multiple transcripts expressed at different levels to be assayed in the same reaction. Figure 10 shows a schematic of how a multiplexed reaction is carried out using the method discussed above for three genes: *GAPDH*, *CXCR4*, and *HMBS*. The three genes are assayed in the same reaction (triplex). Figure 11 shows typical results for a logarithmic titration of the internal standard for these three transcripts against a mix of cDNA representing major tissue/organ systems of humans. Notice that the point at which the *y*-axis crosses 0.50 (point at which cDNA and standard are at

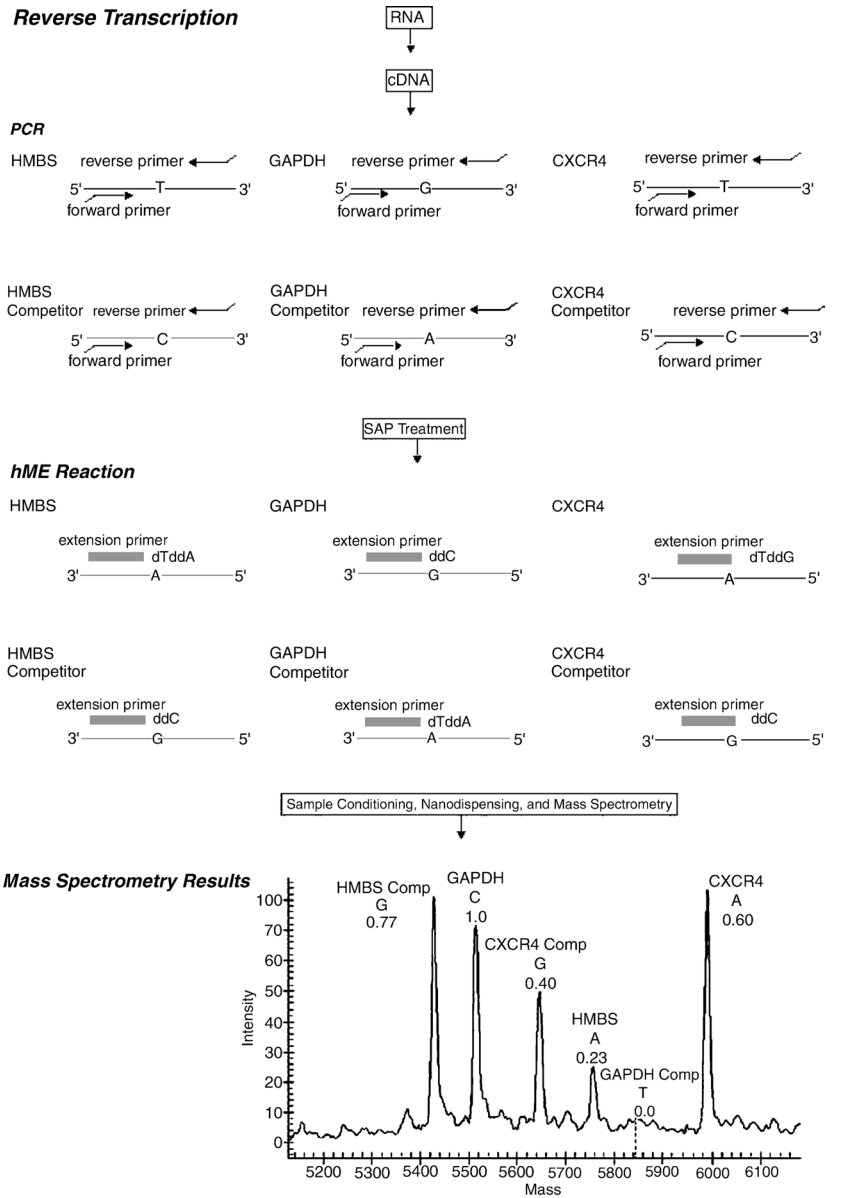


Figure 10 Procedure for multiplexed quantitative gene expression analysis using competitive PCR coupled with MALDI-TOF MS (MassARRAY system). Primers for three transcripts (CXCR4, HMBS, and GAPDH) to be measured simultaneously are depicted. Figure 11 illustrates the results for such an experiment.

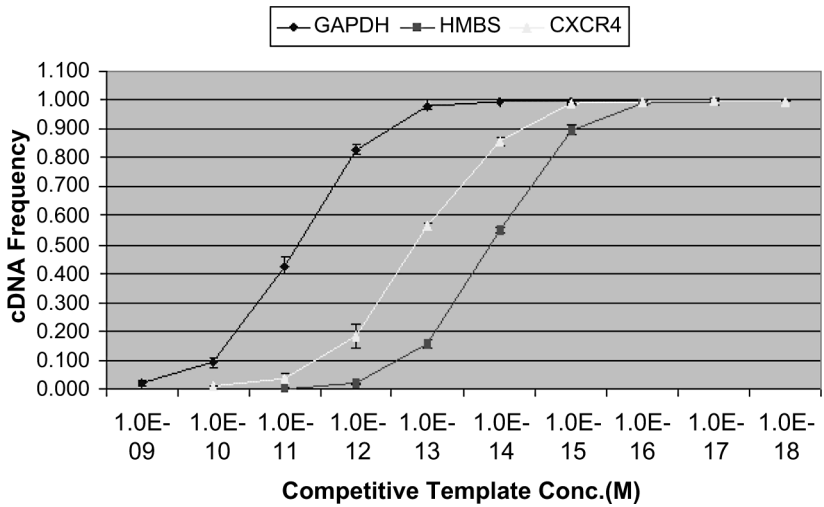


Figure 11 Multiplexing gene expression analysis using MALDI-TOF MS. Line graph plotting the results of nine-log range titration of competitive template (internal standard) to determine relative gene expression levels for three genes: Chemokine (c-x-c motif) receptor 4 (CXCR4), Glyceraldehyde 3 phosphate dehydrogenase (GAP DH) and Hydroxymethylbilane (HMBS). Input cDNA amount was constant for all titration points. The point at which each titration curve crosses the *y*-axis at 0.50 represents the concentration at which a 1:1 ratio of cDNA to internal standard alleles have been observed and therefore represents the concentration of cDNA in that sample. Note that each gene shows an increase in cDNA frequency as the internal standard (competitor) concentration is decreased logarithmically until only cDNA template products are measured on the MALDI-TOF-MS.

1:1 ratio) for the three genes is 2 to 3 logs apart. These results match those of uniplex reactions for each transcript. Results also correlate with those obtained from real time PCR assays (44). The possibility of including gene(s) for sample normalization for each reaction via multiplexing makes this methodology attractive for expression studies needing the throughput and precision of MALDI-TOF MS.

4. COMPARATIVE SEQUENCE ANALYSIS USING BASE-SPECIFIC CLEAVAGE AND MALDI-TOF MS

During the early days of the Human Genome Project, mass spectrometry was proposed as one of the methods potentially being able to accelerate the sequencing efforts. The initial idea combined Sanger sequencing with analysis of the sequencing ladder by MALDI-TOF MS. Replacing the rather slow and laborious process of gel electrophoresis with mass spectrometry readout would have significantly accelerated the process. The use of

molecular weight information for determining the sequence instead of electrophoretic mobility was also thought to increase accuracy.

A decade later, the Human Genome Project has been completed without the use of mass spectrometry. The main obstacle for the use of MALDI-TOF MS was the relatively short sequence read length. Normally, the read length still is rarely longer than 20 bases. This creates significant hurdles in genome assembly and in the handling of repeat regions.

The completion of the Human Genome Project, however, does not imply that there remains no further need for sequencing. Quite to the contrary, the availability of a genetic blueprint allows for large-scale genetic variability studies. This requires efficient sequencing technologies.

In the course of these developments, new methods for MALDI-TOF MS-based sequencing have been implemented. These methods rely on the generation of short oligonucleotide fragments by complete, base-specific endonucleolytic cleavage. Each cleavage product features at least one defined terminal base. This concept resembles the original Sanger method for RNA sequencing. Base-specific cleavage is analogous to the peptide mapping used for protein sequencing by mass spectrometry, because mass signals are matched to amino acid compositions and sequence strings are reconstructed from overlapping compositions. The degradation of a DNA analyte into short oligonucleotide fragments for nucleic acid sequence analysis avoids some of the mass accuracy, resolution and sensitivity issues, which are inherent to primer extension methods. Several methods have been developed to achieve base-specific cleavage. These include enzymatic or chemical DNA-based cleavage methods (45–48) as well as methods using RNA transcription and specific RNAses (49–52). The efficient use of base-specific cleavage and MALDI-TOF MS requires the a priori knowledge of a validated reference sequence and cannot be used for de novo sequencing. Nevertheless, it can be applied to experimental questions for DNA-based identification or re-sequencing, because in these applications an experimentally determined sequence is cross-compared to a known reference sequence. The following sections describe how base-specific cleavage can be applied to answer biological questions in the context of genotype–phenotype correlations.

4.1. SNP Discovery Using MALDI-TOF MS, MassCLEAVE™

The first fully implemented re-sequencing method, Sequenom's MassCLEAVE, specializes in the discovery of SNPs in known target regions. The method relies on base-specific endonucleolytic cleavage of single stranded RNA into short oligonucleotide fragments. Each such fragment represents a subsection of the target sequence and can be analyzed using MALDI-TOF MS. From alterations in the masses of fragments it is possible to determine the corresponding sequence changes (51). The fragmentation products from four base-specific cleavage

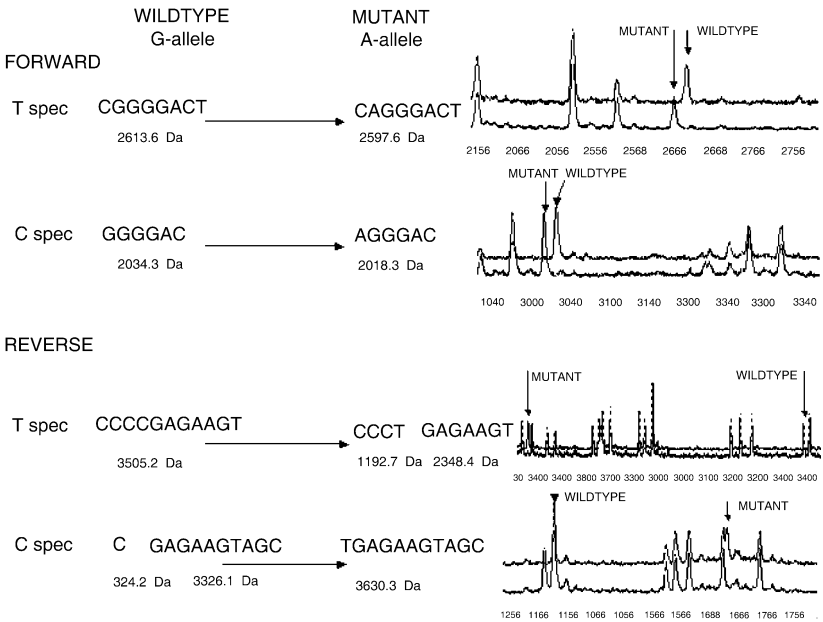


Figure 12 Example of four cleavage reactions obtained from MassCLEAVE reactions. The combination of four cleavage spectra allows usage of multiple observations to confirm the presence of a SNP. Here a substitution G/A SNP from chromosome 5 q23.3 region is used to demonstrate changes in the calculated cleavage products including their masses for the G-allele (wild-type) and the A-allele (mutant), on the left and the corresponding mass spectra, on the right. A G/A substitution results in a 16 Da mass shift in C-specific and T-specific forward cleavage reactions, which represents the mass difference from a G to an A nucleotide. In the cleavage reactions from the reverse strand this substitution is represented as a C/T change. Hence, the polymorphism, shown here, will either introduce a cleavage base (T specific), or remove a cleavage base (C specific). Introduction of a cleavage base (here T-specific reverse) results in fragmentation of one cleavage product into two. Respectively, when a cleavage base is removed (here C-specific reverse) two cleavage products get connected resulting in one fragment.

reactions can be compared and interpreted into accurate sequence information (Fig. 12).

The first step of this process involves PCR amplification of the target region. The use of a primer that carries a T7 promoter sequence at its 5′ end enables incorporation of a transcription start site in the PCR product. This allows for in vitro transcription of the PCR product into a single-stranded RNA molecule in a subsequent step. Endonucleolytic cleavage of the RNA molecule is then performed by adding a ribonuclease. RNase A cleaves pyrimidine residues (Cytosine and Uracil) and therefore it is not

sufficiently base specific. To achieve base specificity non-cleavable nucleotides have to be incorporated during transcription. A mutant T7 RNA polymerase allows for incorporation of non-canonical nucleotides such as dNTPs during in vitro transcription. The incorporation of deoxy cytosine nucleotide (dC) results in base-specific cleavage at uracil residues (U) and incorporation of deoxy thymine nucleotide results in cleavage at ribo cytosine residues (rC). Base-specific cleavage of the other two bases ribo adenine residues (rA) and ribo guanine residues (rG) are addressed indirectly. Tagging the 5′ end of the reverse gene-specific primer with the T7 promoter sequence results in a RNA transcript that represents the reverse strand. Base-specific cleavage after rC on the reverse strand is virtually the same as cleaving the forward strand after G. Thus, the combination of four reactions cleaving at rC and U forward and rC and U reverse allows for the equivalent of base-specific cleavage of all four bases (49).

The mixture of cleavage products generated in each of the four base-specific reactions is analyzed by MALDI-TOF MS. Each cleavage product in this mixture has a distinct mass. This mass is defined by the nucleotide composition (the numerical appearance of the nucleotides A, C, G, and T). With any given reference sequence the expected cleavage products and hence the expected mass signals can be calculated in-silico. Sequence variations in the target regions, such as SNPs, will affect one or more cleavage products. The mass of the affected cleavage products will differ from their expected mass. These alterations in the resulting mass spectrum are then utilized to perform automated SNP discovery. In a first step all new mass signals are collected and their nucleotide composition is calculated from their mass. In a second step a specialized algorithm finds and collects all possible SNPs, which explain the occurrence of the detected nucleotide compositions. Finally all explanations are verified and scored using supporting evidence from the other three cleavage reactions. This leads to a final score that indicates the likelihood of a truly observed sequence variation.

The combination of four cleavage reactions increases the sensitivity of base-specific cleavage for detection of sequence variations. Simulations revealed that about 98% of all SNPs could be detected from amplicon lengths of up to 800 bp (53). The exact detection rate depends on the length of the target regions as well as the sequence context. Longer amplicons yield more cleavage products and therefore increase the likelihood that newly generated signals are masked by preexisting signals and compromise their detection.

Performing SNP discovery by using base-specific cleavage and MALDI-TOF MS offers multiple advantages. The use of molecular weight information provides unambiguous results that cannot be misinterpreted. Also the compounded information of four cleavage reactions allows the verification of each discovered sequence change with up to 10 observations. This redundancy of information coupled with accuracy and speed of

MALDI-TOF MS acquisition is a unique combination that will play an increasingly important role in the discovery of novel sequence variations.

4.2. Signature Sequence Identification Using MALDI-TOF MS

A human fingerprint is the characteristic of the unique skin pattern of the fingertip. The first molecular method for species differentiation and typing became available in 1985. Geneticist Jeffreys et al. (54) introduced the term *DNA-fingerprinting* to describe species-specific DNA fragment patterns obtained from electrophoretic separation and multi-locus probe detection. Pulsed-field-gel electrophoresis, high-density oligonucleotide arrays and Sanger sequencing further extended the panel of available technologies. The ultimate methodology would be amenable to standardization and allow electronic data portability, reproducibility, redundancy, and high processing speed from minute amounts of starting material. Especially for non-culturable cellular species, PCR-based technologies are favorable. Sequencing is currently the most accurate standardized methodology offering acceptable redundancy rates since both strands can be sequenced, although de novo sequencing is not of prime importance when considering species identification.

The redundancy of the four cleavage reactions of MassCLEAVE provides a solution for reproducible species-specific fingerprinting. Characteristic mass signal patterns can be utilized to identify and differentiate microbes at the genus, species or strain-specific level and thus allow pathogen identification. Infectious diseases are an increasingly important area of concern in today's public health environment. The throughput, accuracy and specificity of MALDI-TOF MS are ideally suited for large-scale testing and global data comparability. Software algorithms can compare the peak pattern of the sample to a database set of reference pattern and deliver the homology with the best fit. Base-specific cleavage and MALDI-TOF MS was successfully used to detect signature sequences of Bordetella strains (46,51) and mycobacteria (55) including pathogens such as *Mycobacterium tuberculosis*, a microbe that causes more than 2-million deaths per year worldwide. Twenty-four clinical isolates were identified.

Figure 13 exemplifies the identification of four mycobacterium species on the basis of their mass signal fingerprints of the 16S rRNA hyper variable regions. The C-specific cleavage of the forward RNA-transcript generates unique identifier signals differentiating the species unambiguously.

4.3. Methylation Analysis Using MALDI-TOF MS

The analysis of epigenetic changes, mainly the covalent addition of methyl groups to the 5′ position of Cytosine in CpG sequences, has become an increasingly important field in genetic research. The DNA methylation has

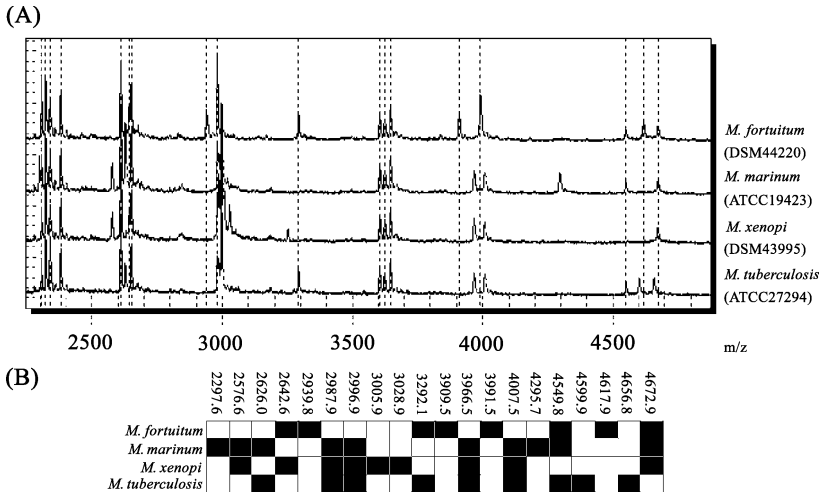


Figure 13 (**A**) Overlay of mass spectra of the 16S rDNA hypervariable regions of *M. tuberculosis* ATCC27294, *M. xenopi* DSM43995, *M. marinum* ATCC19423 and *M. fortuitum* DSM44220 using MassCLEAVE. Mass signals reflect C-specific cleavage products of the forward transcript (**B**) Barcode of in silico clevage fragments of the C-specific forward transcipt.

been found to play a central role in cancer development, genetic imprinting, X chromosomal inactivation, suppression of parasitic DNA and other aspects of gene silencing. Unlike other fields of genetic research there is no "gold standard" method available for DNA methylation analysis. Current methods have limitations in sensitivity, throughput, and the ability to produce results in a quantitative manner (56). Base-specific cleavage and MALDI-TOF MS MassCLEAVE can overcome these challenges.

The assay concept begins with a bisulphate treatment of genomic DNA. During this procedure all non-methylated Cytosine bases are converted to uracil residues, whereas methylated Cytosine remains unaffected (57). Bases substitutions can be detected by MALDI-TOF MS through a MassCLEAVE process (refer to Sec. 4.1). The most apparent change in the mass spectrum can be found in the C-specific cleavage reaction from the forward transcript. Since all the Cytosine residues from a non-methylated template get converted to uracil, no cleavage occurs. Thus, the resulting mass spectrum shows no mass signals other than those resulting from the introduction of cleavage sites within the primer tags (Fig. 14). On the other hand, methylated Cytosine residues are protected from conversion. The corresponding transcription products thus contain cleavage sites and generate corresponding mass signal patterns after C-specific cleavage (for more details refer the MassCLEAVE process section above). These obvious

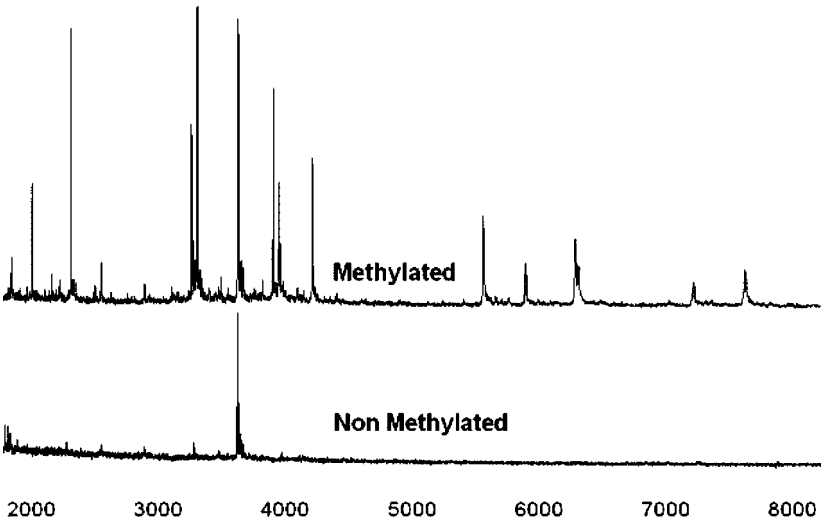


Figure 14 Methylation analysis using MALDI-TOF MS. The spectra shown here were obtained following a MassCLEAVE assay from a methylated DNA sample (*top*) and a non-methylated DNA sample (*bottom*). After bisulphate treatment, the cytosine residues from the Non-methylated DNA are substituted to uracil residues. Following a Cytosine-specific cleavage reaction, the mass spectrum generated from the non-methylated DNA template shows a single peak as expected. The Cytosine residues from the methylated DNA sample are not affected by the bisulfate treatment and the resulting mass spectrum shows expected fragmentation products.

differences in mass spectra allow for a clear segregation of methylated and non-methylated template and enable the discovery of genomic methylation from a single reaction. Because of this clear segregation, it is possible to detect down to 5% methylated DNA in a mixture of methylated and non-methylated sample DNA. However, the C-specific forward reaction alone does not allow quantification, because any amount of methylated DNA will result in the same mass signal pattern. The assessment of methylation ratios is aided by one of the reverse reactions where both methylated and non-methylated DNA species are represented by distinct mass signals. In the reverse reactions, the complementary strand is transcribed to RNA and methylated Cytosine is represented by the incorporation of rG while non-methylated Cytosine corresponds to a rA. The mass difference between rG and rA is 16 Da. Therefore, the mass difference of a cleavage product representing methylated DNA and its equivalent from non-methylated DNA is 16 Da or a multiple thereof, depending on the number of CpGs within the fragment. The close proximity of mass signals facilitates the comparison of their intensities in order to achieve a quantitative readout (Fig. 15). The relative amount of methylated DNA can be determined for

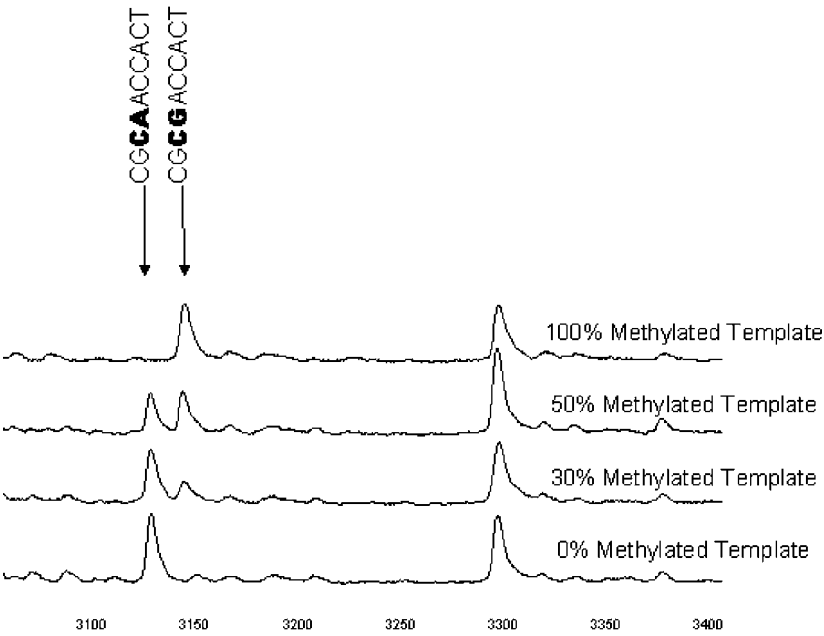


Figure 15 Quantitative methylation analysis using MALDI-TOF MS. Cleavage reactions from the reverse strand allow for determination of methylation ratios. When mixtures of methylated and non-methylated template are analyzed. Typically, in a signal pair one signal represents the methylated and a second nearby signal represents the non-methylated template for defined CpG sites. The ratio of signal intensities reflects the ratio of methylated/non-methylated template DNA.

each CpG individually or it can be averaged for the complete region. Although the reactions described above normally cover $>95\%$ of all CpG sites some applications require the analysis of every CpG in a target region. The aggregate of four base-specific cleavage reactions enables exact methylation pattern analysis. The methylation state for each CpG can be evaluated, and most CpGs would be represented in multiple reactions. This redundancy of information may become an important aspect for diagnostic applications.

The combination of bisulfate treatment and MassCLEAVE makes MALDI-TOF MS a quantitative, high throughput DNA methylation analysis tool that will be useful in future epigenetic research efforts.

4.4. Molecular Haplotyping Using MALDI-TOF MS

Haplotypes are collections of variant allele combinations that can be found in phase on a given chromosome. Phase information can be critical to the

mapping of disease genes. When a disease mutation arises, it does so on a specific chromosomal haplotype. The association between the mutant allele and its ancestral haplotype can be disrupted only by mutation or recombination events in subsequent generations. Thus, it should be easier to track disease mutations through complex haplotypes instead of discrete genotypes. Haplotyping methods have contributed to the identification of genes for Mendelian diseases (58–60), and complex disorders (61–64).

Traditionally, haplotype structure has been deduced by family pedigree analyses. Haplotyping can also be performed through computational analysis of genotyping data at lower cost and labor. However, this offers only haplotype estimations through statistical approximation. Phase determination can be done on a molecular basis on individual patients. Most conventional genotyping methods cannot determine the phase of markers on their own. To obtain molecular haplotypes, one needs to physically isolate a single chromosome prior to genotyping the targeted SNPs. Molecular haplotyping can be done through cost-intensive cloning techniques, by genotyping hemizygous clones. Alternatively, the genotyping of allele-specific PCR (AS-PCR) products is more cost-effective and amenable to automation. AS-PCR can be coupled with MassEXTEND or MassCLEAVE to produce haplotype information. It has also been demonstrated that genotyping PCR products produced from diluting DNA samples down to a single molecule is possible when used in combination with the MassEXTEND reaction and MALDI-TOF MS. Multiplex genotyping of diluted DNA allows molecular haplotype determinations of up to 24 kb in length (65).

Recently, our group demonstrated that it is possible to couple AS-PCR with base-specific cleavage (Fig. 16). The MassCLEAVE reaction can be performed on AS-PCR products so long as an RNA polymerase sequence is incorporated on either PCR primer. Cleavage products can be analyzed by MALDI-TOF MS. Since the reactions are performed on AS-PCR products, the SNPs detected are in phase with one another, which confers the haplotype. This combination offers several advantages: (1) it is possible to interrogate several variations at once; (2) allele discrimination can be controlled internally. Leakage from allele-specific amplification is detectable allowing unambiguous haplotype determination; (3) the discovery of novel variations is a trivial feature with MassCLEAVE, therefore, the analysis is not limited to previously known SNPs; (4) the analysis can be automated and very high throughput performances obtained with the speed of MALDI-TOF MS. Our group applied this method for molecular haplotyping of up to 2 kb long amplicons in single reactions (data not shown).

Potential applications for this method are not limited to haplotyping. This technique can be used to facilitate the characterization of genomic regions with very high polymorphism density. The typing of genomic loci such as the HLA is particularly challenging using conventional genotyping

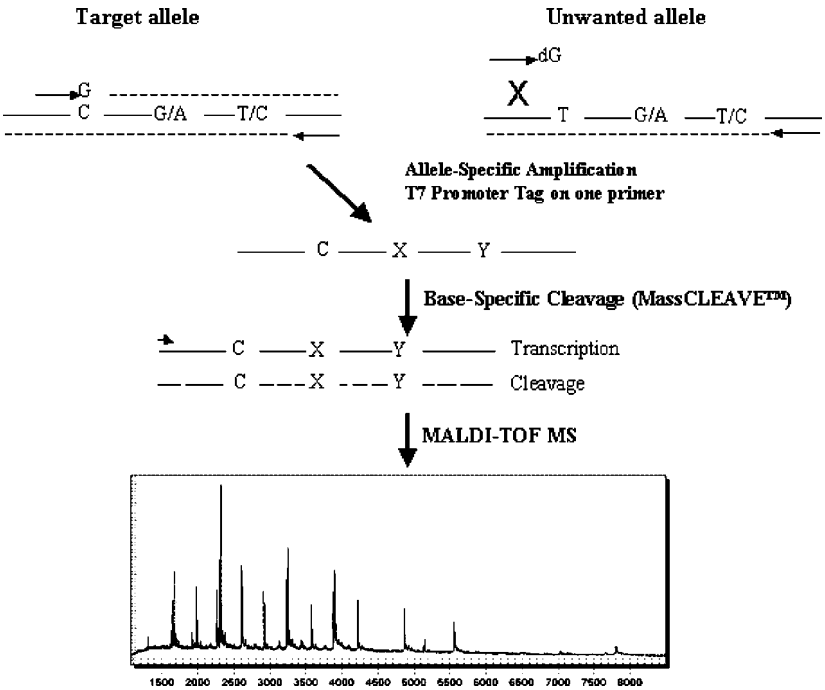


Figure 16 Haplotyping using MALDI-TOF MS. A schematic representation of Allele-specific PCR coupled with MassCLEAVE™ is depicted.

technologies. The combination of AS-PCR with MassCLEAVE may offer a solution.

5. CONCLUSION AND OUTLOOK

Recent developments in MALDI-TOF MS instrumentation such as the implementation of delayed extraction have improved the use of mass spectrometry for nucleic acids analysis. Biochemical assays have adapted to mass spectrometry using homogeneous processes in order to match the analytical speed with suitable processing automation. Software algorithms can interpret mass signals into genomic information automatically at very high speed with real-time control of data quality. The MassARRAY platform includes a MALDI-TOF MS instrument optimized for nucleic acids analysis that can be purchased along with necessary computer algorithms and liquid handling robotics. The suite of applications has extended significantly in recent years. It is currently possible to analyze nucleic acids qualitatively and quantitatively. Comparative sequence analysis using base-specific cleavage allows SNP discovery, mutation detection, pathogen identification,

and methylation analysis. Molecular haplotyping can be done when MassCLEAVE is combined with AS-PCR. This allows the analysis of highly polymorphic loci such as the HLA regions, which are challenging to type using conventional technologies. The vision for the near future is one where even more MALDI-TOF MS applications can be developed to fulfill the needs for the future pharmacogenomics challenges. Applications such as de novo sequencing, multiplexed exon sequencing, methylation haplotype scoring, and drug resistance screening could be developed making MALDI-TOF MS the most versatile analytical platform while maintaining the highest analytical standards. These applications will facilitate drug development efforts and diagnostic analyses to improve individual medicine.

ACKNOWLEDGMENTS

We are grateful to Prof. Dr. Ulf B. Goebel, Universitaetklinkum Charite zu Berlin, Humboldt-Universitaet, Berlin, Germany, for providing us with mycobacteria samples. We are thankful to Dominik Kosman and Nadine Vater for technical assistance, and to Kishor Bhakta, Richard MacDonald and Sebastien Bocker for their efforts on MassARRAY software suite programming.

REFERENCES

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74:5463–5467.
2. Karasand M, Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 Da. Anal Chem 1988; 60:2299–2301.
3. Kaufmann R, Chaurand P, Kirsch D, Spengler B. Post-source decay and delayed extraction in matrix-assisted laser desorption/ionization-reflectron time-of-flight mass spectrometry. Are there trade-offs?. Rapid Commun Mass Spectrom 1996; 10:1199–1208.
4. Juhasz P, Roskey MT, Smirnov IP, Haff LA, Vestal ML, Martin SA. Applications of delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry to oligonucleotide analysis. Anal Chem 1996; 68: 941–946.
5. Braun A, Little DP, Koster H. Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. Clin Chem 1997; 43: 1151–1158.
6. Jurinke C, van den Boom D, Collazo V, Luchow A, Jacob A, Koster H. Recovery of nucleic acids from immobilized biotin–streptavidin complexes using ammonium hydroxide and applications in MALDI-TOF mass spectrometry. Anal Chem 1997; 69:904–910.
7. Nordhoff E, Ingendoh A, Cramer R, Overberg A, Stahl B, Karas M, Hillenkamp F, Crain PF. Matrix-assisted laser desorption/ionization mass

spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared. Rapid Commun Mass Spectrom 1992; 6:771–776.
8. Ono T, Scalf M, Smith LM. 2′-Fluoro modified nucleic acids: polymerase-directed synthesis, properties and stability to analysis by matrix-assisted laser desorption/ionization mass spectrometry. Nucleic Acids Res 1997; 25:4581–4588.
9. Nordhoff E, Cramer R, Karas M, Hillenkamp F, Kirpekar F, Kristiansen K, Roepstorff P. Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry. Nucleic Acids Res 1993; 21:3347–3357.
10. Fitzgeraldand MC, Smith LM. Mass spectrometry of nucleic acids: the promise of matrix-assisted laser desorption-ionization (MALDI) mass spectrometry. Annu Rev Biophys Biomol Struct 1995; 24:117–140.
11. Jurinke C, van den Boom D, Cantor CR, Koster H. The use of MassARRAY technology for high throughput genotyping. Adv Biochem Eng Biotechnol 2002; 77:57–74.
12. Haff LA, Smirnov IP. DNA sequence analysis by matrix-assisted laser desorption ionization MS. Biochem Soc Trans 1996; 24:901–904.
13. Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res 1997; 7:378–388.
14. Little DP, Braun A, Darnhofer-Demar B, Koster H. Identification of apolipoprotein E polymorphisms using temperature cycled primer oligo base extension and mass spectrometry. Eur J Clin Chem Clin Biochem 1997; 35:545–548.
15. Ross P, Hall L, Smirnov I, Haff LA. High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 1998; 16:1347–1351.
16. Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Koster H. Chip-based genotyping by mass spectrometry. Proc Natl Acad Sci USA 1999; 96:10016–10020.
17. Tang K, Opalsky D, Abel K, van den Boom D, Yip P, Mistro G, Braun A, Cantor CR. Single nucleotide polymorphism analyses by MALDI-TOF MS. Int J Mass Spec 2002; 226:37–54.
18. Storm N, Darnhofer-Patel B, van den Boom D, Rodi CP. MALDI-TOF mass spectrometry-based SNP genotyping. Methods Mol Biol 2003; 212:241–262.
19. Rodi CP, Darnhofer-Patel B, Stanssens P, Zabeau M, van den Boom D. A strategy for the rapid discovery of disease markers using the MassARRAY system. Biotechniques 2002; 32(suppl):62–66, 68–69.
20. Breslauer KJ, Frank R, Blocker H, Marky LA. Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci USA 1986; 83:3746–3750.
21. SantaLucia J, Allawi HT Jr, Seneviratne PA. Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry 1996; 35: 3555–3562.
22. Haff LA, Smirnov IP. Multiplex genotyping of PCR products with MassTag-labeled primers. Nucleic Acids Res 1997; 25:3749–3750.
23. Cobb BD, Clarkson JM. A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. Nucleic Acids Res 1994; 22:3801–3805.
24. van den Boom D, Beaulieu M, Oeth P, Roth R, Honisch C, Nelson MR, Jurinke C, Cantor CR. MALDI-TOF MS: a platform technology for genetic discovery. Int J Mass Spec 2004; 238:173–188.

25. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science 2002; 296:2225–2229.
26. Olivier M, Chuang LM, Chang MS, Chen YT, Pei D, Ranade K, de Witte A, Allen J, Tran N, Curb D, Pratt R, Neefs H, de Arruda Indig M, Law S, Neri B, Wang L, Cox DR. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. Nucleic Acids Res 2002; 30:e53.
27. Jurinke C, Oeth P, van den Boom D. MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis. Mol Biotechnol 2004; 26: 147–164.
28. Tang K, Opalsky D, Abel K, van den Boom D, Yip P, Del Mistro G, Braun A, Cantor CR. Single nucleotide polymorphism analyses by MALDI-TOF MS. Int J Mass Spec 2003; 226:37–54.
29. Nelson RW, McLean MA, Hutchens TW. Quantitative determination of proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Anal Chem 1994; 66(9):1408–1415.
30. Ross P, Hall L, Haff LA. Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry. Biotechniques 2000; 29:620–626, 628–629.
31. Le Hellard S, Ballereau SJ, Visscher PM, Torrance HS, Pinson J, Morris SW, Thomson ML, Semple CA, Muir WJ, Blackwood DH, Porteous DJ, Evans KL. SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. Nucleic Acids Res 2002; 30:e74.
32. Tang K, Oeth P, Kammerer S, Denissenko MF, Ekblom J, Jurinke C, van den Boom D, Braun A, Cantor CR. Mining disease susceptibility genes through SNP analyses and expression profiling using MALDI-TOF Mass Spectrometry. J Proteome Res 2004; 3:218–227.
33. Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. Nat Rev Genet 2002; 3:862–871.
34. Shifman S, Pisante-Shalom A, Yakir B, Darvasi A. Quantitative technologies for allele frequency estimation of SNPs in DNA pools. Mol Cell Probes 2002; 16:429–434.
35. Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun A. High-throughput development and characterization of a genome wide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Proc Natl Acad Sci USA 2001; 98:581–584.
36. Mohlke KL, Erdos MR, Scott LJ, Fingerlin TE, Jackson AU, Silander K, Hollstein P, Boehnke M, Collins FS. High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. Proc Natl Acad Sci USA 2002; 99:16928–16933.
37. Werner M, Sych M, Herbon N, Illig T, Konig IR, Wjst M. Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. Hum Mutat 2002; 20:57–64.

38. Bansal A, van den Boom D, Kammerer S, Honisch C, Adam G, Cantor CR, Kleyn P, Braun A. Association testing by DNA pooling: an effective initial screen. Proc Natl Acad Sci USA 2002; 99:16871–16874.
39. Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS. Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. Diabetes 2004; 53:1141–1149.
40. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. Science 2002; 297:1143.
41. Muller S, van den Boom D, Zirkel D, Koster H, Berthold F, Schwab M, Westphal M, Zumkeller W. Retention of imprinting of the human apoptosis-related gene TSSC3 in human brain tumors. Hum Mol Genet 2000; 9:757–763.
42. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat Genet 2003; 33:469–475.
43. Ding C, Cantor CR. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. Proc Natl Acad Sci USA 2003; 100:3059–3064.
44. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol Res 2002; 3:0034.
45. Elso C, Toohey B, Reid GE, Poetter K, Simpson RJ, Foote SJ. Mutation detection using mass spectrometric separation of tiny oligonucleotide fragments. Genome Res 2002; 12:1428–1433.
46. von Wintzingerode F, Bocker S, Schlotelburg C, Chiu NH, Storm N, Jurinke C, Cantor CR, Gobel UB, van den Boom D. Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: a tool for rapid bacterial identification. Proc Natl Acad Sci USA 2002; 99:7039–7044.
47. Shchepinov MS, Denissenko MF, Smylie KJ, Worl RJ, Leppin AL, Cantor CR, Rodi CP. Matrix-induced fragmentation of P3'-N5' phosphoramidate-containing DNA: high-throughput MALDI-TOF analysis of genomic sequence polymorphisms. Nucleic Acids Res 2001; 29:3864–3872.
48. Wolfe JL, Kawate T, Sarracino DA, Zillmann M, Olson J, Stanton VP Jr, Verdine GL. A genotyping strategy based on incorporation and cleavage of chemically modified nucleotides. Proc Natl Acad Sci USA 2002; 99:11073–11078.
49. Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N, Hartmer R, Honisch C, Rodi CP, Bocker S, van den Boom D. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. Genome Res 2004; 14:126–133.
50. Krebs S, Medugorac I, Seichter D, Forster M. RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. Nucleic Acids Res 2003; 31:e37.
51. Hartmer R, Storm N, Boecker S, Rodi CP, Hillenkamp F, Jurinke C, van den Boom D. RNase T1 mediated base-specific cleavage and MALDI-TOF MS for high-throughput comparative sequence analysis. Nucleic Acids Res 2003; 31:e47.

52. Hahner S, Ludemann HC, Kirpekar F, Nordhoff E, Roepstorff P, Galla HJ, Hillenkamp F. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) of endonuclease digests of RNA. Nucleic Acids Res 1997; 25:1957–1964.
53. Bocker S. SNP and mutation discovery using base-specific cleavage and MALDI-TOF mass spectrometry. Bioinformatics 2003; 19(suppl 1):i44–i53.
54. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. Nature 1985; 316:76–79.
55. Lefmann M, Honisch C, Bocker S, Storm N, von Wintzingerode F, Schlotelburg C, Moter A, van den Boom D, Gobel UB. Novel mass spectrometry-based tool for genotypic identification of mycobacteria. J Clin Microbiol 2004; 42:339–346.
56. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 2003; 3:253–266.
57. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994; 22:2990–2997.
58. Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 1992; 2:204–211.
59. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. Science 1989; 245:1073–1080.
60. Puffenberger EG, Hosoda K, Washington SS, Nakao K, deWit D, Yanagisawa M, Chakravart A. A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. Cell 1994; 79:1257–1266.
61. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001; 411: 603–606.
62. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001; 411:599–603.
63. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, Kocher K, Miller K, Guschwan S, Kulbokas EJ, O'Leary S, Winchester E, Dewar K, Green T, Stone V, Chow C, Cohen A, Langelier D, Lapointe G, Gaudet D, Faith J, Branco N, Bull SB, McLeod RS, Griffiths AM, Bitton A, Greenberg GR, Lander ES, Siminovitch KA, Hudson TJ. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 2001; 29:223–228.
64. Vermeire S, Wild G, Kocher K, Cousineau J, Dufresne L, Bitton A, Langelier D, Pare P, Lapointe G, Cohen A, Daly MJ, Rioux JD. CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure. Am J Hum Genet 2002; 71:74–83.
65. Ding C, Cantor CR. Direct molecular haplotyping of long-range genomic DNA with M1-PCR. Proc Natl Acad Sci USA 2003; 100:7449–7453.

17

Gene Expression Analysis in Pharmacogenetics and Pharmacogenomics

Julius V. Halaschek-Wiener, Marco A. Marra, and Angela R. Brooks-Wilson

Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

1. INTRODUCTION

Pharmacogenomics combines genomics ;investigations into the structure or function of very large numbers of genes undertaken in a simultaneous fashion, with the study of phenotypes and traits specifically related to drug responses. Gene expression analysis is an important and widely applied genomics tool. Gene expression analyses, either at the single gene level or by means of whole genome approaches, are integral to our understanding of how drugs work, whether they are effective, and whether a subset of the human population will either fail to benefit from a given drug or experience adverse affects from it.

In this review, we focus primarily on three types of investigations that use gene expression technology to study pharmacogenomics. The first is the study of how human genetic variation affects the expression of genes, particularly genes relevant to drug response. The second topic is the use of gene expression analyses in the optimization of treatment, and the third type of investigation is the use of gene expression analyses in the discovery of pharmacologically important genes not previously implicated in drug response. In each case, we attempt to summarize the major types of

techniques used for these investigations, and provide examples of both single gene (genetic) and multiple gene (genomic) experiments in these areas.

2. HUMAN GENETIC VARIATION AND GENE EXPRESSION

Genetic variation affects gene function in several ways. Nonsynonymous variants in translated regions cause amino acid changes that may alter protein function. Variants in exons or in introns may affect splicing, whilst those in the untranslated regions (UTRs) of genes can affect transcript stability. Finally, variants outside transcribed regions but within regulatory elements may affect the frequency of transcription of a gene and its level of expression, its tissue distribution, or its developmental or temporal expression. Given the ability of expression level changes to cause modest (and therefore developmentally viable) differences in phenotype without complete loss of gene function, differences in the expression levels of genes may be among the most important types of variation for common complex genetic diseases. Indeed, promoter variants have been shown to influence complex traits such as blood pressure (1) and susceptibility to diseases such as rheumatoid arthritis (2). Genetic variants that affect gene regulation may include either single-nucleotide polymorphisms (SNPs) or insertions or deletions (“indels”). The SNPs that affect gene regulation are sometimes referred to as regulatory SNPs or “rSNPs.” Inclusion of such variants in genetic association studies will be important for the assessment of candidate genes with regard to pharmacologically relevant phenotypes and other complex genetic disorders.

Studies of the effect of genetic variation on gene expression have lagged behind studies of variants that alter protein structure or function. This is due in part to the more straightforward interpretation of the effects of many nonsynonymous changes and a lack of effective assays for detecting genetically determined gene expression differences, particularly in the normal chromosomal context of each gene. Furthermore, we lack a full understanding of the sequence elements necessary for the expression and regulation of individual genes or groups of genes. In addition, it is known that sequence elements such as enhancers may be found at relatively large distances (of the order of 100 kb) from the transcriptional start of a gene, making it difficult to define precisely the genomic region containing important regulatory elements. For these and other reasons, variants in the regulatory regions of human genes have been more difficult to study and are now the subject of considerable attention, a key aspect of which is the development of high-throughput assays for the effects of specific genetic variants on gene expression.

2.1. The Effect of Individual Sequence Variants and Haplotypes on Gene Expression and Phenotype

Inter-individual sequence-based variation in gene expression is now considered to be a major cause of phenotypic variation. Pharmacogenetic examples of such effects include, variation in the number of repeats in the promoter region of the thiopurine methyltransferase (TPMT) gene (reviewed by Coulthard and Hall, (3)), which have been shown to affect gene expression *in vitro.* Sensitivity to thiopurine drugs such as immunosuppressants and cytotoxic compounds is affected by the inter-individual genetic variation in this gene.

Alternate splicing also has clear functional effects on genes, many of which express isoforms with distinct functions (reviewed by Bracco and Kearsey, (4)). CYP4F3, for example, hydroxylates the pro-inflammatory molecule leukotriene B4. Alternate splicing of CYP4F3 alters the substrate specificity, tissue distribution, and function of this enzyme (Christmas et al. (5)). Examples of genetic variants that affect splicing are a SNP in codon 242 of the cyclin D1 gene that affects the age of onset of hereditary nonpolyposis colon cancer (6) and a VNTR (variable number of tandem repeats) polymorphism that prevents multiple alternate splicing of the cystathione beta-synthase gene (7).

A particularly well-characterized example of genetic variation affecting gene function, expression, and phenotype is the β2-adrenergic receptor gene (ADRB2). ADRB2 is a G protein coupled receptor for catecholamines (epinephrine and norepinephrine), and a major drug target for the treatment of heart failure and asthma. The functional or phenotypic consequences of several common individual ADRB2 variants are well characterized (reviewed by Liggett (8,9)) including a Gly16 variant that is associated with increased agonist-promoted down-regulation of the receptor and nocturnal asthma; a Glu27 variant markedly associated with obesity; and an Ile164 variant that decreases the coupling, binding, and sequestration of substrate. Ile164, Gly16, and the Gly16/Gln27 combinations are also associated with depressed exercise performance in patients with heart failure (10). *In vitro* expression studies using allele-specific constructs of ADRB2 or employing a reporter gene have shown that a Cys19 variant in the leader peptide of the gene increases expression at the level of protein without affecting mRNA levels.

A seminal paper by Drysdale et al. (11) used 13 SNPs at ADRB2 to characterize 12 common haplotypes that differ in frequency between ethnic groups. They then correlated gene and protein expression, as well as *in vivo* response to albuterol, a β2-adrenergic receptor agonist, with haplotype pairs for a series of individuals, and found that both gene expression and response to the drug were significantly related to specific ADRB2 haplotype pairs. Importantly, ADRB2 haplotype pairs were predictive of drug

response, whereas individual SNP variants were not, showing that unique interactions of multiple SNPs within a haplotype can affect biological and therapeutic phenotypes. If this observation holds true for many other human genes, haplotype analysis will be critical for pharmacogenetics, and pharmacogenomics.

2.2. Cross-Species Conservation: Indirect Evidence of Regulatory Elements

Cross-species comparison of genomic sequence, or comparative genomics, has become a widely used technique for the identification of evolutionarily and presumably functionally important DNA sequence elements. These computational methods use a variety of algorithms, strategies, and programs to align and compare sequences to identify conserved elements (12–15). Analyses such as phylogenetic footprinting (Refs. (16–20). to list just a few) and phylogenetic shadowing (21) have been used to computationally identify the binding specificities of transcription factors. DNA sequence variants located within conserved or other computationally identified sequences are candidates for rSNPs.

2.3. In Vitro Functional Characterization of SNPs

Functional assessment of individual non-coding SNPs can be done by a variety of methods, including gel-shift assays, DNA footprinting, and allele-specific promoter constructs with reporter gene assays. Although these in vitro assays are valuable methods of assessing the effects of different genetic variants on gene expression, they have several major limitations. One is the difficulty of assessing the consequences of variation in regulatory elements located at a distance from the gene of interest. These methods also do not assess the effects of regulatory sequence variation within the gene's normal chromosomal context in the presence of other genetic and epigenetic effects within a mammalian genome. To overcome these limitations, two methods of assessing variants in their normal chromosomal/chromatin context have been developed. One measures allele-specific transcript abundance (22,23); the other assays allele-specific binding of polymerase II, which is a reflection of the frequency of gene transcription (24).

2.4. Assays of Allele-Specific Transcript Abundance

An effective and elegantly uncomplicated method for directly assessing allele-specific transcript abundance within a normal chromosomal context has been used by several groups (22,23,25) to examine allele-specific expression in human samples. Cowles et al. (26) have used this method to examine such variation between mouse strains. These groups use SNPs in the transcribed regions of genes as tools to measure allele-specific expression levels of genes

of interest, or expression "allelic imbalance" (Ref. 23; reviewed by Olivier (27)). Here, we will use the term "expression allelic imbalance (EAI)" to avoid confusion with tumor allelic imbalance; the latter is generally used to describe events such as loss of heterozygosity in cancers. Allele-specific differences in gene expression are likely due to the presence of either allele-specific *cis*-acting factors physically linked to the transcribed "indicator" SNP, or caused by epigenetic factors such as imprinting. The method does not generally discriminate between imbalances caused by unequal allele-specific transcription, allele-specific splicing, or differential transcript stability.

In this EAI method, single base extension (SBE) genotyping assays are used with either fluorescently labeled (22) or acycloprime-labeled (23) terminator nucleotides in assays using cDNA. Hudson and co-workers (23) validate each assay using genomic DNA, for which robust assays give the expected 50:50 allele ratio in heterozygous individuals. When cDNA from heterozygous lymphoblastoid cell lines from unrelated CEPH (Centre d'Etude du Polymorphisme Humain) individuals was used, however, they found that 23 of 129 inflammation-related genes (18%) deviated from the expected equimolar ratio in one or more individuals tested. The frequency of EAI in different genes varied from 6% to 50% of the informative unrelated individuals tested. Kinzler and co-workers (22) observed EAI in 6 of 13 genes tested, with 3–30% of individuals displaying EAI in different genes. Bray et al. (25) find that roughly half (7 of 15) of genes tested in human brain cDNA show EAI. The Hudson group also demonstrated EAI in heterogeneous nuclear RNA (hnRNA) of three out of three genes tested, demonstrating that the imbalance originated at the level of transcription and, for these three genes at least, was not due to allelic differences in transcript stability.

A group led by Lee and Buetow (28) describes an efficient approach to EAI analysis that uses Affymetrix HuSNP chips to assay for both sample genotype and allele-specific expression level. Of 602 informative (heterozygous) genes expressed in kidney or liver, 54% were found to show EAI in at least one of seven individuals tested.

A subset of genes showing EAI was also tested for segregation of the EAI phenotype in CEPH pedigrees (22,23). Several genes showed transmission patterns of EAI that were compatible with Mendelian inheritance, one with incomplete penetrance of the EAI phenotype. *BTN3A2*, a gene for which EAI was consistently associated with a particular allele in different families, showed co-segregation of the EAI effect with a 15-kb haplotype composed of nine SNPs in three pedigrees as well as in unrelated individuals (23). The indicator SNPs in *BTN3A2* are likely to be in linkage disequilibrium with the rSNP underlying its EAI. Other genes displayed apparently random expression of either one allele or the other in different individuals; this random monoallelic expression that may result from imprinting or other epigenetic phenomena.

Interestingly, most EAIs were bidirectional, with increased expression of a specific allele in some samples, but increased expression of the other allele in other samples. Pastinen et al. (23) noted that this could be explained by allelic heterogeneity of *cis*-acting regulatory variants, epigenetic effects, or incomplete linkage disequilibrium between the indicator SNP and the rSNP underlying the EAI. They point out that the indicator SNPs chosen are on average 34 kb downstream of the start of transcription, a distance that is greater than the average size (20 kb, Ref. 29) of a linkage disequilibrium block.

Several laboratory groups have used EAI assays to demonstrate that allele-specific expression is a widespread and heterogeneous phenomenon, and is therefore likely to have a key role in complex genetic disease. Experimental strategies such as those described above can be adapted for the high-throughput systematic discovery of the effects of rSNPs in whole genomes. Though the method does not identify the causative rSNPs per se, it can provide an invaluable snapshot of the effect of genetic variation on gene expression in the genome. If comprehensively applied across many genes in different tissues, this would be of particular value for genes of pharmacogenetic significance.

2.5. HaploChIP: Allele-Specific Protein–DNA Interactions Identify Putative rSNPs

Knight et al. (Ref. 24; reviewed by Hudson (30)) have developed a novel technique based on chromatin immunoprecipitation (ChIP) that detects the allele-specific biases in the "loading" of phosphorylated RNA polymerase II (Pol II) onto promoter sequences. It is known that transcription begins when Pol II is phosphorylated on specific serine residues and is released from the initiation complex to begin transcript synthesis (31,32). ChIP involves the treatment of cells to cross-link proteins to DNA, followed by fragmentation of the chromatin by sonication. Protein–DNA complexes are then immunoprecipitated with antibodies that recognize DNA binding proteins. In the case of HaploChIP, an antibody specific for particular phosphorylated serine residues of Pol II is used to isolate Pol II-bound DNA fragments. The DNA–protein cross-links are then reversed and the protein digested away. The relative abundance of different alleles of the promoter region of a specific gene in the Pol II-bound DNA fraction is then measured using PCR with primers flanking a nearby SNP followed by primer extension genotyping and detection by quantitative MALDI-TOF mass spectrometry. Knight et al. (24) used their technique to demonstrate haplotype-specific loading of Pol II at the lymphotoxin alpha (LTA) gene.

The HaploChIP approach does not require the presence of an indicator SNP in a transcribed region; SNPs in chromatin outside transcribed

regions can be used to demonstrate and measure allele- or haplotype-specific loading of Pol II, a surrogate measure for gene transcription. This method allows the high-throughput identification of allele-specific transcription for genes that lack transcribed SNPs, but does not determine the particular SNP or SNPs responsible for the allele- or haplotype-specific effect.

Neither EAI nor HaploChIP determine which SNP is causing an effect on gene expression; additional techniques are required for this. As noted by Hudson (30), such variants may be targets for manipulation of expression levels to restore full gene function.

2.6. Gene Expression Profiles as Quantitative Traits (eQTLs)

The experimental approaches described above, though amenable to high-throughput analyses, require a gene-by-gene assessment of expression levels. Another powerful approach marries large-scale microarray-based expression profiling to genome-wide linkage analysis. The method, first termed "genetical genomics" by Jansen and Nap (33,34), is referred to as "expression quantitative trait locus" (eQTL) analysis by Friend and co-workers [Refs. 35 and 36; commentary by Darvasi (37)] who describe its application to the genome-wide genetic analyses of gene expression levels in maize, mouse, and man.

An example of the eQTL approach would be analysis of a series of F2 progeny of a cross between two inbred strains of mice (Fig. 1). The genomes of each of these F2 mice would be a patchwork of blocks of sequence derived from one parental inbred strain or the other. In classical quantitative trait locus (QTL) analysis, genetic linkage analysis is applied to traits such as weight or size that vary quantitatively between mice of such an F2 series, to identify genomic loci that are responsible for that quantitative trait. In eQTL analysis, the quantitative traits used are the gene expression levels from gene expression profiles of the F2 mice. These F2 mice are subject to microsatellite-based genome scanning to provide the genetic data for linkage analysis. If a gene expression chip representing, for example, 10,000 genes was used, this would provide 10,000 quantitative traits to analyze for linkage using the genome scan data, i.e., genome-wide linkage analysis of 10,000 phenotypes!

Schadt et al. (35) analyzed 23,574 profiled genes and found that 7861 (33%) were differentially expressed in the livers of the parental (C57BL/6J and DBA/2J) mouse strains or their F2 progeny. The eQTL mapping revealed 2123 genes with LOD scores (log of the odds in favor of linkage) greater than 4.3, and 965 genes with LOD scores greater than 7.0. eQTL hotspots, where multiple eQTLs clustered in a clearly non-random way, were observed on several mouse chromosomes. Such hotspots are likely to represent loci that affect the expression of multiple genes, or clusters of genes for which expression is under genetic control. An observation that

the expression of many genes was linked to more than one eQTL locus, and that some were linked to three loci, demonstrated that gene expression is a complex trait. Schadt et al. (35) note that the *Nnmt* drug metabolism gene was known to differ in expression between the DBA and B6 parental mouse strains (38) and they also observed a corresponding eQTL (LOD score 15.3) at the position of the *Nnmt* gene. Perhaps most interesting of all was that while many genes' expression levels map to an eQTL corresponding to the position of that gene, others identified eQTLs elsewhere in the genome that are likely to correspond to *trans*-acting genetic factors influencing gene expression.

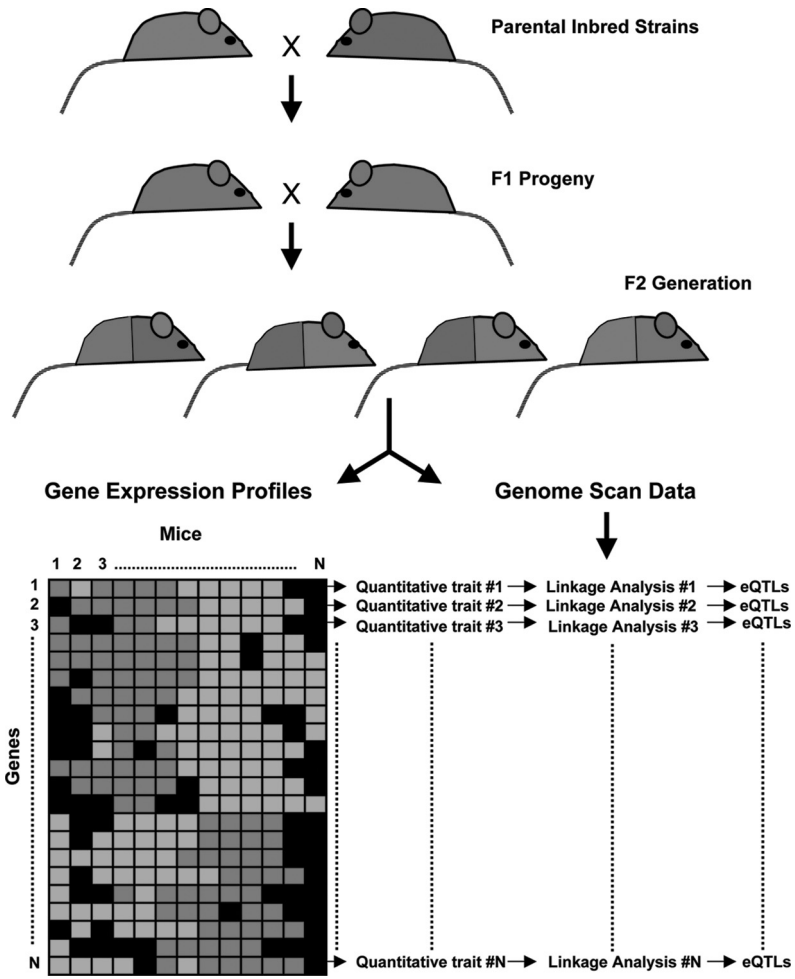


Figure 1 (*Caption on Facing page*)

Schadt et al. (35) also performed initial eQTL analyses using gene expression profiles from the immortalized lymphocytes of members of four CEPH pedigrees. Almost one-third of the differentially expressed genes that were analyzed in this way showed inheritance of their EAI. Given that only 16 "normal" individuals were studied, it is remarkable that genetic effects could be detected in such a high proportion of genes. This result underscores the importance of genetic effects on gene expression in the determination of human phenotypes such as susceptibility to complex diseases and drug responses.

The eQTL analyses in humans, either based on linkage studies of pedigrees or on association studies in case/control groups, are likely to be valuable for those pharmacologically relevant genes that are expressed in accessible tissues such as lymphocytes. For tissues such as liver, which is the site of many steps in the metabolism of various drugs, studies in model organisms like mice will be particularly fruitful for pharmacogenomics research.

Early examples of pharmacogenetics represent Mendelian or near-Mendelian traits. A true picture of pharmacogenomics will be multifactorial and will more closely resemble complex genetic disease. As for other complex phenotypes, studies of drug responses will benefit from new genomic methodologies and resources such as inexpensive genotyping methods and the HapMap project (39). New assays for gene regulatory processes such as EAI and particularly new genome-wide methods for relating expression levels to phenotypes, such as eQTL analyses, will be critical for an under-

Figure 1 (*Facing page*) An example of eQTL analysis based on analyses of F2 mice from crosses of two strains of inbred mice is shown. Homozygous inbred parental strains that differ in phenotypes of interest are indicated in black and light gray. The parental strains are crossed to give F1 progeny (shown in dark gray), which are heterozygous for all markers that differ in the two parental strains. The F1 mice are mated to each other to produce F2 animals that have different compositions of the parental genomes (they are homozygous for one strain's genotypes in some parts of the genome, heterozygous in others, and homozygous for the second strain's markers in other regions); these mice are indicated as a patchwork of black, light gray, and dark gray. The F2 mice also differ with respect to various phenotypes, particularly those that were recognizably different in the parental strains. A tissue of interest is isolated from each of the F2 mice and subjected to gene expression profiling using, for example, Affymetrix arrays. The F2 mice are also subjected to genome scanning using, for example, polymorphic microsatellite markers from throughout the genome. The expression profile of each gene is then treated as a quantitative trait (varying from high to low) in a QTL analysis using the genome-wide genetic marker data. If N genes have been profiled, N linkage analyses are conducted, each of which may result in the identification of eQTL loci with significant LOD scores indicative of linkage.

standing of the genetic basis of drug response and other medically important phenotypes.

3. GENE EXPRESSION ANALYSES FOR TAILORING AND OPTIMIZING DRUG TREATMENT

Gene expression analysis has broad applications in pharmacogenomics including molecular profiling of patients for optimizing drug treatment, selection of target genes, and pathways for drug development and toxicological profiling of novel drugs (e.g., determination of the activation patterns of drug-metabolizing enzymes).

Molecular profiling of tumors and tissues to assess the optimal treatment strategy and to adjust drug dosage prior to drug administration is advantageous, and is available for some therapeutic compounds. Implementation of such measures is expected to lead to a new era of individualized molecular medicine (40–42). Patients in general, and cancer patients in particular, can be profiled by gene expression analysis using advanced high throughput technologies. In particular, advances in DNA microarray technology, high throughput DNA sequencing, and proteomic assays accelerate the generation of high quality patient profiles at ever-decreasing costs. The combination of these methods with sophisticated bioinformatics tools for data analysis has enhanced the basis of drug discovery (identifying targets genes and pathways) and molecular patient profiling (identification of drug responders and non-responders).

It is widely hoped that the broad implementation of DNA microarray technology will revolutionize future pharmacologic investigations. Monitoring gene expression profiles may provide insights into (a) the molecular fingerprints of different diseases including cancer, diseases of the central nervous system, and the cardiovascular system, (b) therapeutic treatments, (c) environmental agents, (d) avoiding toxicity and achieving personalized treatment, and (e) methods of achieving the ultimate goal of preventing these diseases altogether (43).

3.1. High-Throughput Gene Expression Analysis

DNA microarrays are often used to analyze genome-wide changes in gene expression patterns. Either cDNA microarrays or oligonucleotide-based gene chips can be used for gene expression analysis. A cDNA microarray contains a large number of cDNA fragments physically spotted onto glass slides or nylon membranes (44). Alternatively, oligonucleotides corresponding to known genes or ESTs can be synthesized in situ on a miniature matrix using a photolithographic process to produce oligonucleotide-based microarrays (45). Oligonucleotide-based DNA chips can also be used to screen

individuals for DNA mutations and polymorphisms by analyzing variations in genomic DNA (46).

The traditional gene-by-gene approach to assess gene expression induced by pharmaceutical compounds includes Northern blotting of mRNAs, reverse-transcription-coupled polymerase chain reaction (RT-PCR) of mRNAs, Western blotting of proteins, and various enzymatic assays. These techniques, however, allow evaluation of only a few to tens or at most hundreds of genes/gene products per study. With the advent of DNA microarray technology, expression of thousands or tens of thousands of genes can be queried simultaneously, including various drug metabolizing enzyme (DME) genes and a battery of genes that may be relevant pharmacologically (as potential therapeutic targets) or toxicologically (leading to undesirable or toxic effects of the drugs) to drug response in humans. Microarray data have also been used recently to analyze gene expression changes in response to environmental toxins, chemotherapeutic agents, and cytokines (47–49). The use of DNA microarrays to interrogate the gene expression patterns of cells exposed to pharmacologic small molecules and environmental toxins may yield insights into the mechanisms of drug- or chemical-induced toxicity and carcinogenesis.

An important use of microarray-based gene expression analysis is the sub-classification of cancers. Two methods are currently used: supervised and unsupervised analyses. A supervised analysis first defines sample subcategories and then searches for genes that are differentially expressed between these two groups. Unsupervised analysis (or clustering) of expression data identifies gene expression ''clusters'' in phenotypically uncharacterized samples that are significantly related in terms of their expression profiles. Samples that share expression profile features might be expected to share phenotypic features, treatment outcome for example, that can be clearly defined pathologically (50).

Expanding our knowledge of basic biological processes and developing a detailed understanding of the signaling pathways that participate in the development of disease has led to opportunities for diagnostic and therapeutic intervention. Validation of drug targets within patient groups is critical, and for a successful clinical application it is also critical to determine whether the putative target can be identified as present or absent in a consistent, practical, and inexpensive manner.

3.2. Molecular Profiling of Cancers

Examples of specialized molecular therapeutics that have been developed from discoveries made using expression profiling experiments include chimeric human/murine antibodies for the treatment of various cancers. Rituximab (Rituxan®; Genentech Inc.; South San Francisco, CA), an anti-CD20 antibody in use since 1997, was the first chimeric monoclonal antibody to

receive FDA approval for treatment of CD20-positive Non-Hodgkin lymphomas. Cells from most Non-Hodgkin lymphomas express this antigen on their surface (51). Shortly thereafter Trastuzumab (Herceptin®; Genentech Inc.), which targets the human epidermal growth factor receptor 2 (HER2) was approved for breast cancer treatment; it is now part of the clinical management of metastatic disease. Further examples of monoclonal antibodies in clinical use include Cetuximab (Erbitux®; ImClone Systems Inc.; New York, NY), which targets the epidermal growth-factor receptor (EGFR) and Bavacizumab (Avastin®; Genentech Inc.), which reacts with vascular endothelial growth factor (VEGF). Both have been approved for treatment of metastatic colorectal cancer and are in clinical testing for use with other types of tumors. A comprehensive review of monoclonal antibodies in cancer therapy was performed by Harris (52).

From a pharmacological viewpoint these molecular therapies change the "one drug fits all" approach that was applied previously. As with a plethora of classical therapeutics, where the actual drug targets and the mode of action are not entirely understood, these new-generation therapeutics are only effective when the target protein is present in the malignant tissue. The specific investigation of one gene product does not require the sophistication of complex microarray analysis. Specific assays or small-customized arrays, consisting of tens rather than thousands of genes or customized real-time quantitative PCR platforms represent realistic alternatives for clinical use.

Examples of diagnostic tests in clinical use in cancer therapy include evaluation of HER2 and EGFR expression levels. The HER2 diagnostic tests marketed by Dako (Copenhagen, Denmark) and Abbott Laboratories (Abbott Park, IL) detect overproduction of HER2 protein and gene amplification, respectively. Dako also markets an FDA approved test kit, EGFR-pharmDx that identifies EGFR-positive tumors and thus helps determine whether a patient can benefit from Cetuximab therapy. These single gene expression-based diagnostic tools are among the first steps towards applied pharmacogenetics/genomics and individualized medicine. Future challenges for pharmacogenomics in clinical settings will be the transformation from single-protein diagnostics to detection of multiple biomarkers and, if necessary, the integration of whole genome analysis.

High-density microarrays allow the parallel investigation of all transcripts in a given cellular system. This global gene expression analysis will advance drug-associated gene and pathway discovery, functional characterization of genes, and tumor sub-classification. Although molecular testing has been increasingly used in clinical practice, the precise diagnosis, and prognosis of most human cancers still relies heavily on descriptive histopathological data. Identification of robust molecular markers associated with distinctive morphological parameters will assist in diagnostic and prognostic assessment. In the past several years, DNA microarray

technology has been widely used in the context of cancer research, resulting in a deluge of new information that can be used to identify molecular alterations common to all tumors of a specific type as well as signatory profiles unique to a subcategory of cancers. Expression-profiling studies have been performed in almost every major type of cancer (53).

Although many of these studies were initially designed for cancer classification, a large proportion of the published data contains information that can be used to identify cancer-specific markers. Rhodes et al. (54) demonstrated a successful model for performing meta-analysis of independent microarray datasets. Analysis of four prostate cancer gene expression data sets (55–58) revealed striking similarities, despite the different sample preparation methods and array platforms used. Results from this and other studies (59–62) have generated a number of promising cancer specific markers that will probably out-perform other prostate cancer markers previously identified. Based on existing well-illustrated examples such as AMACR (α-methylacyl-coA racemase), which is commonly up-regulated in prostate cancer specimens, the complete molecular distinction between normal and cancer cells can be surprisingly easy to achieve, if a combination of a few robust markers is used (63,64).

3.3. Tumor Classification

Sub-classification of tumors originating from a specific tissue holds much promise for accurate cancer diagnosis and individualized treatment. Expression profiles unique to each subtype of cancer can be used for development of subtype specific therapies and for monitoring their therapeutic efficacy. Although advanced cytogenetics and molecular analysis tools have been used in subclassification of some cancers (e.g., acute leukemia), molecular classification based solely on gene expression has proven to be a viable alternative and may provide further information regarding the disease state. Proof of this principle was first illustrated by Golub et al. (65) in a study involving samples from 72 acute leukemia patients. Supervised learning methods based on known identities of acute myelogenous leukemia (AML) and acute lymphocytic leukemia (ALL) generated a profile that was successfully used to classify a new group of samples into the correct category. This study established the feasibility of expression-based tumor classification.

Genetic microarray analysis was also extensively studied to classify and predict prognosis in a series of hematological malignancies (66). Some of the earliest clinical studies that used gene array analysis studied acute leukemia, a disease for which one has access to essentially pure tumor cell populations. An important initial question was whether expression arrays could be used diagnostically to distinguish AML from ALL (67). Using unsupervised learning based solely on gene expression, 36 of 38 AML and

ALL patients were correctly assigned, demonstrating that microarrays can be used to classify acute leukemias without prior clinical knowledge. Using supervised learning, 50 genes were identified that were differentially expressed between AML and ALL. Profiling of these genes in unknown leukemia samples classified 29 of 34 samples correctly. Gene expression signatures have also been defined for prognostic subgroups of diffuse large B-cell lymphoma (DLBCL) that show statistically significant differences in overall survival. The sub-classification into germinal center B-like DLBCL (GC-DLBCL) and activated B-like DLBCL involved gene expression patterns of approximately 3000 genes identifying each group (68). Although the entire patient group showed an overall 52% survival at 5 years, 76% of GC B-like DLBCL patients were alive after 5 years in contrast to only 16% of activated B-like DLBCL patients, suggesting that these two subgroups represent distinct disease entities. Germinal center B-like DLBCL has a more favorable response to standard therapies, whereas activated B-like DLBCL is less responsive to them. More aggressive and/or investigational treatment approaches may be justified for the latter group. This DLBCL subgroup classification resulted from the development of a custom microarray-based prognostic tool, the LymphoChip (69) capable of identifying these two distinct subgroups of DLBCL.

Rituximab, which targets the B cell surface protein CD20, is therapeutically useful for treating follicular lymphoma (FL). A recent study has examined the expression of more than 20,000 genes in tumor samples from patients who received rituximab for treatment of FL in order to identify differentially regulated genes in responding and non-responding patients (70). Based on their expression profiles, the tumors fell into two groups, one clustered with normal lymphoid tissues from spleen and tonsil, and the other with more malignant samples. Rituximab non-responders were found in the normal lymphoid tissue group, whereas rituximab responders were mainly found in the more malignant group.

Other hematological malignancies have also been studied using DNA microarrays, particularly with respect to prognosis, including chronic lymphocytic leukemia (71), mantle cell and marginal zone lymphoma (72), and multiple myeloma (73). All these studies used high-density microarrays in class prediction experiments to identify diagnostic markers and to improve the characterization of and therapeutic strategies for these malignancies.

A key clinical classification of breast cancer tumors is estrogen receptor (ER) expression. The more ERs are present on tumor cells, the more likely an anti-estrogen therapy such as tamoxifen can be successfully applied. Classification of each tumor is important, as only about 60% of breast cancers are ER-positive. Initial studies set out to demonstrate that breast cancers with distinct pathological features could be separated by microarrays. Several groups demonstrated that supervised data analysis could be used to distinguish ER-positive from ER-negative tumors (74–76).

Surprisingly, these studies also demonstrated that the gene expression profiles of ER-positive and ER-negative breast cancers differ both qualitatively and quantitatively in expression of a large number of genes. The fact that ER-positive and ER-negative tumors are different at the level of gene expression suggests that these molecular subtypes are entirely different disease entities that may have arisen from distinct precursor cell types. Furthermore, only a few of the genes that discriminate between ER-positive and ER-negative tumors appear to be part of the ER signaling pathway, adding further weight to the concept of distinct lineages for ER-positive and ER-negative tumors (74). In a landmark study, van't Veer et al. (75) performed gene expression microarray analysis on 78 sporadic lymph-node-negative tumors arising in women under the age of 55 years. Microarray analysis of these primary tumors identified a list of 70 discriminatory genes whose expression patterns, in an internal validation, identified a group of patients who had not developed metastases at 5 years from diagnosis. A method such as this that can identify patients at very low risk of developing metastases allows chemotherapy to be targeted only to those women who are likely to benefit from it.

If accurate determination of chemosensitivity were predictable by gene expression analysis, the overall number of patients receiving cytotoxic treatment unnecessarily would decrease, and the overall survival benefit derived per patient treated would increase accordingly. Large increases in the absolute survival of patients diagnosed with breast cancer, however, will require the development of novel therapeutic agents.

Microarray-based gene expression profiling for gene and pathway discovery, functional classification of genes, and new tumor sub-classifications have also been applied to various other malignancies including lung, bladder and prostate cancer, and cutaneous melanoma (reviewed in Refs. 77–80).

4. IDENTIFICATION OF DRUG TARGETS AND TOXICOLOGICAL PROFILES BY GENE EXPRESSION ANALYSIS

In addition to molecular profiling of diseases to identify drug responders/non-responders and to classify sub-groups within a given disease, large scale gene expression analysis is also broadly used to determine the effects of drugs on target cells or tissues. It is well known that genes control drug effects, but there is increasing evidence that drugs also affect gene function. Accordingly, microarray analysis can be used to monitor gene expression changes that directly reflect the cellular response to drug exposure. This kind of information is valuable in several ways: (a) altered gene expression patterns of drug-metabolizing enzymes highlight specific drug response pathways, (b) the single genes or whole signaling pathways activated represent potential novel drug targets, and (c) toxicological information,

especially induction of apoptosis, stress response genes, and multi drug-resistance genes (toxicogenomics), can be assessed with genome-wide gene expression analyses.

4.1. Gene Expression Analysis of Drug-Treated Cells

Rae et al. (81) used microarrays to measure the effects of rifampin, an antibacterial agent, on the mRNA expression of drug-metabolizing enzymes in human hepatocytes. They showed increases of cytochromes P450 CYPC9, CYP2C8, and CYP3A4 but not of CYP2C18, CYP2E1, or CYP2C9. This cytochrome P450 expression pattern resembles the specific drug metabolizing enzyme profile for rifampin.

To investigate the toxicological profile and to identify novel downstream mediators of tumor cell response, two recent studies investigated MCF-7 breast cancer cells that were treated with different chemotherapeutic agents. Maxwell et al. (82) used DNA microarrays to identify genes that are transcriptionally activated in MCF-7 cells by 5-fluorouracil (5-FU) treatment. They found specific genes that were consistently up-regulated in the treated group (spermine/spermidine acetyl transferase, annexin II, thymosin-β-10, chaperonin-10, and MAT-8). As the major limitation to the clinical use of 5-FU is acquired or inherent resistance, these identified genes may prove valuable in overcoming these limitations.

Kudoh et al. (48) used DNA microarrays to monitor the expression profile of human MCF-7 cells that were either transiently treated with doxorubicin or selected for doxorubicin resistance. They identified a subset of genes (microsomal epoxide hydrolase 1, 26S proteasome regulatory subunit 4, and XRCC1) that were constitutively over-expressed in both the short-term treated and resistant groups and may therefore be functionally relevant for drug resistance. Identification of drug-target genes and the underlying signaling pathways using advanced gene expression technologies can indeed facilitate the development of new therapeutic strategies to improve the efficacy, safety, and tolerance of novel or existing drugs.

In the field of toxicogenomics (i.e., the study of how genomes respond to environmental stressors or toxins) large-scale gene expression analysis of toxin-treated cells and animals has demonstrated its highly accurate capacity to recognize the toxic potential of novel drug candidates (83–86), resulting in an increase in the efficiency, quality, and safety of drug selection in drug development pipelines.

Gunther et al. (87) used human primary neurons treated with multiple classes of antidepressant drugs, antipsychotic drugs, and opioid receptor agonists to generate DNA microarray gene expression data representative of these classes of treatment. They investigated whether gene expression profiles from these drug treatments could be used to construct statistical models capable of predicting drug efficacy. They showed that supervised

classification schemes could be used to accurately predict the functional category of members of each of these drug classes, based on analysis of the drug-induced gene expression profile.

4.2. Gene Expression Analysis of Combination Therapy

Recently, Cheok et al. (88) examined the gene expression profile of a combination therapy. They examined the DNA microarray-based gene expression profile of childhood acute lymphoblastic leukemia bone marrow cells 24 hr after the randomized initiation of patient treatment with mercaptopurine (a purine antagonist), methotrexate (a dihydrofolate reductase inhibitor), or a combination of these two agents. This study addressed two important questions: (a) do mercaptopurine and methotrexate elicit similar changes in gene expression, as they both affect purine metabolism in similar ways, and (b) does the combination of mercaptopurine and methotrexate treatment result in gene expression changes that simply reflect the sum of the two parts? Interestingly, the effects of mercaptopurine and methotrexate when given as single agents were largely non-overlapping. Only 14% of the genes regulated by either mercaptopurine or methotrexate alone were similarly regulated by the combination therapy (88). These data support a model in which the combination does not simply function as a sum of the component parts. Whether this combination treatment represents synergistic action impinging on a single molecular pathway, or modulation of independent pathways that subsequently converge remains to be determined.

4.3. The Reverse Pharmacogenomics Approach

A new twist in the use of gene expression as a drug discovery tool was reported by Stegmaier et al. (89), who devised a new screening strategy for identifying compounds that force the myeloid differentiation of AML cells. Their approach is based on the observation that all-*trans* retinoic acid produces clinical remissions by inducing differentiation of acute promyelocytic leukemias harboring a mutated retinoic acid receptor alpha (90). They developed a surrogate marker approach that uses post-treatment gene expression signatures in AML cell lines as the read-out for screening the ability of candidate compounds to induce differentiation of AML cells. This ''reverse pharmacogenomics'' strategy inverts the classical model of assessing therapeutic agents by first defining the drug-induced pattern of gene expression that is the therapeutic goal, and then screening compounds that can induce this endpoint. In some ways, this can be viewed as a modern, mechanistically insightful reimplementation of classical pharmacology, using a whole-cell response to drive drug discovery. The ultimate tests would be in vivo experiments illustrating that compounds identified by this approach actually induce AML differentiation in preclinical models, and clinical trials.

5. SUMMARY

Gene expression analysis is an important genomic tool that has wide applications in pharmacogenomics. Successes to date include the subclassification of many types of tumors to optimize treatment or prognosis, as well as the use of drug-induced expression profiles in target discovery and toxicity studies. New techniques that enhance our understanding of how human genetic variation affects gene expression contributes to drug response will be important in predicting individual responses to drugs, both for efficacy and for the prediction of adverse events.

ACKNOWLEDGMENTS

We thank Karen Novik, Payal Sipahimalani, and Stephen Montgomery for their helpful comments. J.H.W. is supported by an Erwin-Schrödinger Fellowship from the Austrian Science Foundation (FWF). M.A.M. is supported by a Career Investigator Award from the Michael Smith Foundation for Health Research. We are grateful for the support of the British Columbia Cancer Agency, the British Columbia Cancer Foundation, and the Chan Sisters Foundation. We thank all the staff of the Canada's Michael Smith Genome Sciences Centre for their interest and enthusiasm.

REFERENCES

1. Jiang Z, Akey JM, Shi J, Xiong M, Wang Y, Shen Y, Xu X, Chen H, Wu H, Xiao J, Lu D, Huang W, Jin L. A polymorphism in the promoter region of catalase is associated with blood pressure levels. Hum Genet 2001; 109(1):95–98.
2. Waldron-Lynch F, Adams C, Amos C, Zhu DK, McDermott MF, Shanahan F, Molloy MG, O'Gara F. Tumour necrosis factor 5′ promoter single nucleotide polymorphisms influence susceptibility to rheumatoid arthritis (RA) in immunogenetically defined multiplex RA families. Genes Immun 2001; 2(2):82–87.
3. Coulthard SA, Hall AG. Recent advances in the pharmacogenomics of thiopurine methyltransferase. Pharmacogenomics J 2001; 1:254–261.
4. Bracco L, Kearsey J. The relevance of alternative RNA splicing to pharmacogenomies. Trends Biotechnol 2003; 21:346–353.
5. Christmas P, Jones JP, Patten CJ, Rock DA, Zheng Y, Cheng SM, Weber BM, Carlesso N, Scadden DT, Rettie AE, Soberman RJ. Alternative splicing determines the function of CYP4F3 by switching substrate specificity. In J Biol Chem 2001; pp. 38166–38172.
6. Kong S, Amos CI, Luthra R, Lynch PM, Levin B, Frazier ML. Effects of cyclin D1 polymorphism on age of onset of hereditary nonpolyposis colorectal cancer. Cancer Res 2000; 60:249–252.
7. Yang F, Hanson NQ, Schwichtenberg K, Tsai MY. Variable number tandem repeat in exon/intron border of the cystathionine beta-synthase gene: a single nucleotide substitution in the second repeat prevents multiple alternate splicing. Am J Med Genet 2000; 95:385–390.

8. Liggett SB. Beta(2)-adrenergic receptor pharmacogenetics. Am J Respir Crit Care Med 2000; 161(3 Pt 2):S197–S201.
9. Liggett SB. Pharmacogenetic applications of the human genome project. Nat Med 2001; 7:281–283.
10. Wagoner LE, Craft LL, Singh B, Suresh DP, Zengel PW, McGuire N, Abraham WT, Chenier TC, Dorn GW 2nd, Liggett SB. Polymorphisms of the beta(2)-adrenergic receptor determine exercise capacity in patients with heart failure. Circ Res 2000; 86:834–840.
11. Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci USA 2000; 97:10483–10488.
12. Loots G, Ovcharenko I, Pachter L, Dubchak I, Rubin E. rVISTA for comparative sequence-based discovery of functional transcription factor binding sites. Genome Res 2002; 12:832–839.
13. Schwartz S, Elnitski L, Li M, Weirauch M, Riemer C, Smit A, Green ED, Hardison RC, Miller W. NISC Comparative Sequencing Program. MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Res 2003; 31:3518–3524.
14. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. NISC Comparative Sequencing Program. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 2003; 13:721–731.
15. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin SQ, Maduro VV, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghighi P, Ho SL, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang LH, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED. Comparative analyses of multi-species sequences from targeted genomic regions. Nature 2003; 424:788–793.
16. Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT. Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Biol 1988; 203:439–455.
17. Hardison RC, Oeltjen J, Miller W. Long human–mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. Genome Res 1997; 7:959–966.
18. Duret L, Bucher P. Searching for regulatory elements in human noncoding sequences. Curr Opin Struct Biol 1997; 7:399–406.
19. Wasserman WW, Fickett JW. Identification of regulatory regions, which confer muscle-specific gene expression. J Mol Biol 1998; 278:167–181.

20. Wasserman WW, Palumbo M, Thompson W, Fickett JW, Lawrence CE. Human–mouse genome comparisons to locate regulatory sites. Nat Genet 2000; 26:225–228.
21. Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science 2003; 299:1391–1394.
22. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. Science 2002; 297:1143.
23. Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brandstrom H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnett D, Hudson TJ. A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics 2003; 16:184–193.
24. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat Genet 2003; 33:469–475.
25. Bray NJ, Buckland PR, Owen MJ, O'Donovan MC. *Cis*-acting variation in the expression of a high proportion of genes in human brain. Hum Genet 2003; 113:149–153.
26. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. Detection of regulatory variation in mouse genes. Nat Genet 2002; 32:432–437.
27. Olivier M. From SNPs to function: the effect of sequence variation on gene expression. Focus on "a survey of genetic and epigenetic variation affecting human gene expression." Physiol Genomics 2004; 16:182–183.
28. Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. Allelic variation in gene expression is common in the human genome. Genome Res 2003; 13:1855–1862.
29. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science 2002; 296:2225–2229.
30. Hudson TJ. Wanted: regulatory SNPs. Nat Genet 2003; 33:439–440.
31. O'Brien T, Hardin S, Greenleaf A, Lis JT. Phosphorylation of RNA poly merase II C-terminal domain and transcriptional elongation. Nature 1994; 370:75–77.
32. Komarnitsky P, Cho EJ, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev 2000; 14:2452–2460.
33. Jansen RC, Nap JP. Genetical genomics: the added value from segregation. Trends Genet 2001; 17:388–391.
34. Jansen RC, Nap JP. Regulating gene expression: surprises still in store. Trends Genet 2004; 20:223–225.
35. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. Genetics of gene expression surveyed in maize, mouse and man. Nature 2003; 422:297–302.
36. Schadt EE, Monks SA, Friend SH. A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. Biochem Soc Trans 2003; 31:437–443.

37. Darvasi A. Genomics: gene expression meets genetics. Nature 2003; 422:269–270.
38. Yan L, Otterness DM, Kozak CA, Weinshilboum RM. Mouse nicotinamide N-methyltransferase gene: molecular cloning, structural characterization, and chromosomal localization. *DNA cell Biol.* 1998; 17(8):639–667.
39. The International HapMap Consortium. The International HapMap Project. Nature 2003; 426:789–796.
40. Cleator S, Ashworth A. Molecular profiling of breast cancer: clinical implications. Br J Cancer 2004; 90:1120–1124.
41. Kohn EC, Lu Y, Wang H, Yu Q, Yu S, Hall H, Smith DL, Meric-Bernstam F, Hortobagyi GN, Mills GB. Molecular therapeutics: promise and challenges. Semin Oncol 2004; 31:39–53.
42. Kumar-Sinha C, Chinnaiyan AM. Molecular markers to identify patients at risk for recurrence after primary treatment for prostate cancer. Urology 2003; 62(suppl 1):19–35.
43. Chin KV, Kong AN. Application of DNA microarrays in pharmacogenomics and toxicogenomics. Pharm Res 2002; 19:1773–1778.
44. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995; 270:467–470.
45. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High-density synthetic oligonucleotide arrays. Nat Genet 1999; 21:20–24.
46. Hacia JG, Collins FS. Mutational analysis using oligonucleotide microarrays. J Med Genet 1999; 36:730–736.
47. Afshari CA, Nuwaysir EF, Barrett JC. Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. Cancer Res 1999; 59:4759–4760.
48. Kudoh K, Ramanna M, Ravatn R, Elkahloun AG, Bittner ML, Meltzer PS, Trent JM, Dalton WS, Chin KV. Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. Cancer Res 2000; 60:4161–4166.
49. Voehringer DW, Hirschberg DL, Xiao J, Lu Q, Roederer M, Lock CB, Herzenberg LA, Steinman L. Gene microarray identification of redox and mitochondrial elements that control resistance or sensitivity to apoptosis. Proc Natl Acad Sci USA 2000; 97:2680–2685.
50. Quackenbush J. Computational analysis of microarray data. Nat Rev Genet 2001; 2:418–427.
51. Anderson KC, Bates MP, Slaughenhoupt BL, Pinkus GS, Schlossman SF, Nadler LM. Expression of human B cell-associated antigens on leukemias and lymphomas: a model of human B cell differentiation. Blood 1984; 63: 1424–1433.
52. Harris M. Monoclonal antibodies as therapeutic agents for cancer. Lancet Oncol 2004; 5:292–302.
53. Ramaswamy S, Golub TR. DNA microarrays in clinical oncology. J Clin Oncol 2002; 20:1932–1941.
54. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Res 2002; 62:4427–4433.

55. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM. Delineation of prognostic biomarkers in prostate cancer. Nature 2001; 412:822–826.
56. Luo J, Duggan DJ, Chen Y, Sauvageot J, Ewing CM, Bittner ML, Trent JM, Isaacs WB. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. Cancer Res 2001; 61:4683–4688.
57. Magee JA, Araki T, Patil S, Ehrig T, True L, Humphrey PA, Catalona WJ, Watson MA, Milbrandt J. Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res 2001; 61:5692–5696.
58. Welsh JB, Sapinoso LM, Su AI, Kern SG, Wang-Rodriguez J, Moskaluk CA, Frierson HF Jr, Hampton GM. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res 2001; 61:5974–5978.
59. Burger MJ, Tebay MA, Keith PA, Samaratunga HM, Clements J, Lavin MF, Gardiner RA. Expression analysis of delta-catenin and prostate-specific membrane antigen: their potential as diagnostic markers for prostate cancer. Int J Cancer 2002; 100:228–237.
60. Ernst T, Hergenhahn M, Kenzelmann M, Cohen CD, Bonrouhi M, Weninger A, Klaren R, Grone EF, Wiesel M, Gudemann C, Kuster J, Schott W, Staehler G, Kretzler M, Hollstein M, Grone HJ. Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue. Am J Pathol 2002; 160: 2169–2180.
61. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Res 2002; 62:4499–4506.
62. Luo JH, Yu YP, Cieply K, Lin F, Deflavia P, Dhir R, Finkelstein S, Michalopoulos G, Becich M. Gene expression analysis of prostate cancers. Mol Carcinog 2002; 33:25–35.
63. Jiang Z, Woda BA, Rock KL, Xu Y, Savas L, Khan A, Pihan G, Cai F, Babcook JS, Rathanaswami P, Reed SG, Xu J, Fanger GR. P504S: a new molecular marker for the detection of prostate carcinoma. Am J Surg Pathol 2001; 25:1397–1404.
64. Luo J, Zha S, Gage WR, Dunn TA, Hicks JL, Bennett CJ, Ewing CM, Platz EA, Ferdinandusse S, Wanders RJ, Trent JM, Isaacs WB, De Marzo AM. Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. Cancer Res 2002; 62:2220–2226.
65. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286:531–537.
66. Levene AP, Morgan GJ, Davies FE. The use of genetic microarray analysis to classify and predict prognosis in haematological malignancies. Clin Lab Haematol 2003; 25:209–220.
67. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. Nature 2000; 405:827–836.

68. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000; 403: 503–511.
69. Alizadeh A, Eisen M, Davis RE, Ma C, Sabet H, Tran T, Powell JI, Yang L, Marti GE, Moore DT, Hudson JR Jr, Chan WC, Greiner T, Weisenburger D, Armitage JO, Lossos I, Levy R, Botstein D, Brown PO, Staudt LM. The lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. Cold Spring Harb Symp Quant Biol 1999; 64:71–78.
70. Bohen SP, Troyanskaya OG, Alter O, Warnke R, Botstein D, Brown PO, Levy R. Variation in gene expression patterns in follicular lymphoma and the response to rituximab. Proc Natl Acad Sci USA 2003; 100:1926–1930.
71. Stratowa C, Loffler G, Lichter P, Stilgenbauer S, Haberl P, Schweifer N, Dohner H, Wilgenbus KK. CDNA microarray gene expression analysis of B-cell chronic lymphocytic leukemia proposes potential new prognostic markers involved in lymphocyte trafficking. Int J Cancer 2001; 91:474–480.
72. Thieblemont C, Chettab K, Felman P, Callet-Bauchu E, Traverse-Glehen A, Berger F, Dumontet C, Cerutti C, Baseggio L, Salles G, Paultre C, Coiffier B, McGregor JL. Identification and validation of seven genes, as potential markers, for the differential diagnosis of small B cell lymphomas (small lymphocytic lymphoma, marginal zone B cell lymphoma and mantle cell lymphoma) by cDNA macroarrays analysis. Leukemia 2002; 16:2326–2329.
73. Zhan F, Hardin J, Kordsmeier B, Bumm K, Zheng M, Tian E, Sanderson R, Yang Y, Wilson C, Zangari M, Anaissie E, Morris C, Muwalla F, van Rhee F, Fassas A, Crowley J, Tricot G, Barlogie B, Shaughnessy J Jr. Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. Blood 2002; 99: 1745–1757.
74. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, Ferno M, Peterson C, Meltzer PS. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 2001; 61: 5979–5984.
75. Van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415: 530–536.
76. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR, Nevins JR. Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci USA 2001; 98: 11462–11467.
77. Carr KM, Bittner M, Trent JM. Gene-expression profiling in human cutaneous melanoma. Oncogene 2003; 22:3076–3080.

78. Huppi K, Chandramouli GV. Molecular profiling of prostate cancer. Curr Urol Rep 2004; 5:45–51.
79. Meyerson M, Franklin WA, Kelley MJ. Molecular classification and molecular genetics of human lung cancers. Semin Oncol 2004; 31:4–19.
80. Sanchez-Carbayo M, Cordon-Cardo C. Applications of array technology: identification of molecular targets in bladder cancer. Br J Cancer 2003; 89: 2172–2177.
81. Rae JM, Johnson MD, Lippman ME, Flockhart DA. Rifampin is a selective, pleiotropic inducer of drug metabolism genes in human hepatocytes: studies with cDNA and oligonucleotide expression arrays. J Pharmacol Exp Ther 2001; 299:849–857.
82. Maxwell PJ, Longley DB, Latif T, Boyer J, Allen W, Lynch M, McDermott U, Harkin DP, Allegra CJ, Johnston PG. Identification of 5-fluorouracil-inducible target genes using cDNA microarray profiling. Cancer Res 2003; 63:4602–4606.
83. Hamadeh HK, Bushel PR, Jayadev S, DiSorbo O, Bennett L, Li L, Tennant R, Stoll R, Barrett JC, Paules RS, Blanchard K, Afshari CA. Prediction of compound signature using high density gene expression profiling. Toxicol Sci 2002; 67:232–240.
84. Hamadeh HK, Bushel PR, Jayadev S, Martin K, DiSorbo O, Sieber S, Bennett L, Tennant R, Stoll R, Barrett JC, Blanchard K, Paules RS, Afshari CA. Gene expression analysis reveals chemical-specific profiles. Toxicol Sci 2002; 67:219–231.
85. Lakkis MM, DeCristofaro MF, Ahr HJ, Mansfield TA. Application of toxicogenomics to drug development. Expert Rev Mol Diagn 2002; 2:337–345.
86. Waring JF, Halbert DN. The promise of toxicogenomics. Curr Opin Mol Ther 2002; 4:229–235.
87. Gunther EC, Stone DJ, Gerwien RW, Bento P, Heyes MP. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. Proc Natl Acad Sci USA 2003; 100:9608–9613.
88. Cheok MH, Yang W, Pui CH, Downing JR, Cheng C, Naeve CW, Relling MV, Evans WE. Treatment-specific changes in gene expression discriminate in vivo drug response in human leukemia cells. Nat Genet 2003; 34:85–90.
89. Stegmaier K, Ross KN, Colavito SA, O'Malley S, Stockwell BR, Golub TR. Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. Nat Genet 2004; 36:257–263.
90. Huang ME, Ye YC, Chen SR, Chai JR, Lu JX, Zhoa L, Gu LJ, Wang ZY. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. Blood 1988; 72:567–572.

18

Proteomics

Frank A. Witzmann

Department of Cellular and Integrative Physiology, Indiana University School of Medicine, Indianapolis, Indiana, U.S.A.

Mu Wang

Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana, U.S.A.

1. OVERVIEW

Proteomics continues to grow in relevance and technical complexity as its importance in the investigation and understanding of protein expression in disease and drug development is recognized. Advances in proteomics have been supported by the continued expansion and availability of pharmacologically relevant genome information and driven by numerous and diverse technological and bioinformatic developments that underscore the complexity of the proteome. This chapter addresses some of the various technologies included in proteomic approaches, and provides salient examples where these technologies have been applied to pharmacology and/or the process of drug development.

2. INTRODUCTION

Whether it is pharmacology, toxicology, environmental science, or molecular pathology to name a few, researchers are increasingly cognizant of the importance of studying the proteome, in favor of profiling gene expression exclusively. This is true both from a quantitative and qualitative standpoint,

and it is partly the result of limitations of the functional genomic approach. Side-by-side comparison of differential mRNA and protein expression, when attempted, typically has demonstrated a weak or poor correlation between them (1–5). In addition, significant fluctuations in a cell's proteome are unrelated to up- or down-regulation; they are instead the temporary result of a myriad of posttranslational modifications to existing proteins. To accurately assess changes in cell function, one must be in a position to determine: (a) the actual (or at least relative) quantities of the proteins expressed at a given time, and (b) the extent of critical posttranslational modifications that regulate protein activity. Relative quantitation of mRNA, even on a global scale, cannot provide this information nor can it describe the dynamics of protein–protein interactions.

Proteomics thus measures the quantitative and qualitative changes in cellular or tissue protein expression and explores protein–protein and protein–ligand interactions. Though portrayed here as an alternative to pharmacogenomics, pharmacoproteomics is fundamentally a complementary approach for the identification and validation of protein targets and characterization of drug effects, good and bad. Accordingly, the pharmaceutical industry has developed an interest in proteomics that continues to increase, with the expectation that this approach will lead to the discovery and development of viable drug candidates (6–8).

Compared to functional genomic approaches, many proteomic technologies are new and rapidly developing. Because each expressed protein is unique and the final, a fully functional protein product rarely resembles the gene(s) from which it was translated. Therefore, a comprehensive approach including a variety of techniques is necessary to detect, quantify, and identify individual proteins and to elucidate protein activities and determine protein–protein interactions. Unlike mRNA, protein analysis is severely limited by the tremendous dynamic range of protein expression and the large number of relevant proteins whose cellular expression or abundance in body fluids is minute. Thus, advances in proteomic research have been driven by new technologies and approaches that address this issue by increasing the sensitivity and accuracy of measurement. This chapter will briefly address several of the core technologies that comprise contemporary proteomics and, in some cases, include examples where each has been applied to pharmacology. Additional literature addressing this topic can be found in a number of review articles (9–19).

3. PHARMACOPROTEOMIC APPROACHES

General consensus has sub-divided proteomics into three main areas, Expression Proteomics, Functional Proteomics, and Structural Proteomics. Expression Proteomics (sometimes called differential-expression proteomics) involves the analysis of differential protein expression by protein

quantitation and identification. Differential protein analysis compares the expression profiles of proteins (the proteomes) of cells, tissues, or organisms in one condition (e.g., disease, injury, drug therapy, or intoxication) to a standard or "normal" proteome. Differences in the proteomes indicate the biological response, hence the utility of this approach in the development of drugs.

In contrast, Functional Proteomics concerns the manner in which proteins interact and, in turn, how these interactions determine function, both normal and abnormal. This approach is less reductionist than Expression Proteomics as proteins are studied in the context of their complex cellular interactions. Finally, Structural Proteomics is concerned with the primary through tertiary structure of proteins, and modifications therein, largely determined by x-ray and NMR analysis of protein crystals (20). In this chapter we will concern ourselves only with Expression and Functional Proteomics.

3.1. Expression Proteomics

3.1.1. Two-Dimensional Gel Electrophoresis and Peptide Mass Fingerprinting

The classical combination of two-dimensional gel electrophoresis (2DE) and mass spectrometry (MS) (Fig. 1) remains the most widely used approach in proteomic analysis. This figure also illustrates the additional use of tandem

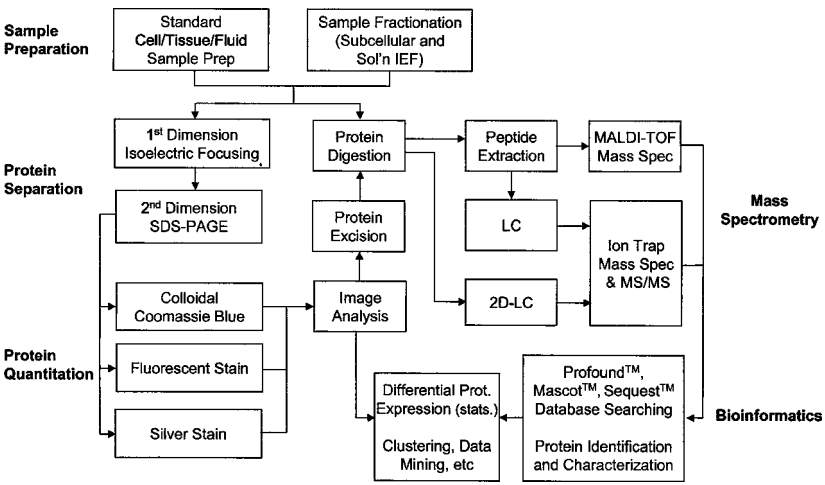


Figure 1 Schematic of the typical workflow of 2DE gel-based proteomics. This strategy includes flexible steps that can bypass 2DE, when desired, by directly proteolyzing cell lysates and cell fractions and analyzing by 2DLC-MS/MS.

MS, *either* after electrophoretic separation and proteolysis *or* independently of 2DE [e.g., as in Multidimensional Protein Identification Technology (MudPIT), Ref. 21, also see Sec. 3.1.3), as a stand-alone system for protein identification.

3.1.1.1. Two-dimensional gel electrophoresis: In 2DE, proteins are separated first by charge using isoelectric focusing (IEF) on tube gels or immobilized pH-gradient (IPG) strips, then by molecular weight using SDS-polyacrylamide slab gel electrophoresis, and finally stained for image analysis or visual inspection. Because the dynamic range of protein expression in most whole cell or tissue lysates is huge, estimated to be up to 10 orders of magnitude in body fluids (22), and only the most abundant proteins from 2D gels can be analyzed, far too many proteins are overlooked when 2DE is used. Typically, only 2000 or so of the most abundant proteins in an appropriately solubilized sample can be reliably separated and identified. Therefore, to analyze the less abundant proteins in a sample, prefractionation of the starting material is required for protein enrichment. To that end, novel sample preparation steps aimed at increasing protein recovery and resolution have been reported (23–27). Additional sample complexity reduction aimed at digging deeper into the broad dynamic range of protein expression has been accomplished using subcellular fractionation by differential centrifugation, yielding cytosol, mitochondria, microsomes, or nuclei/membranes, whose proteomes are subsequently separated on individual 2D gels (28–30). These types of prefractionation approaches have significantly improved the utility of 2D gel-based proteomics in liver analysis and should become standard operating procedure when analyzing any tissues or organs, electrophoretically or otherwise.

Another strategy to reduce sample complexity and increase analytical depth of field is through the serial fractionation of subcellular components into more manageable compartments by solution IEF (31,32). For example, subcellular protein fractions can be subdivided into ~0.8–3 pH unit fractions using the ZoomIEF™ apparatus (Invitrogen Corp., Carlsbad, CA). The resulting fractions can then be resolved on large format 2D gels (25×20 cm) whose individual horizontal dimension (pI) is defined by linear pH unit gradients of matching pH, and 24-cm long. The overall approach is illustrated in Figure 2 and an example of whole liver lysate separated into a 0.8 pH unit is shown in Figure 3. Initial experiments using liver cell fractions (authors' unpublished data) suggest that a single tissue sample, fractionated into four subcellular fractions and five solution IEF fractions, will yield 2DE differential expression data for at least ~20,000 protein spots, i.e., ~1000 on 20 individual gels. The application of solution IEF fractionation to liver subcellular fractions, or those from any target tissue, as described above will specifically address the issue of dynamic range and significantly improve proteomic comparisons using 2DE.

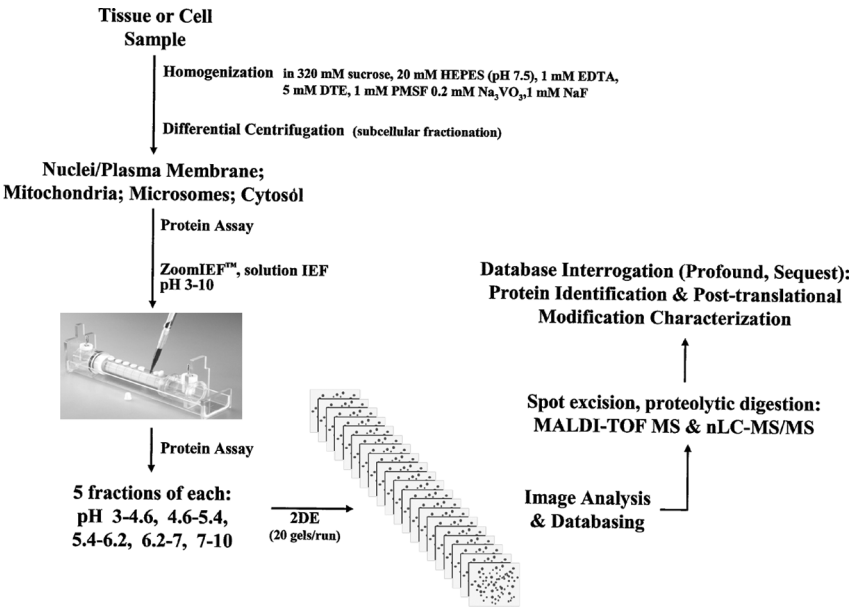


Figure 2 Schematic of a strategy for reduction of sample complexity. By combining the fractionation of cell lysates by differential centrifugation with subsequent solution IEF, followed by parallel 2DE on a 20-gel platform, significantly more proteins per sample can be resolved and analyzed.

Despite some of its limitations, 2D gel-based proteomics remains an exceptional approach for assessing differential protein expression, particularly in studies with pharmacological considerations. Since its inception, 2DE has been a useful tool in drug development, especially in the analysis of adverse drug effects. For example, methapyrilene, a putative antihistamine and sleep aid withdrawn from the market after it was discovered to be a potent liver carcinogen in rodents, was shown by 2DE to target the hepatic mitochondrion by forming protein adducts specifically in this organelle (33–35). The nature of the protein modification indicated that it was covalent and resulted from binding of a negatively charged adduct. This study raised the likelihood that through a putative reactive metabolite, by attacking mitochondrial but not nuclear DNA, exerted methapyrilene's genotoxicity in an unconventional way. As many other studies have demonstrated (36–44), detection of protein posttranslational modifications as conspicuous and quantifiable charge variants (with heterogeneous pI) demonstrates the distinctiveness of 2DE, above all other proteomic methods, in detecting and quantifying drug-related processes that result in the formation reactive metabolites.

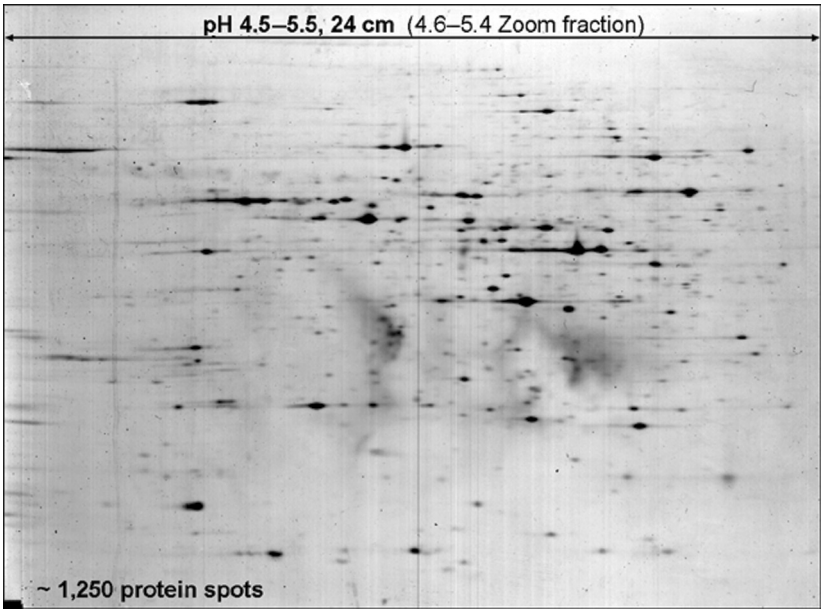


Figure 3 The 2DE gel image of Sprague–Dawley rat liver sample that was fractionated by solution IEF using the ZoomIEF Fractionator (Invitrogen Corp.). This image depicts the pH 4.6–5.4 fraction that was isolated and used for IEF on a 24 cm, immobilized pH 4.5–5.5 gradient strip (Amersham, Piscataway, NJ) followed by SDS–PAGE separation on an 11–17% acrylamide gradient slab gel. The pattern was stained with colloidal Coomassie blue and scanned. Using PDQuest® Image Analysis software, ~1250 protein spots were detected in this narrow pH range fraction (authors' unpublished data).

Other examples of studies in which 2DE has been used as a pharmacoproteomic tool include the characterization of Cyclosporine A toxicity in rat kidney (45–47) and brain (48), and the protein-level characterization of histopathologies observed in the liver during preclinical assessment of drug candidates (substituted pyrimidine derivatives) (49). The hepatic effects of cholesterol-lowering statin drugs (HMG-CoA reductase inhibitors) such as lovastatin (e.g., Mevacor®, a natural product) and fluvastatin (e.g., Lescol®, a totally synthetic product) were studied (50,51) to investigate the response of other proteins (enzymes) in the pathways; to evaluate statin hepatotoxicity at high doses; and to generate potential markers for use in a high-throughput screen (HTS) assay format to compare the therapeutic windows of different members of the statin family.

More recently, a similar approach was used successfully to investigate farnesyl transferase inhibitor treatment of ovarian cancer (52,53) and in describing the protein molecular basis for drug-resistance development in pancreatic cancers (5,54).

Another of the utilities of 2D gel-based proteomics lies in biomarker development, e.g., generating sets of proteins that can be used as indicators (5,55,56) and even predictors of chemical effects on cells and tissues (57), as in candidate screening in drug or chemical safety evaluation and efficacy studies. Recognition and validation of sets of appropriate protein biomarkers must first be accomplished before truly functional, high-throughput drug toxicity screening systems such as protein chips containing these proteins can be developed. The 2D gel-based proteomics can serve as a way to identify such a pool of biomarkers, as a recent application of this approach to characterize the nephrotoxicant gentamicin demonstrates (58).

3.1.1.2. Peptide mass fingerprinting: The MS is the method of choice for identifying 2D gel-separated proteins. Separated protein spots can be identified after digestion with trypsin and analysis by any one of several MS methods. A tryptic digest can be analyzed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS, a process referred to as peptide mass fingerprinting (PMF) (59). The measured and optimized monisotopic mass data then are compared with theoretically derived peptide mass databases, generated by applying specific enzymatic cleavage rules to predicted/known protein sequences. MALDI-TOF-based PMF enables high-throughput, accurate, and sensitive mass detection. For unambiguous identification of 2D gel-separated proteins, spectral data derived by MS/MS can be compared to spectra predicted from sequence databases using search engines such as Mascot (60) and algorithms such as SEQUEST (61). In some cases, peptide sequence tags (62) from either MALDI or electrospray ionization ion sources can also be used for protein identification. The sequence of long peptides (20+ amino acids) can be determined using chemically assisted fragmentation (CAF) MALDI post-source decay (PSD) MS (63). This is especially useful when trying to identify proteins isolated from organisms with unsequenced genomes. Finally, with the advent of the TOF/TOF mass spectrometer (64), protein identification is entering a new era in which all of the above methods will become faster and more accurate.

3.1.2. Isotope-Coded Affinity Tags—Quantitative Alternatives to 2DE

A novel technique known as Isotope-Coded Affinity Tagging (ICAT) has been developed to improve and expand relative protein quantitation in differential expression proteomics studies where MS methods are used in place of 2DE. Using light or heavy isotope-labeled peptide "tagging" reagents that differ only in molecular mass, proteins derived from normal/diseased, or untreated/treated samples can be quantified, compared, and identified using LC–MS/MS (3). The tagging reagents contain a cysteine-reactive alkylating group at one end and a biotin tag at the other. After mixing the differently labeled proteins together, they are digested by

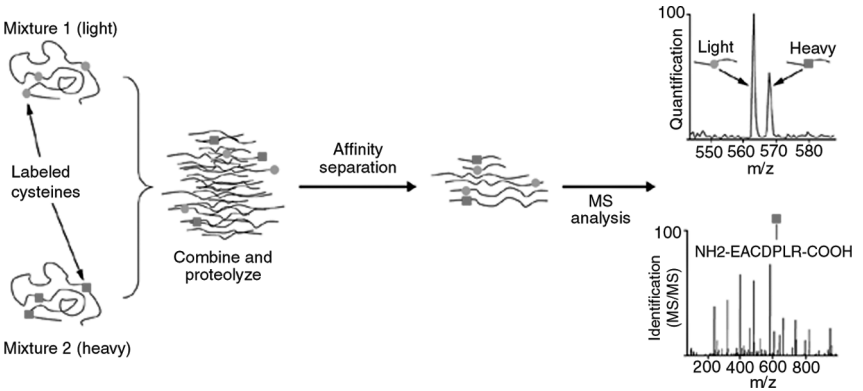


Figure 4 The ICAT reagent strategy for quantifying differential proteins. Two protein mixtures representing two different cell states are treated with the isotopic ally light (•) or heavy (■) ICAT reagents, respectively. The labeled protein mixtures are then combined and proteolyzed; tagged peptides are selectively isolated and analyzed by MS. The relative abundance is determined by the ratio of signal intensities of the tagged peptide pairs. Every other scan is devoted to fragmenting a peptide. The CID spectra are recorded and searched against large protein sequence databases to identify the protein. Therefore, in a single operation, the relative abundance and sequence of a peptide are determined (140).

the addition of trypsin and the biotin-tagged, cysteine-containing peptides, are purified over an avidin column. These peptides are then separated on a C18 reversed phase column that is directly coupled to an MS/MS instrument. As Figure 4 illustrates, the relative amounts of the various peptides in the original sample are determined from the ratio of the isotope-labeled ion pairs *and* proteins are identified from the fragmentation pattern. This method seems to work best when the starting materials are partially purified protein mixtures and/or protein complexes (65).

A number of variations to the standard method exist. To reduce both the complexity of the sample and the computing resources necessary for data analysis, intact ICAT-labeled proteins have been fractionated initially on a 2D gel, and then digested, quantitated, and identified by MS (66,67). Similarly, MS methods using microcapillary electrospray ionization time-of-flight MS to determine those peptides that differ in abundance followed by their identification using tandem MS have recently been used by Griffin et al. (68,69) to study the tumorigenic potential of human prostate epithelial cells. Arnott et al. (70) used a modified ICAT reagent to verify the presence and measure the amount of membrane proteins that were predicted to be present from gene expression studies. Finally, the development of second-generation ICAT reagents such as a cleavable ICAT reagent (71) has helped to overcome some of the technical limitations of

the original reagents and expand their utility. Different chemistries and approaches to tag protein samples have been developed (72–75), and stable isotope incorporation methods for laser-capture samples have been devised (76).

3.1.3. Multidimensional Protein Identification Technology

A technique that addresses "global" protein analysis head-on is MudPIT. This is a "shotgun" approach to protein identification where mixtures of intact proteins are proteolyzed to form a complex peptide mixtures and analyzed directly by multidimensional liquid chromatography and tandem MS (LC/LC–MS/MS) (77). Acquired peptide fragmentation spectra are correlated with predicted amino acid sequences in translated genomic databases using the SEQUEST algorithm (61). This "gel-free" approach avoids the manual data interpretation from each tandem mass spectrum and provides a rapid method to process the spectrometric data. Figure 5 shows a general experimental flowchart of MudPIT. Once a complex peptide mixture is loaded onto the system, no additional sample handling is required because the peptides are eluted directly off the column and into the mass spectrometer. This technology has been applied to large-scale analysis of the yeast proteome (21), membrane proteomics (78), identification of protein modifications (79), and most recently quantitative analysis using ICAT and stable metabolic labeling strategies (80).

3.1.4. Surface-Enhanced Laser Desorption–Ionization Protein Profiling

Another useful MS-based method for proteomic analysis makes use of the ProteinChip® System from Ciphergen Biosystems, Fremont, CA. The effectiveness of this system resides in the surface-enhanced laser desorption and ionization (SELDI) ProteinChip surface technology that performs separation, detection, and analysis of proteins at femtomole levels, directly from complex (unfractionated) biological samples. In essence, specifically modified slides bearing various surface chemistries (cationic, anionic, hydrophobic, hydrophilic, etc.) or biochemistries (antibody, receptor, DNA, etc.) that bind and selectively purify proteins from a complex biological sample. For a given ProteinChip, various buffer and elution conditions can be used to further fractionate the sample. The slide then can be analyzed using a SELDI mass spectrometer (essentially a MALDI-TOF instrument) as portrayed in Figure 6. The ease and speed of screening samples has made this a popular method for biomarker detection for clinical and toxicological samples (81–83). In a classic paper illustrating the potential of SELDI in the study of cancer, Petricoin et al. (84) identified a proteomic pattern in serum that is diagnostic for ovarian cancer. Using the ProteinChip/SELDI technology and a custom algorithm to recognize protein patterns in the spectra, they first used sera from 50 unaffected women and 50 patients with ovarian

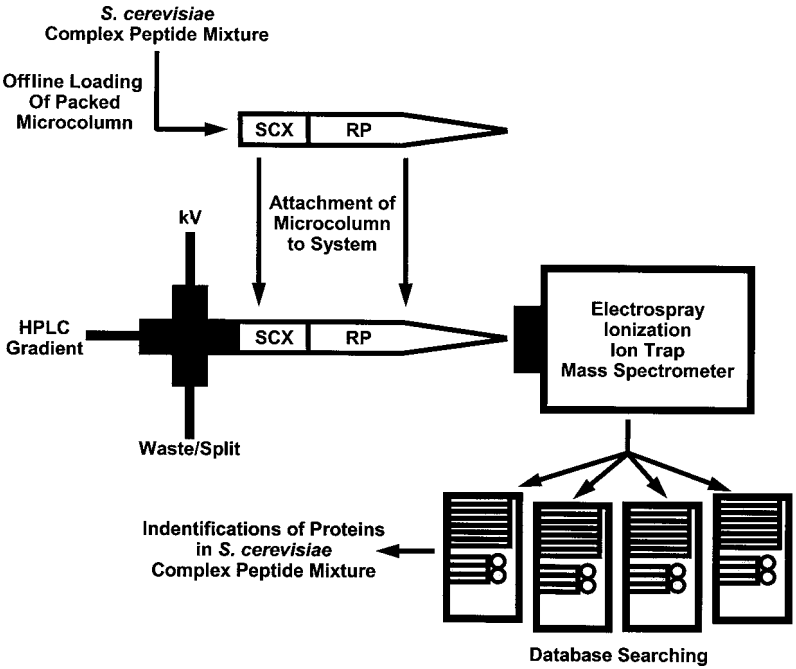


Figure 5 Schematic of the MudPIT approach (21). Complex peptide mixtures from different fractions of a *Saccharomyces cerevisiae* whole-cell lysate are loaded separately onto a biphasic microcapillary column packed with strong cation exchange (SCX) and reverse-phase (RP) materials. After loading the complex peptide mixture into the microcapillary column, the column was inserted into the instrumental setup. Xcalibur software, HPLC, and mass spectrometer were controlled simultaneously by means of the user interface of the mass spectrometer. Peptides directly eluted into the tandem mass spectrometer because a voltage (kV) supply is directly interfaced with the microcapillary column. As described in the experimental protocol, peptides were first displaced from the SCX to the RP by a salt gradient and eluted off the RP into the MS/MS. In an iterative process, the microcolumn was re-equilibrated and an additional salt step of higher concentration displaced peptides from the SCX to the RP. Peptides were again eluted by an RP gradient into the MS/MS, and the process was repeated. The tandem mass spectra generated were correlated to theoretical mass spectra generated from protein or DNA databases by the SEQUEST algorithm.

cancer to identify a proteomic pattern that completely discriminated cancer from non-cancer. Using this pattern for a set of 116 masked serum samples, they then correctly identified 50/50 ovarian cancer cases and correctly identified 63/66 non-cancerous samples. In a similar application, Rosty et al. (85) used the ProteinChip/SELDI technology to screen for

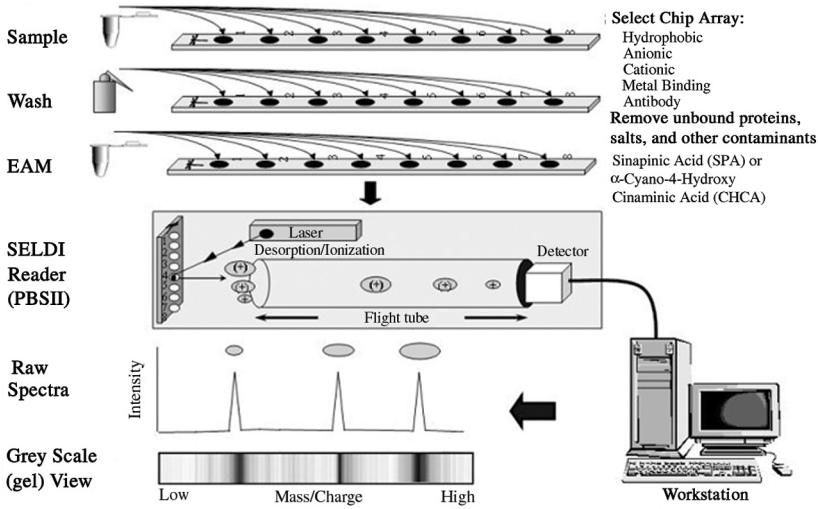


Figure 6 The SELDI technology. This type of proteomic analytical tool is a class of mass spectroscopy instrument that is useful in high-throughput proteomic fingerprinting of serum. Using a robotic sample dispenser, 1 μL of serum is applied to the surface of a protein-binding chip. A subset of the proteins in the sample binds to the surface of the chip. The bound proteins are treated with a matrix-assisted laser desorption and ionization matrix and are washed and dried. The chip, which contains multiple patient samples, is inserted into a vacuum chamber where it is irradiated with a laser. The laser desorbs the adherent proteins and causes them to be launched as ions. The TOF of the ion before detection by an electrode is a measure of the mass-to-charge (m/z) value of the ion. The ion spectra can be analyzed by computer-assisted tools that classify a subset of the spectra by characteristic patterns of relative intensity (adapted from www.evmsdoctors.com).

differentially expressed proteins in pancreatic juice from patients with and without pancreatic adenocarcinoma. After finding a protein with a molecular weight of ~16,500 that was elevated significantly in the affected individuals, they were able to identify this protein using an immunoassay.

Despite its clever utility, SELDI suffers from several limitations. The SELDI MS instrumentation usually is capable of accurately detecting proteins with molecular weights less than 45,000, the detected proteins cannot be identified using this technique alone, and reproducibility in complicated experiments is suspect (86). Improvements in next-generation instruments using ProteinChip tandem MS techniques that enable direct protein identification (87), improved surface chemistries (88), and improved experimental design (89,90) should all greatly enhance SELDI's effectiveness as a powerful proteomic tool (91,92).

3.1.5. Protein (Micro) Arrays

Analogous to but not yet as technically developed as the genomic microarray, the protein microarray is another emerging technique that holds great analytical promise (93,94), especially with respect to throughput. In general, two types of microarrays are used in proteomics: the antibody array (95) and the functional protein microarray (96) (Fig. 7).

In antibody microarrays, antibodies to proteins of interest (or otherwise—specific protein capture agents) are prepared in high-throughput or semi-high-throughput fashion, and then spotted onto a specially treated surface (e.g., glass, silica, etc.). The principle is similar to the enzyme-linked immunosorbent assay (ELISA) immunoassay and the rate-limiting step often is obtaining enough useful antibodies or protein binding agents to produce a truly functional array. Protein detection and quantitation can be accomplished using several methods, but fluorescence is the most common and detection is very sensitive, at picogram per milliliter levels. For example, cytokine antibody arrays were used to establish the profiles of cytokine release from THP-1 monocytes exposed to different amphotericin B (AMB) drug delivery systems (97). Arrays with up to 500 different antibodies are now commercially available from various vendors and arrays targeting specific cellular conditions, e.g., apoptosis, cell adhesion, cancer growth, and signal transduction, to name a few, are also available.

Functional protein microarrays are useful for direct analysis of biochemical activity (e.g., substrate specificity, Refs. 98 and 99), protein–protein interaction (100), or small molecule–protein interaction. Proteins of interest are over expressed and applied in pure form on addressable arrays, and assays performed (101).

For example, Haab et al. (102) were some of the first to demonstrate the feasibility of using protein microarrays to simultaneously detect antibody/antigen pairs with selectivity and sensitivity. Kukar et al. (103) immobilized proteins onto derivatized glass or nitrocellulose-coated slides and then detected these proteins in a quantitative manner by probing with fluorescent proteins known to be binding partners. Finally, Cahill (104) and collaborators have produced and purified more than 10,000 unique human proteins and arranged them onto PVDF filters and glass slides. These high-density arrays are being used to screen sera from patients with autoimmune diseases in order to identify the proteins to which the body is making autoantibodies.

3.1.6. Activity-Based Probes

The isolation and analysis of specific classes of proteins within complex mixtures, particularly those in low abundance, is based on the unique biochemical properties that are characteristic of that class of proteins. These types of approaches take advantage of small molecule/protein binding

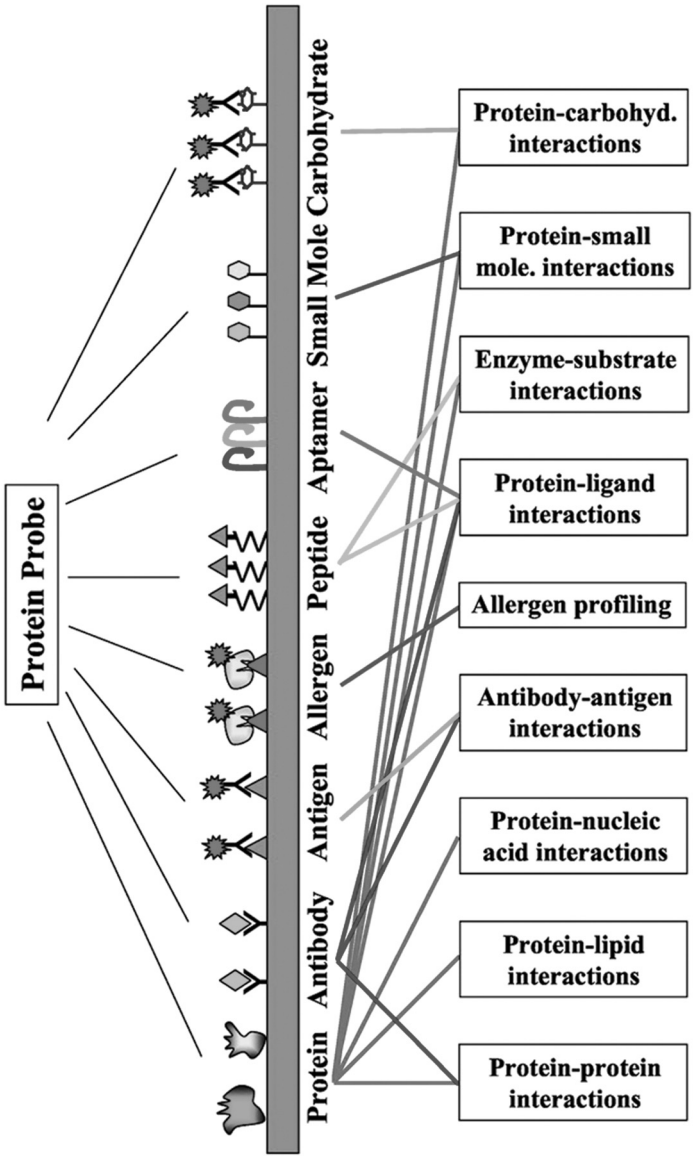


Figure 7 Protein microarrays and their applications. Ligands, such as proteins, peptides, antibodies, antigens, allergens, and small molecules, are immobilized in high density on modified surfaces to form functional and analytical protein microarrays. These protein microarrays can also be used for various kinds of biochemical analysis.

specificity to tag proteins of interest so that they can be readily isolated, identified, and further studied. In effect, specific components of a proteome are profiled based on activity rather than quantity. For example, this approach has been used to study a series of serine hydrolase inhibitors that covalently react with active proteases, lipases, and esterases tagged with biotin or fluorescent labels (105,106), as well as a class I aldehyde dehydrogenase inhibited by various sulfonates (107). These activity-based probes have been used (108) to profile the active hydrolases in a number of human breast and melanoma cancer cell lines. The proteomic signatures observed in many of these cell lines suggested novel targets for the diagnosis and treatment of human cancer. Similarly, activity-based probes that specifically tag active cysteine proteases have been developed (109). Binding partners that interact in a non-covalent manner can be studied by modification of these methods as well. For example, specific non-covalent complexes formed between small molecule ligands and the proteins of interest can be converted into stable covalent complexes by photo-crosslinking (e.g., with an azido group, Ref. 110). In addition to target discovery, the activity-based approach has significant potential in accelerating the discovery of potential drug candidates (111,112).

3.2. Functional Proteomics

As mentioned earlier, Functional Proteomics is concern with the manner in which proteins interact and, in turn, how these interactions ultimately determine function and dysfunction. This approach is less reductionist than Expression Proteomics because proteins are studied in the context of their complex cellular interactions.

3.2.1. Isolation of Protein Complexes

Most major tasks within the cell, from gene transcription and mRNA splicing to protein degradation, signal transduction, and cell cycle regulation, tend to be carried out by multiprotein complexes rather than individual proteins (113). Accordingly, examining protein complexes often will yield information about protein function. One way to approach this problem is to isolate large multiprotein complexes from the cell, such as ribosomes. Using multidimensional LC and tandem MS, Link et al. (77) showed that yeast ribosomes contain over 80 components, many of which were not previously known to be associated with ribosomes. Several of these proteins provide a link of the translational apparatus to other processes in the cell such as transcription, RNA processing, metabolic signaling, etc.

Essentially all proteins in the cell interact with other proteins as members of multiprotein complexes. However, most of these complexes cannot easily be isolated for study. Antibodies against one member of a complex sometimes are useful for purification, but often the epitope is not

accessible when the protein is in a complex. In general, the use of molecular biology techniques to insert a tag (e.g., His-tag, Flag-tag, GST, etc.) into the gene coding for the protein of interest has been more successful. This approach has been successful, as demonstrated by two recent important studies. Gavin et al. (114) used a novel tandem-affinity purification (TAP) tagging technique (Fig. 8) to isolate over 200 distinct multiprotein complexes from yeast. The size of the complexes varied from 2 to 83 different proteins, with an average of 5 to 10 proteins per complex. Most proteins were members of (on average) about five distinct complexes. Over 300 proteins were assigned new roles based on their association with proteins of known function. When done on a large scale, such work begins to unravel the connectedness of proteins and the interconnected biochemical networks in which they participate. Ho et al. (115), using high-throughput mass spectrometric protein complex identification (HMS-PCI) very analogous in

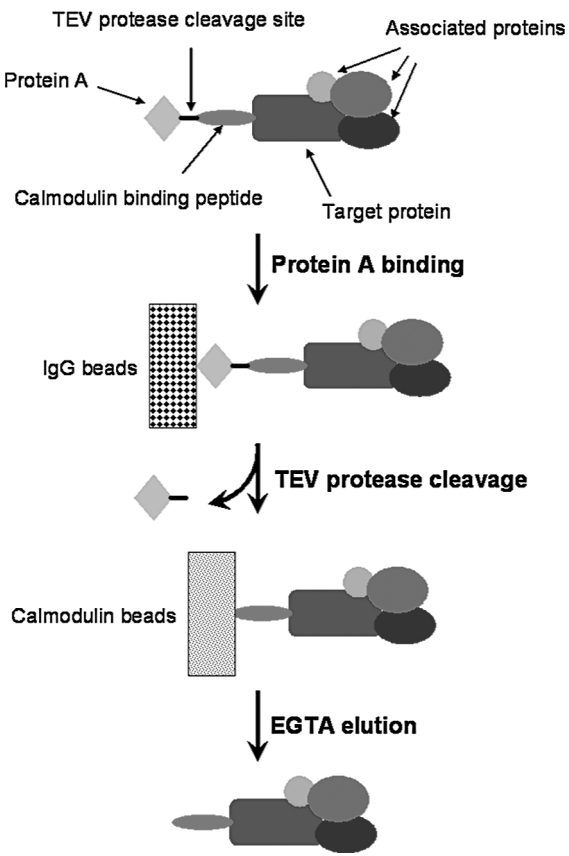


Figure 8 Overview of the TAP strategy. *Source*: Ref. 138.

many ways to Gavin's approach, elucidated similar protein complexes and networks. Many of the identified yeast proteins and protein complexes have human homologs, emphasizing the relevance to human disease and drug discovery.

3.2.2. Yeast 2-Hybrid

The yeast 2-hybrid system is a unique genetic-based approach for analyzing protein–protein, protein–nucleic acid, and protein–ligand interactions (116,117). Basically, a DNA binding domain and an activation domain that normally reside together in a single transcription factor are cloned into separate vectors. The bait, or the protein of interest, usually is cloned as a fusion protein into the binding domain vector. The prey consists of all members of a cDNA library that are cloned as fusion proteins into the activation domain vector. When the bait and prey proteins interact, the binding and activation domains of the original transcription factor are brought close enough together to promote transcription, and yeast clones containing the interacting proteins are selected. A detailed description of the yeast 2-hybrid system can be found elsewhere (118).

Using the yeast 2-hybrid system, Uetz et al. (119) and Ito et al. (120) determined several thousand interactions between proteins in yeast. These were direct interactions between protein pairs, making this information complementary to the protein interaction data that is inferred from having several proteins present in the same complex (see Sec. 2.2.1). A similar approach was used recently, as part of a comprehensive proteomic analysis of human breast cancer cell membranes (121). Tumor cell-derived membrane proteins were isolated, separated on 1D gels, digested with trypsin, and identified by MS. Nearly one-third of the 501 proteins identified were previously uncharacterized (e.g., ESTs of unknown function). Three of these uncharacterized proteins were analyzed further for interacting protein partners using the yeast 2-hybrid system. All three were shown to interact with known (i.e., well-characterized) proteins, thus leading to a clearer understanding of their EST-encoded partners and their potential roles in cancer.

One problem associated with the yeast 2-hybrid system is that protein interactions that are dependent on posttranslational modifications that can occur only in mammalian cells, would not be detected. For example, monoubiquitinylation and serine phosphorylation of a Fanconi anemia complementation group D2 protein (FANC-D2) are both required for mediating cellular resistance to ionizing radiation (122), an interaction that would be missed using this approach.

3.2.3. Biomolecular Interaction Analysis

Another commonly used method for protein–protein interaction studies is BIA-MS (123,124). Surface plasmon resonance (SPR) based biomolecular

interaction analysis (BIA) in combination with MALDI-TOF-MS is an excellent application in the real-time investigation of biomolecular interaction events (125), and represents an extremely powerful approach for the analysis of biomolecular recognition in drug development (126). The general concept of such analysis (BIA-MS) is illustrated in Figure 9. In the simplest form of BIA-MS, SPR-BIA is used to monitor an affinity interaction occurring on the surface of the biosensor chip. The association and dissociation kinetics of the interaction are determined and used to derive a dissociation constant for the event. After SPR-BIA analysis, MALDI-TOF is used to detect the presence of the target ligand(s) on the biosensor chip. In situations where more than just the targeted ligand is detected, MALDI-TOF is used to determine the identity of the unknown components that are retained

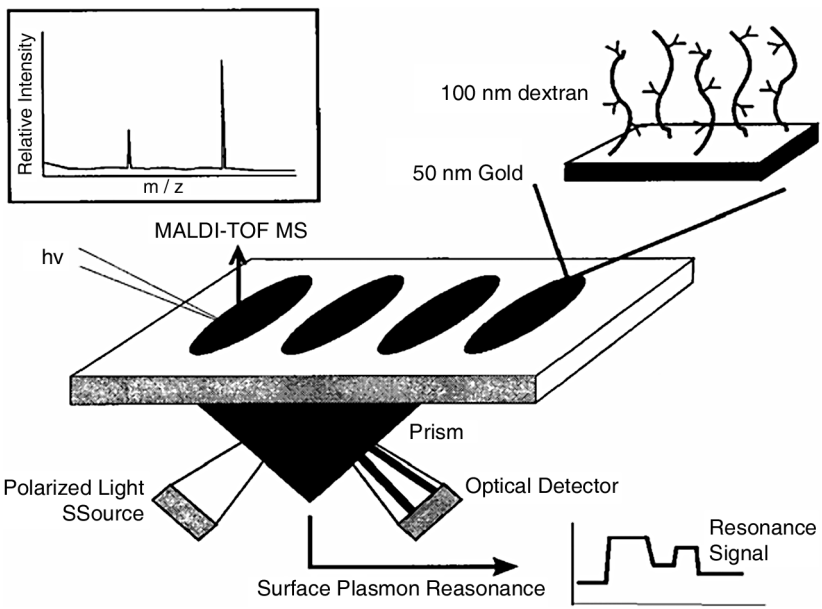


Figure 9 Illustration of the combined SPR-based BIA/MS approach (139). Derivatized biosensor chips, having multiple (2–4) flow cells each, are used in the real-time SPR-BIA analysis of interactions between surface-bound receptors and solution-phase ligands. The sensor chips are removed from the biosensor after SPR-BIA, with ligands still retained within the flow cells, and prepared for MALDI-TOF by application of an appropriate matrix to the flow cells. The matrix solution disrupts the receptor–ligand interaction, liberating the ligand into solution for incorporation into the matrix crystals. With proper application of the matrix, the crystals settle onto the original location of the interaction and spatial resolution between flow cells is preserved. The flow cells are targeted individually during MALDI-TOF and the retained ligand(s) are detected at precise and characteristic m/z values.

through either non-specific interaction with the sensor chip, or through specific interaction with immobilized receptor. The BIA-MS has been applied to antigen/antibody interaction mapping (127), ligand fishing (128), monitoring the actions of enzymes (129), and drug screening (130).

3.2.4. Organellar Proteomics

Given the complexity of eukaryotic cells, it is exceedingly unlikely that any current proteomic methods can separate and analyze the tens (or hundreds) of thousands of proteins and their modified forms that are present in a cell. As an alternative, one can exploit the fact that cells are highly compartmentalized. Animal cells are divided into compartments called organelles that include the mitochondrion, lysozome, phagosome, endoplasmic reticulum (ER), golgi, nucleus, nucleolus, and others. Using various biochemical, centrifugation, and electrophoretic methods, these organelles can be purified or greatly enriched, generally giving at least a 10-fold purification. The proteomes of many of these organelles from several species have been analyzed (131).

For example, the protein composition of phagosomes and phagosomal membranes has been examined using 2DE and MS (132,133). Through the identification of over 500 proteins it was noted that numerous proteins previously thought to be associated with the ER also were found in phagosomes. After ruling out cross-contamination, a detailed model for phagolysozome biogenesis in which the ER is the membrane source during phagocytosis.

The human nucleolar proteome also has been characterized (134–136). The results of these studies portray this organelle as extremely metabolically active, with a role in RNA transport and modification, and cell cycle signaling.

Finally, a clever use of stable isotope labeling by amino acids in cell culture (SILAC) was used to analyze the difficult-to-isolate lipid raft proteome, representing the first functional proteomic analysis of these physiologically significant membrane components (137). In essence, these investigators used quantitative MS by encoding all of the proteins in one of two HeLa cell populations by metabolically labeling with deuterium-substituted leucine. One of the populations was treated with nystatin or methyl-β-cyclodextrin (MβCD) to disrupt the lipid rafts in those cells and then the treated and untreated cells were then combined and biochemically fractionated. Because of raft disruption, labeled lipid-raft proteins from that population were absent (or severely reduced) in the analyzed fraction, as reflected in the quantitative isotopic ratios between treated and untreated peptides. Peptides (and corresponding identified proteins) with large ratios thus comprised the lipid-raft proteome, while other, non-raft proteins were present in roughly equal proportions.

4. CONCLUSION

This moderately brief look at proteomics reflects the technological diversity that characterizes this new science and the accompanying examples of where the various tools of proteomics have been applied demonstrate that "no one size fits all." The strengths and weaknesses of each approach necessitate clever experimental design and application on the part of the researcher. No single proteomic technique or approach is universally applicable. On the contrary, choosing a proteomic approach to better understand physiological function, disease, or efficacious and adverse drug effects relies on the careful implementation of various combinations of these techniques. Their application to pharmaceutical discovery in animals and humans, while fraught with complexity and labor, is sure to be successful as the growing proteomic literature attests.

REFERENCES

1. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 1997; 18:533–537.
2. Tew KD, Monks A, Barone L, Rosser D, Akerman G, Montali JA, Wheatley JB, Schmidt DE Jr. Glutathione-associated enzymes in the human cell lines of the National Cancer Institute Drug Screening Program. Mol Pharmacol 1996; 50:149–159.
3. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999; 19:1720–1730.
4. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001; 292:929–934.
5. Moller A, Soldan M, Volker U, Maser E. Two-dimensional gel electrophoresis: a powerful method to elucidate cellular responses to toxic compounds. Toxicology 2001; 160:129–138.
6. He QY, Chiu JF. Proteomics in biomarker discovery and drug development. J Cell Biochem 2003; 89:868–886.
7. Jain KK. Proteomics and drug discovery. Contrib Nephrol 2004; 141:308–327.
8. Wilson JF. The promise of disease proteomics: faster detection, diagnosis, and drug development. Ann Intern Med 2004; 140:317–319.
9. Guttman A, Varoglu M, Khandurina J. Multidimensional separations in the pharmaceutical arena. Drug Discov Today 2004; 9:136–144.
10. Witzmann FA, Grant RA. Pharmacoproteomics in drug development. Pharmacogenomics J 2003; 3:69–76.
11. Voshol H, Glucksman MJ, van Oostrum J. Proteomics in the discovery of new therapeutic targets for psychiatric disease. Curr Mol Med 2003; 3:447–458.
12. Kumble KD. Protein microarrays: new tools for pharmaceutical development. Anal Bioanal Chem 2003; 377:812–819.
13. Hewick RM, Lu Z, Wang JH. Proteomics in drug discovery. Adv Protein Chem 2003; 65:309–342.

14. Parsons AB, Geyer R, Hughes TR, Boone C. Yeast genomics and proteomics in drug discovery and target validation. Prog Cell Cycle Res 2003; 5:159–166.
15. Valle RP, Jendoubi M. Antibody-based technologies for target discovery. Curr Opin Drug Discov Devel 2003; 6:197–203.
16. Hanash S. Disease proteomics. Nature 2003; 422:226–232.
17. Weigl BH, Bardell RL, Cabrera CR. Lab-on-a-chip for drug development. Adv Drug Deliv Rev 2003; 55:349–377.
18. Jeffery DA, Bogyo M. Chemical proteomics and its application to drug discovery. Curr Opin Biotechnol 2003; 14:87–95.
19. Krieg RC, Paweletz CP, Liotta LA, Petricoin EF 3rd. Clinical proteomics for cancer biomarker discovery and therapeutic targeting. Technol Cancer Res Treat 2002; 1:263–272.
20. Sali A, Glaeser R, Earnest T, Baumeister W. From words to literature in structural proteomics. Nature 2003; 422:216–225.
21. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 2001; 19:242–247.
22. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002; 1:845–867.
23. Molloy MP, Herbert BR, Walsh BJ, Tyler MI, Traini M, Sanchez JC, Hochstrasser DF, Williams KL, Gooley AA. Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. Electrophoresis 1998; 19:837–844.
24. Herbert B, Righetti PG. A turning point in proteome analysis: sample prefractionation via multicompartment electrolyzers with isoelectric membranes. Electrophoresis 2000; 21:3639–3648.
25. Cordwell SJ, Nouwens AS, Verrills NM, Basseal DJ, Walsh BJ. Subproteomics based upon protein cellular location and relative solubilities in conjunction with composite two-dimensional electrophoresis gels. Electrophoresis 2000; 21:1094–1103.
26. Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS. Prefractionation techniques in proteome analysis. Proteomics 2003; 3:1397–1407.
27. Jiang L, He L, Fountoulakis M. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. J Chromatogr A 2004; 1023:317–320.
28. Krapfenbauer K, Fountoulakis M, Lubec G. A rat brain protein expression map including cytosolic and enriched mitochondrial and microsomal fractions. Electrophoresis 2003; 24:1847–1870.
29. Fountoulakis M, Berndt P, Langen H, Suter L. The rat liver mitochondrial proteins. Electrophoresis 2002; 23:311–328.
30. Fountoulakis M, Schlaeger EJ. The mitochondrial proteins of the neuroblastoma cell line IMR-32. Electrophoresis 2003; 24:260–275.
31. Shang TQ, Ginter JM, Johnston MV, Larsen BS, McEwen CN. Carrier ampholyte-free solution isoelectric focusing as a prefractionation method for the proteomic analysis of complex protein mixtures. Electrophoresis 2003; 24:2359–2368.

32. Zuo X, Speicher DW. Comprehensive analysis of complex proteomes using microscale solution isoelectrofocusing prior to narrow pH range two-dimensional electrophoresis. Proteomics 2002; 2:58–68.
33. Anderson NL, Copple DC, Bendele RA, Probst GS, Richardson FC. Covalent protein modifications and gene expression changes in rodent liver following administration of methapyrilene: a study using two-dimensional electrophoresis. Fundam Appl Toxicol 1992; 18:570–580.
34. Richardson FC, Strom SC, Copple DM, Bendele RA, Probst GS, Anderson NL. Comparisons of protein changes in human and rodent hepatocytes induced by the rat-specific carcinogen, methapyrilene. Electrophoresis 1993; 14:157–161.
35. Richardson FC, Horn DM, Anderson NL. Dose-responses in rat hepatic protein modification and expression following exposure to the rat hepatocarcinogen methapyrilene. Carcinogenesis 1994; 15:325–329.
36. Steinberg RA, Coffino P. Two-dimensional gel analysis of cyclic AMP effects in cultured S49 mouse lymphoma cells: protein modifications, inductions and repressions. Cell 1979; 18:719–733.
37. Witzmann FA, Carpenter RL, Ritchie GD, Wilson CL, Nordholm AF, Rossi J 3rd. Toxicity of chemical mixtures: proteomic analysis of persisting liver and kidney protein alterations induced by repeated exposure of rats to JP-8 jet fuel vapor. Electrophoresis 2000; 21:2138–2147.
38. Witzmann FA, Daggett DA, Fultz CD, Nelson SA, Wright LS, Kornguth SE, Siegel FL. Glutathione S-transferases: two-dimensional electrophoretic protein markers of lead exposure. Electrophoresis 1998; 19:1332–1335.
39. Witzmann FA, Fultz CD, Mangipudy RS, Mehendale HM. Two-dimensional electrophoretic analysis of compartment-specific hepatic protein charge modification induced by thioacetamide exposure in rats. Fundam Appl Toxicol 1996; 31:124–132.
40. Witzmann FA, Jarnot BM, Clack JW. Charge modification in rodent hepatic Grp78/BiP following exposure to structurally diverse peroxisome proliferators. Appl Theor Electrophor 1994; 4:81–88.
41. Witzmann FA, Jarnot BM, Parker DN, Clack JW. Modification of hepatic immunoglobulin heavy chain binding protein (BiP/Grp78) following exposure to structurally diverse peroxisome proliferators. Fundam Appl Toxicol 1994; 23:1–8.
42. Silva CM, Cidlowski JA. The effect of oxidation/reduction on the charge heterogeneity of the human glucocorticoid receptor. J Steroid Biochem Mol Biol 1992; 41:1–10.
43. Jenkins LW, Peters GW, Dixon CE, Zhang X, Clark RS, Skinner JC, Marion DW, Adelson PD, Kochanek PM. Conventional and functional proteomics using large format two-dimensional gel electrophoresis 24 hours after controlled cortical impact in postnatal day 17 rats. J Neurotrauma 2002; 19:715–740.
44. Paron I, D'Elia A, D'Ambrosio C, Scaloni A, D'Aurizio F, Prescott A, Damante G, Tell G. A proteomic approach to identify early molecular targets of oxidative stress in human epithelial lens cells. Biochem J 2004; 378:929–937.

45. Aicher L, Wahl D, Arce A, Grenet O, Steiner S. New insights into cyclosporine A nephrotoxicity by proteome analysis. Electrophoresis 1998; 19:1998–2003.
46. Steiner S, Aicher L, Raymackers J, Meheus L, Esquer-Blasco R, Anderson NL, Cordier A. Cyclosporine A decreases the protein level of the calcium-binding protein calbindin-D 28 kDa in rat kidney. Biochem Pharmacol 1996; 51:253–258.
47. Aicher L, Meier G, Norcross AJ, Jakubowski J, Varela MC, Cordier A, Steiner S. Decrease in kidney calbindin-D 28 kDa as a possible mechanism mediating cyclosporine A- and FK-506-induced calciuria and tubular mineralization. Biochem Pharmacol 1997; 53:723–731.
48. Varela MC, Arce A, Greiner B, Schwald M, Aicher L, Wahl D, Grenet O, Steiner S. Cyclosporine A-induced decrease in calbindin-D 28 kDa in rat kidney but not in cerebral cortex and cerebellum. Biochem Pharmacol 1998; 55:2043–2046.
49. Newsholme SJ, Maleeff BF, Steiner S, Anderson NL, Schwartz LW. Two-dimensional electrophoresis of liver proteins: characterization of a drug-induced hepatomegaly in rats. Electrophoresis 2000; 21:2122–2128.
50. Steiner S, Gatlin CL, Lennon JJ, McGrath AM, Seonarain MD, Makusky AJ, Aponte AM, Esquer-Blasco R, Anderson NL. Cholesterol biosynthesis regulation and protein changes in rat liver following treatment with fluvastatin. Toxicol Lett 2001; 120:369–377.
51. Steiner S, Gatlin CL, Lennon JJ, McGrath AM, Aponte AM, Makusky AJ, Rohrs MC, Anderson NL. Proteomics to display lovastatin-induced protein and pathway regulation in rat liver. Electrophoresis 2000; 21:2129–2137.
52. Hu W, Wu W, Verschraegen CF, Chen L, Mao L, Yeung SC, Kudelka AP, Freedman RS, Kavanagh JJ. Proteomic identification of heat shock protein 70 as a candidate target for enhancing apoptosis induced by farnesyl transferase inhibitor. Proteomics 2003; 3:1904–1911.
53. Hu W, Wu W, Yeung SC, Freedman RS, Kavanagh JJ, Verschraegen CF. Increased expression of heat shock protein 70 in adherent ovarian cancer and mesothelioma following treatment with manumycin, a farnesyl transferase inhibitor. Anticancer Res 2002; 22:665–672.
54. Moller A, Malerczyk C, Volker U, Stoppler H, Maser E. Monitoring daunorubicin-induced alterations in protein expression in pancreas carcinoma cells by two-dimensional gel electrophoresis. Proteomics 2002; 2:697–705.
55. Leak LV, Petricoin EF, Liotta LA, Jones M. Proteomic technologies to study lymphangiogenesis. FASEB J 2003; 17:A554–A554.
56. Jones MB, Krutzsch H, Shu HJ, Zhao YM, Liotta LA, Kohn EC, Petricoin EF. Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. Proteomics 2002; 2:76–84.
57. Bajzer Z, Randic M, Plavsic D, Basak SC. Novel map descriptors for characterization of toxic effects in proteomics maps. J Mol Graph Model 2003; 22:1–9.
58. Charlwood J, Skehel JM, King N, Camilleri P, Lord P, Bugelski P, Atif U. Proteomic analysis of rat kidney cortex following treatment with gentamicin. J Proteome Res 2002; 1:73–82.

59. Cottrell JS. Protein identification by peptide mass fingerprinting. Pept Res 1994; 7:115–124.
60. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999; 20:3551–3567.
61. Eng JK, McCormack AL, Yates I, John R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 1994; 5:976–989.
62. Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Boucherie H, Mann M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci USA 1996; 93:14440–14445.
63. Keough T, Lacey MP, Fieno AM, Grant RA, Sun Y, Bauer MD, Begley KB. Tandem mass spectrometry methods for definitive protein identification in proteomics research. Electrophoresis 2000; 21:2252–2265.
64. Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. Anal Chem 2000; 72:552–558.
65. McDonald WH, Yates JR 3rd. Shotgun proteomics and biomarker discovery. Dis Markers 2002; 18:99–105.
66. Smolka M, Zhou H, Aebersold R. Quantitative protein profiling using two-dimensional gel electrophoresis, isotope-coded affinity tag labeling, and mass spectrometry. Mol Cell Proteomics 2002; 1:19–29.
67. Smolka MB, Zhou H, Purkayastha S, Aebersold R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. Anal Biochem 2001; 297:25–31.
68. Griffin TJ, Han DK, Gygi SP, Rist B, Lee H, Aebersold R, Parker KC. Toward a high-throughput approach to quantitative proteomic analysis: expression-dependent protein identification by mass spectrometry. J Am Soc Mass Spectrom 2001; 12:1238–1246.
69. Griffin TJ, Lock CM, Li XJ, Patel A, Chervestsova I, Lee H, Wright ME, Ranish JA, Chen SS, Aebersold R. Abundance ratio-dependent proteomic analysis by mass spectrometry. Anal Chem 2003; 75:867–874.
70. Arnott D, Kishiyama A, Luis EA, Ludlum SG, Marsters JC Jr, Stults JT. Selective detection of membrane proteins without antibodies: a mass spectrometric version of the Western blot. Mol Cell Proteomics 2002; 1:148–156.
71. Zhou H, Ranish JA, Watts JD, Aebersold R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. Nat Biotechnol 2002; 20:512–515.
72. Hansen KC, Schmitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL. Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-Isotope-coded affinity tag and multidimensional chromatography. Mol Cell Proteomics 2003; 2:299–314.
73. Bottari P, Aebersold R, Turecek F, Gelb MH. Design and synthesis of visible isotope-coded affinity tags for the absolute quantification of specific proteins in complex mixtures. Bioconjug Chem 2004; 15:380–388.

74. Olsen JV, Andersen JR, Nielsen PA, Nielsen ML, Figeys D, Mann M, Wisniewski JR. HysTag—a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. Mol Cell Proteomics 2004; 3:82–92.
75. Sethuraman M, McComb ME, Heibeck T, Costello CE, Cohen RA. Isotope-coded affinity tag approach to identify and quantify oxidant-sensitive protein thiols. Mol Cell Proteomics 2004; 3:273–278.
76. Li C, Hong Y, Tan YX, Zhou H, Ai JH, Li SJ, Zhang L, Xia QC, Wu JR, Wang HY, Zeng R. Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. Mol Cell Proteomics 2004; 3:399–409.
77. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol 1999; 17:676–682.
78. Wu CC, Yates JR 3rd. The application of mass spectrometry to membrane proteomics. Nat Biotechnol 2003; 21:262–267.
79. MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A, Clark JI, Yates JR 3rd. Shotgun identification of protein modifications from protein complexes and lens tissue. Proc Natl Acad Sci USA 2002; 99:7900–7905.
80. MacCoss MJ, Wu CC, Liu H, Sadygov R, Yates JR 3rd. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. Anal Chem 2003; 75:6912–6921.
81. He QY, Yip TT, Li M, Chiu JF. Proteomic analyses of arsenic-induced cell transformation with SELDI-TOF ProteinChip technology. J Cell Biochem 2003; 88:1–8.
82. Dare TO, Davies HA, Turton JA, Lomas L, Williams TC, York MJ. Application of surface-enhanced laser desorption/ionization technology to the detection and identification of urinary parvalbumin-alpha: a biomarker of compound-induced skeletal muscle toxicity in the rat. Electrophoresis 2002; 23:3241–3251.
83. Austen BM, Frears ER, Davies H. The use of solid ProteinChip arrays to monitor production of Alzheimer's beta amyloid in transfected cells. J Pept Sci 2000; 6:459–469.
84. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002; 359:572–577.
85. Rosty C, Christa L, Kuzdzal S, Baldwin WM, Zahurak ML, Carnot F, Chan DW, Canto M, Lillemoe KD, Cameron JL, Yeo CJ, Hruban RH, Goggins M. Identification of hepatocarcinoma–intestine–pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. Cancer Res 2002; 62:1868–1875.
86. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004; 20:777–785.

87. Weinberger SR, Viner RI, Ho P. Tagless extraction-retentate chromatography: a new global protein digestion strategy for monitoring differential protein expression. Electrophoresis 2002; 23:3182–3192.
88. Thulasiraman V, Wang Z, Katrekar A, Lomas L, Yip TT. Simultaneous monitoring of multiple kinase activities by SELDI-TOF mass spectrometry. Methods Mol Biol 2004; 278:205–214.
89. Cordingley HC, Roberts SL, Tooke P, Armitage JR, Lane PW, Wu W, Wildsmith SE. Multifactorial screening design and analysis of SELDI-TOF ProteinChip array optimization experiments. Biotechniques 2003; 34:364–365, 368–373.
90. Wiesner A. Detection of tumor markers with ProteinChip(R) technology. Curr Pharm Biotechnol 2004; 5:45–67.
91. Petricoin EF, Liotta LA. Clinical applications of proteomics. J Nutr 2003; 133:2476S–2484S.
92. Reddy G, Dalmasso EA. SELDI ProteinChip(R) array technology: protein-based predictive medicine and drug discovery applications. J Biomed Biotechnol 2003; 2003:237–241.
93. Schweitzer B, Predki P, Snyder M. Microarrays to characterize protein interactions on a whole-proteome scale. Proteomics 2003; 3:2190–2199.
94. Templin MF, Stoll D, Schwenk JM, Potz O, Kramer S, Joos TO. Protein microarrays: promising tools for proteomic research. Proteomics 2003; 3:2155–2166.
95. Wiese R, Belosludtsev Y, Powdrill T, Thompson P, Hogan M. Simultaneous multianalyte ELISA performed on a microarray platform. Clin Chem 2001; 47:1451–1457.
96. Zhu H, Snyder M. Protein arrays and microarrays. Curr Opin Chem Biol 2001; 5:40–45.
97. Turtinen LW, Prall DN, Bremer LA, Nauss RE, Hartsel SC. Antibody array-generated profiles of cytokine release from THP-1 leukemic monocytes exposed to different amphotericin B formulations. Antimicrob Agents Chemother 2004; 48:396–403.
98. Zhu H, Snyder M. ''Omic'' approaches for unraveling signaling networks. ''Omic'' approaches for unraveling signaling networks. Curr Opin Cell Biol 2002; 14:173–179.
99. Kafadar KA, Zhu H, Snyder M, Cyert MS. Negative regulation of calcineurin signaling by Hrr25p, a yeast homolog of casein kinase I. Genes Dev 2003; 17:2698–2708.
100. Lee Y, Lee EK, Cho YW, Matsui T, Kang IC, Kim TS, Han MH. ProteoChip: a highly sensitive protein microarray prepared by a novel method of protein immobilization for application of protein–protein interaction studies. Proteomics 2003; 3:2289–2304.
101. Angenendt P, Nyarsik L, Szaflarski W, Glokler J, Nierhaus KH, Lehrach H, Cahill DJ, Lueking A. Cell-free protein expression and functional assay in nanowell chip format. Anal Chem 2004; 76:1844–1849.
102. Haab BB, Dunham MJ, Brown PO. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol 2001; 2:RESEARCH0004.

103. Kukar T, Eckenrode S, Gu Y, Lian W, Megginson M, She JX, Wu D. Protein microarrays to detect protein–protein interactions using red and green fluorescent proteins. Anal Biochem 2002; 306:50–54.
104. Cahill DJ. Protein and antibody arrays and their medical applications. J Immunol Methods 2001; 250:81–91.
105. Cravatt BF, Sorensen EJ. Chemical strategies for the global analysis of protein function. Curr Opin Chem Biol 2000; 4:663–668.
106. Patricelli MP, Giang DK, Stamp LM, Burbaum JJ. Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes. Proteomics 2001; 1:1067–1071.
107. Adam GC, Cravatt BF, Sorensen EJ. Profiling the specific reactivity of the proteome with non-directed activity-based probes. Chem Biol 2001; 8:81–95.
108. Jessani N, Liu Y, Humphrey M, Cravatt BF. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. Proc Natl Acad Sci USA 2002; 99:10335–10340.
109. Greenbaum D, Medzihradszky KF, Burlingame A, Bogyo M. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. Chem Biol 2000; 7:569–581.
110. Parang K, Kohn JA, Saldanha SA, Cole PA. Development of photo-crosslinking reagents for protein kinase–substrate interactions. FEBS Lett 2002; 520:156–160.
111. Leung D, Hardouin C, Boger DL, Cravatt BF. Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. Nat Biotechnol 2003; 21:687–691.
112. Greenbaum D, Baruch A, Hayrapetian L, Darula Z, Burlingame A, Medzihradszky KF, Bogyo M. Chemical approaches for functionally probing the proteome. Mol Cell Proteomics 2002; 1:60–68.
113. Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell 1998; 92:291–294.
114. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002; 415:141–147.
115. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002; 415:180–183.

116. Fields S, Bartel PL. The two-hybrid system. A personal view. Methods Mol Biol 2001; 177:3–8.
117. Fields S, Song O. A novel genetic system to detect protein–protein interactions. Nature 1989; 340:245–246.
118. Topcu Z, Borden KL. The yeast two-hybrid system and its pharmaceutical significance. Pharm Res 2000; 17:1049–1055.
119. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 2000; 403:623–627.
120. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 2001; 98:4569–4574.
121. Adam PJ, Boyd R, Tyson KL, Fletcher GC, Stamps A, Hudson L, Poyser HR, Redpath N, Griffiths M, Steers G, Harris AL, Patel S, Berry J, Loader JA, Townsend RR, Daviet L, Legrain P, Parekh R, Terrett JA. Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. J Biol Chem 2003; 278:6482–6489.
122. Taniguchi T, Garcia-Higuera I, Xu B, Andreassen PR, Gregory RC, Kim ST, Lane WS, Kastan MB, D'Andrea AD. Convergence of the Fanconi anemia and ataxia telangiectasia signaling pathways. Cell 2002; 109:459–472.
123. Szabo A, Stolz L, Granzow R. Surface plasmon resonance and its use in biomolecular interaction analysis (BIA). Curr Opin Struct Biol 1995; 5:699–705.
124. Myszka DG. Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. Curr Opin Biotechnol 1997; 8:50–57.
125. Nelson RW, Krone JR. Advances in surface plasmon resonance biomolecular interaction analysis mass spectrometry (BIA/MS). J Mol Recognit 1999; 12:77–93.
126. Gambari R. Biospecific interaction analysis: a tool for drug discovery and development. Am J Pharmacogenomics 2001; 1:119–135.
127. Malmqvist M. Epitope mapping by label-free biomolecular interaction analysis. Methods 1996; 9:525–532.
128. Nedelkov D, Rasooly A, Nelson RW. Multitoxin biosensor-mass spectrometry analysis: a new approach for rapid, real-time, sensitive analysis of staphylococcal toxins in food. Int J Food Microbiol 2000; 60:1–13.
129. Buckle M, Williams RM, Negroni M, Buc H. Real time measurements of elongation by a reverse transcriptase using surface plasmon resonance. Proc Natl Acad Sci USA 1996; 93:889–894.
130. Taremi SS, Prosise WW, Rajan N, O'Donnell RA, Le HV. Human interleukin 4 receptor complex: neutralization effect of two monoclonal antibodies. Biochemistry 1996; 35:2322–2331.
131. Jung E, Heller M, Sanchez JC, Hochstrasser DF. Proteomics meets cell biology: the establishment of subcellular proteomes. Electrophoresis 2000; 21:3369–3377.
132. Gagnon E, Duclos S, Rondeau C, Chevet E, Cameron PH, Steele-Mortimer O, Paiement J, Bergeron JJ, Desjardins M. Endoplasmic reticulum-mediated

phagocytosis is a mechanism of entry into macrophages. Cell 2002; 110: 119–131.
133. Garin J, Diez R, Kieffer S, Dermine JF, Duclos S, Gagnon E, Sadoul R, Rondeau C, Desjardins M. The phagosome proteome: insight into phagosome functions. J Cell Biol 2001; 152:165–180.
134. Scherl A, Coute Y, Deon C, Calle A, Kindbeiter K, Sanchez JC, Greco A, Hochstrasser D, Diaz JJ. Functional proteomic analysis of human nucleolus. Mol Biol Cell 2002; 13:4100–4109.
135. Andersen JS, Lyon CE, Fox AH, Leung AK, Lam, YW, Steen H, Mann M, Lamond Al. Directed proteomic analysis of the human nucleolus. Curr Biol 2002; 12:1–11.
136. Leung AK, Andersen JS, Mann M, Lamond AI. Bioinformatic analysis of the nucleolus. Biochem J 2003; 376:553–569.
137. Foster LJ, De Hoog CL, Mann M. Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. Proc Natl Acad Sci USA 2003; 100:5813–5818.
138. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999; 17:1030–1032.
139. Nelson RW, Krone JR. Advances in surface plasmon resonance biomolecular interaction analysis mass spectrometry (BIA/MS). J Mol Recognit 1999; 12:77–93.
140. Tao WA, Aebersold R. Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. Curr Opin Biotechnol 2003; 14:110–118.

19

Haplotype Structure and Pharmacogenomics

Oliver G. McDonald

Department of Molecular Physiology and Biological Physics, University of Virginia Health Sciences Center, Charlottesville, Virginia, U.S.A.

William E. Evans

St. Jude Children's Research Hospital, Colleges of Pharmacy and Medicine, University of Tennessee, Memphis, Tennessee, U.S.A.

1. INTRODUCTION

The inheritance of genetic information in humans (and other diploid organisms) requires that two copies of each chromosome, one paternal and one maternal, be passed on to offspring at conception. This process effectively combines the autosomal genomic DNA sequence from two genetically different parents into their single-diploid progeny. This reproductive strategy continually produces genetic diversity within and among human populations, because the chromosomal DNA sequences of the parents differ from each other and from others in the species or population at various nucleotide positions.

Genome locations are referred to as "polymorphic" if they are present in the population at a frequency of at least 1%. The results of polymorphism from a genetic standpoint are twofold. First, because the human genome is polymorphic at numerous locations, any two individuals are genetically different at some of these polymorphic nucleotides (i.e., they have different

alleles). Second, because the DNA sequences inherited from the maternal and paternal chromosomes can differ, heterozygosity at polymorphic locations within an individual is not uncommon. These two principles, polymorphism and heterozygosity, give rise to the existence of haplotypes.

Unlike a genotype, which is the identity of a single polymorphic location on both chromosomal alleles, a haplotype is the specific combination of nucleotides present at all of the polymorphic locations within a single chromosomal allele. All of the genetic variation in a population or species can be described simply as the sum of all haplotypes that are present among the individuals of that population or species. Nucleotide differences between these haplotypes are responsible for heterozygous genotypes; they also provide information that is useful in ascertaining the identity or structure of a haplotype. For this reason, haplotyping nucleotide polymorphisms requires two steps: first, identifying the polymorphisms, and second, determining which polymorphisms are allelic to each other. Most polymorphisms in the human genome are single-nucleotide polymorphisms (SNPs); thus, this chapter will concentrate on haplotypes defined by multiple SNPs.

Although the primary tool used in pharmacogenomic association studies has traditionally been the genotyping of SNPs, recent evidence shows that determining haplotypes may be more informative than genotyping single variants (1). Therefore, recently there has been great interest in defining the haplotype structure of the human genome (e.g., the human "HapMap" project). The HapMap project (2) focuses on SNPs that are relatively common among human populations; assessing these SNPs at an appropriate density (i.e., number and position across the human genome) will provide new insights into the polymorphic nature of the human genome.

Because common SNPs are phylogenetically older than rare SNPs, they have arisen from recombination events of ancestral haplotypes (3). Focusing on these common SNPs will thus allow reconstruction of these ancestral haplotypes so that human evolutionary history can be traced. Most important, however, is the fact that using common SNPs to map the structure of human haplotypes will allow identification of haplotypes that constitute the majority (perhaps as much as 90%) of human variation and hence will be the most informative source available for disease and pharmacogenomic association studies.

Recent evidence suggests that genotyping just 6 to 8 "haplotype tag" SNPs per 10–100 kb of genomic DNA may provide enough information to determine an individual's haplotype for that region (3,4). This finding suggests that, once the human HapMap project has been completed, genotyping these haplotype tag SNPs will be the preferred method of haplotyping individual patients for common variations. Indeed, commercially available DNA arrays already permit high-throughput determination of 100,000 SNPs distributed across the human genome, thereby providing coverage of haplotype blocks with a length of approximately 30 kb (5,6).

The first half of this chapter focuses on the large-scale structure of human haplotypes, an explanation of why common SNPs can be used to map this structure, and a clinical example demonstrating how haplotype tag SNPs can provide pharmacogenomic information. The second half of the chapter focuses on the importance of rare SNPs and haplotypes in pharmacogenomics and on strategies for their molecular determination. Clearly, haplotyping common SNPs is the best approach for the human HapMap project. However, using these common SNPs or haplotypes may not always be the best approach for individualizing drug therapy and disease diagnosis for complex traits that are genetically heterogeneous in the human population, because rare haplotypes that contribute to such genetic heterogeneity may not be in linkage disequilibrium (LD) with SNPs used in the human HapMap to construct common haplotypes. Rare SNPs are phylogenetically younger than common SNPs and have most likely arisen by mutation rather than by recombination (3). Thus, they will probably represent no more than 5–10% of human variation, and each may be confined to one racial population. However, these rare SNPs and the haplotypes on which they reside could be important in subsets of ethnic populations whose disease susceptibilities and drug responses differ from those of the general population. The second half of this chapter will therefore focus on this possibility, with a discussion of SNP discovery, a description of methods for determining rare as opposed to common haplotypes, and a clinical example highlighting the importance of rare SNPs in pharmacogenomics.

2. HAPLOTYPES COMPOSED OF COMMON SNPs

The human HapMap project will focus on common SNPs (defined as SNPs with a minor allele frequency of at least 10%) and will thus capture most human haplotype variation. Recent experimental evidence tentatively suggests that most of the common haplotype variation in humans has limited diversity. For example, Drysdale et al. (7) examined 13 SNPs within the β_2-adrenergic receptor gene. This study not only was the first to illustrate that determining haplotype can have greater predictive power for pharmacogenomic association studies than genotyping single SNPs (described in detail later), but it also found that these SNPs were organized into only 12 haplotypes of the possible 8192 haplotype combinations. Of these 12, only five appeared with a frequency of more than 10% in the multiethnic cohorts studied (7). This finding is very encouraging to researchers hoping to use information from the HapMap project for disease and pharmacogenomic association studies, because it indicates that they may need to genotype only a few SNPs within a given genomic region to be able to use the HapMap to infer the haplotype structure of that region in as many as 90% of individuals (4). These SNPs are referred to as haplotype tag SNPs.

Once the human HapMap project has been completed, haplotype tag SNPs will probably be a primary focus of association studies that attempt to link certain haplotypes to phenotypic variations in drug response and disposition. Furthermore, if this approach proves successful, haplotype tag SNPs may then become the focus of pretreatment diagnostic procedures that will use haplotype data to assist in designing drug treatment regimens for individual patients, thereby fulfilling the "bench to bedside" paradigm. To gain a full understanding of how these data can be used, we will first review our current, although limited, understanding of haplotype structure in the human genome.

3. HAPLOTYPE STRUCTURE IN THE HUMAN GENOME

3.1. Haplotype Classification

It is useful to divide haplotypes into two classes. *Global haplotypes* are the specific combinations of nucleotides present along the entire length of a single chromosome; these combinations of nucleotides are typically found in only one human population. *Local haplotype blocks* are the specific combinations of nucleotides present along a shorter region (usually 10–100 kb) within a chromosome; some of these blocks may be found in multiple human populations, whereas others are specific to one population. Thus, global haplotypes are simply a collection of intra-allelic local haplotype structures in tandem from one end of a chromosome to the other. We make the distinction between global haplotypes and local haplotype blocks for two reasons.

The first reason for this distinction is that in two situations haplotype determination provides important assistance in predicting pharmacogenomic traits. The consideration of local haplotypes is presently economically feasible and is sufficient to provide predictive power for simple monogenic traits, which are controlled by a single gene and within a genomic region small enough that determination of global haplotypes is not necessary. Two examples of such traits are the genetic polymorphisms of the thiopurine S-methyltransferase (TPMT) enzyme and of the β-adrenergic receptor; these examples will be discussed in this chapter. Conversely, when accurate and rapid haplotyping of entire chromosomes is economically feasible, determination of global haplotypes will be more informative for cases in which a multitude of genes and polymorphisms that are separated by large distances contribute to a complex, multigenic phenotype, such as the influence of multiple pharmacokinetic or pharmacodynamic processes on drug response (which is typical). Making this classification simply clarifies which haplotyping approach, local or global, is needed in these two situations. Because global haplotyping is currently impossible, most of our discussion will focus on local haplotypes.

The second reason for making the distinction between global haplotypes and local haplotypes is to facilitate our description of LD in the next section. It should be kept in mind that this classification system is an idealized model of human haplotype structure and is based on our current understanding. Although these concepts may be generally applicable to many circumstances, there are undoubtedly principles of human haplotype structure that have yet to be discovered, and the descriptions given here should not be taken as universal in light of the rapidly advancing science of human genome structure.

3.2. Patterns of Linkage Disequilibrium in the Human Genome

Central to the distinction between global and local haplotypes is the concept of LD, which is responsible for the existence of the many blocks of local haplotype structure found within a global haplotype. The SNPs are in LD (or "linked") if they occur together on the same allele at a frequency greater than that which would be expected by chance. If SNPs in LD are found on the same allele, they are said to be "in phase." If they are found on the opposite allele, they are said to be "out of phase." The fact that the SNPs can be in LD but out of phase suggests that the fact that two SNPs are in LD does not mean that they will always occur on the same global or local haplotype of a single individual. Such occurrence would require that they always be in phase. Furthermore, if two SNPs are always in phase, they can be expected not only to be part of the same global haplotype but also to be part of the same local haplotype. Therefore, in a given individual, SNPs that are in phase must be part of the same local or global haplotype structure, but SNPs that are in LD may not be part of the same local or global haplotype.

Furthermore, phasing (i.e., determining the local haplotype) and LD tend to dissipate as a function of increasing genomic distance, and the rate of phase erosion is higher than that of LD erosion. An illustration will clarify this point: we would expect that two common SNPs located 0.5 kb apart will probably be in LD and in phase, that two SNPs located 20 kb apart will probably be in LD but may not be in phase, and that two SNPs located 100 kb apart are probably neither in phase nor in LD. This illustration suggests that the haplotype structure of the human genome is organized into distinct "blocks," each approximately 10–100 kb in length, situated in tandem along chromosomes, and that the combination of these local haplotype blocks forms the global haplotype. Indeed, this concept is supported by several recent studies (3,4,6,7). The fact that LD erodes more slowly than phase suggests that between the point of complete phase dissipation and the point of complete LD dissipation there is LD overlap between two adjacent blocks of local haplotype structure. In other words, some of the SNPs that are detected or observed in this range are part of one local

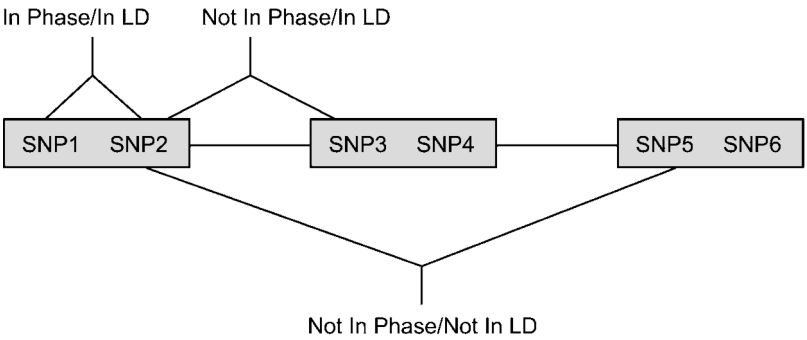


Figure 1 Gray blocks represent local haplotype blocks, and lines connecting the blocks are genomic areas in which recombination hotspots are found. Closely spaced SNPs are probably found on the same block (e.g., SNP1 and SNP2) and therefore are probably in phase and in LD. However, SNPs spaced far apart are probably on adjacent haplotype blocks (e.g., SNP2 and SNP3); there is a low probability that these SNPs are in phase, but there is still a moderate probability that they will be in LD (i.e., the same haplotype blocks are found on the same global haplotype of an individual at least 50% of the time). However, when SNPs are spaced very far apart (e.g., SNP2 and SNP5) they will be separated by more than one complete haplotype block; thus, the probability that they are in LD is greatly reduced.

haplotype block, whereas other SNPs are part of another, adjacent haplotype block. This overlap explains why SNPs separated by these distances are always in LD but may or may not be in phase. These concepts are depicted in Figure 1.

If the human genome is partitioned into haplotype blocks, the definition of LD can be refined. Because a given local haplotype block has multiple variants, the probability that SNPs that are in LD but are not located on the same haplotype block will be in phase (i.e., part of the same local haplotype structure) is a function of increasing distance. For instance, in a hypothetical situation in which two variants or forms are available for every haplotype block, the probability that pairwise SNPs in LD will not be part of the same local haplotype (in phase) is 0.5^{n+1}, where n represents the number of haplotype blocks that separate the two SNPs. For example, if the SNPs are on adjacent blocks (zero block separation), the probability that they will be part of the same local haplotype is 0.5 or 50%. This is the upper limit for SNPs not present on the same haplotype block in this situation. If SNPs are separated by four blocks, the probability that they are part of the same local haplotype falls to 0.03 or 3%, because now there are more blocks and hence more possibilities for variation between the SNPs.

Because LD is defined as the occurrence of two SNPs on an allele at a frequency greater than that predicted by chance, which for two SNPs is $0.5^{1+1} = 0.25$ or 25%, we can modify the definition of LD: LD is the

occurrence on an allele of two SNPs that are not separated by more than one complete haplotype block, because at distances greater than this the probability that two SNPs can exist in the same local haplotype structure is less than that predicted by chance. Using this definition of LD, we find that, because the probability of phasing decreases below that predicted by chance after separation by more than one complete haplotype block, and because LD dissipates after 10–100 kb, the average length of one haplotype block is 10–100 kb, and local haplotype structure is a consideration of this distance. Indeed, recent experimental evidence (3,4,6,7) has strongly suggested that, on average, haplotype blocks span a distance within this range. It should be noted that this distance is only an average and that variability in block length has been observed. For example, blocks as short as 1 kb and as long as 804 kb have been discovered (3,8).

The preceding discussion should not be considered all-inclusive. Rather, it is a simplified version of our current knowledge and is meant to clarify subsequent sections of this chapter. The haplotype block-based model of human LD is based on the concept that most meiotic recombination is not random but occurs across the genome within "recombination hotspots" between haplotype blocks (3). However, recombination rates vary greatly in different regions of the genome, and this model will probably never be able to predict certain elements of LD because it will always include stochastic recombination in areas outside of hotspots. Finally, most of the studies that identified haplotype blocks also showed that at least half of the genomic area investigated did not fall into a discernable block. This finding may be due to the fact that there was a relatively long distance between markers, in these studies, or it may be due to haplotype arrangements that have not yet been elucidated. Hence, much research remains to be done before human haplotype variation is completely understood.

4. METHODS OF DETERMINING COMMON HAPLOTYPES

Until the advent of molecular haplotyping, family pedigree studies were the traditional method of analyzing linkage. However, since the development of molecular biology techniques, especially polymerase chain reaction (PCR) and DNA sequencing, traditional pedigree analysis has largely been replaced by molecular methods that directly examine the DNA sequence. Implementation of these newer DNA-based methods of determining molecular haplotype is possible if two prerequisites are met: (a) SNP discovery and (b) collection of SNP population frequencies. Once these two criteria have been satisfied, multiple SNPs are traditionally genotyped by methods based on PCR or sequencing, and haplotypes are assigned by statistical methods (e.g., Clark's algorithm) derived from the population frequencies. This approach has proved especially useful for local haplotype assignment within genes or within genomic distances that do not eclipse the limits

imposed by LD erosion. Although these methods are straightforward, they have limitations. First, population data are required but have not yet been fully defined. Second, because these methods rely on probabilities, ambiguity arises when the probability that two SNPs lie on the same haplotype approaches 50% (9).

Although most studies have used genotyping and statistics as described above, other studies have used molecular methods that separate parental alleles from each other; such separation allows precise determination of molecular haplotypes without reliance on statistical probability. The PCR amplification of a genomic region is typically followed by cloning the PCR products into a cloning vector and transforming the plasmid vector into competent bacteria. In this way, the DNA sequences from both parental alleles are separated from each other, and haplotypes can be determined unambiguously by sequencing the two alleles. Other methods that are not commonly used are sperm genotyping (10), dilution of genomic DNA to a concentration that carries a high probability of obtaining a hemizygous sample (11), and allele-specific amplification of SNPs that are located in close proximity (less than 1 kb apart) (12). Although these methods can provide unambiguous haplotype assignments, they are generally laborious and may be limited in their ability to haplotype SNPs that are located long distances apart. Thus, these methods are particularly suited for instances in which statistical analysis yields ambiguous haplotype data or in which adequate SNP population frequencies are not available.

Currently, only the local haplotype structure can be determined with the methods described above. However, the human HapMap project may provide new methods of obtaining haplotype data and may also introduce the possibility of determining the global haplotype. Emerging experimental evidence suggests that the recombination frequency across the human genome may not be uniform (13). Rather, the genome is partitioned into local haplotype blocks by unevenly distributed recombination hotspots (13). In addition, evidence indicates that the diversity of these blocks is limited to 2 to 4 very common haplotypes with a combined population frequency as high as 0.9 (4). These two pieces of evidence indicate that mapping human haplotype structure will require only identifying and genotyping SNPs that are phased to each of these haplotypes but not to the others. Identifying and genotyping these "haplotype tag" SNPs is a primary focus of the human HapMap project.

Indeed, one can envision DNA arrays that are coated with oligonucleotides containing these tag SNPs. Once the map has been completed, intense pharmacogenomic association studies will probably be undertaken with the goal of linking these haplotype tag SNPs to phenotypic variations in drug metabolism and disposition. Once this linkage has been established, it will be possible simply to genotype a patient for these tag SNPs and refer back to the HapMap to assign the correct haplotype(s) for a given genomic

locus (or loci). This information not only will quickly and efficiently provides data for tailoring drug therapy (once appropriate association studies are completed) but also will, when combined with DNA arrays, allow economical determination of global haplotype structure (although only a rough estimate).

It must be remembered that molecular methods of determining the global haplotype must be able to provide unambiguous haplotyping of polymorphisms across adjacent haplotype blocks after the array analysis has been completed (Fig. 2). This determination will overcome the limits imposed on other methods by LD erosion, in that haplotype blocks can be connected to each other along the length of a chromosome. With our limited knowledge regarding the function of many areas of the human genome, especially that of intergenic regions, it is not yet clear how determination of global haplotypes will further aid pharmacogenomic studies. However, it is likely that this information will increase the power of haplotype data for pharmacogenomic association studies. For example, recent evidence has indicated that genomes from a variety of organisms, including humans, are organized into very large chromatin domains that contain many genes (and potentially several local haplotype blocks). Expression of these genes is often regulated by DNA sequences within these domains; these sequences are known as locus control regions (LCRs) (14). Heterozygous polymorphisms within LCRs could have a profound impact on the regulation of the genes within this domain. Therefore, it would be necessary to apply the global haplotyping approach to determine which chromosomal haplotype blocks are allelic to the mutated LCR, because these blocks may be expressed at higher or lower levels than haplotype blocks located on the other allele (global haplotype).

5. CLINICAL EXAMPLE OF HAPLOTYPES COMPOSED OF COMMON SNPs

Inhalation of β_2-adrenergic agonists (e.g., albuterol) is a widely used and effective treatment for reversing acute bronchospasm among patients with asthma. This effect is mediated through ligand binding to β_2-adrenergic receptors on the surface of smooth muscle cells in the airway. The β_2-adrenergic receptor is a G-protein-coupled, 7-transmembrane spanning receptor. Upon ligand (e.g., catecholamine) binding, the receptor mediates increased intracellular levels of cAMP by activating adenylate cyclase. The increase in cAMP facilitates the action of myosin light chain phosphatase, which dephosphorylates myosin, thereby disrupting the interactions between myosin and actin and causing relaxation of smooth muscle (15). Because there are marked interindividual phenotypic differences in the degree of airway smooth muscle responses to albuterol (16), it has been suggested that polymorphisms in this receptor might be a principal cause of this variability (7).

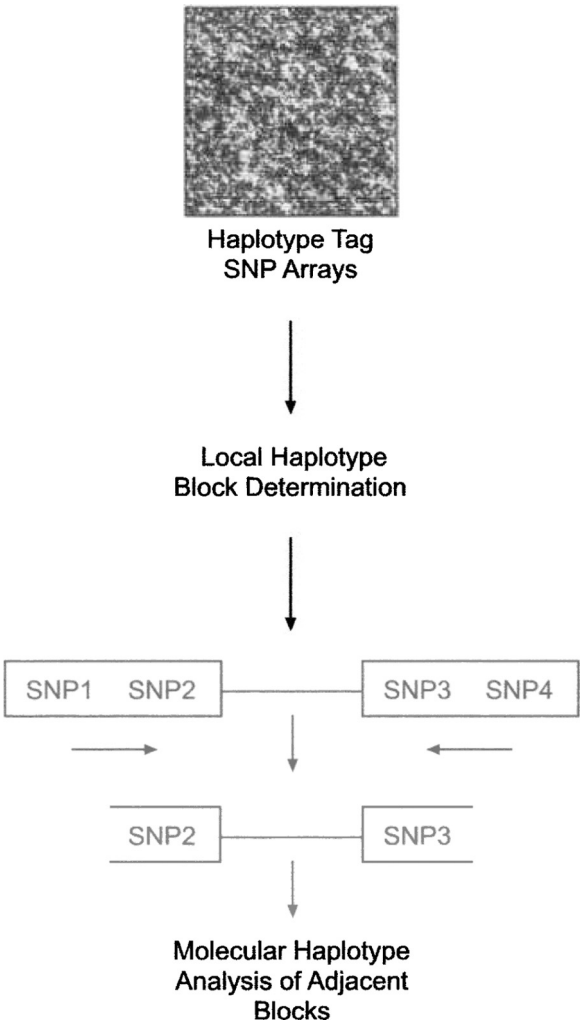


Figure 2 Strategy for obtaining global haplotype information. As a first step, high-throughput SNP detection arrays are used to screen for certain SNPs. The SNPs detected by the array are then haplotyped by conventional methods or, when possible, by referencing the HapMap. This process determines the local haplotypes. Global haplotypes can then be determined by using molecular methods, such as LR-PCR combined with intramolecular ligation, which haplotype SNPs located at the ends of adjacent haplotype blocks (e.g., SNP2 and SNP3).

There are at least 13 known SNPs in the 5′-untranslated region (UTR; i.e., the promoter) and the open reading frame (ORF) of the intronless *ADBR2* gene, which codes for the β_2-adrenergic receptor (1). Several studies have analyzed isolated single SNPs for association with asthma and the response to β_2 agonists. Although some of these studies have demonstrated associations between some SNPs in isolation and the β_2 agonist response, others have yielded mixed results (7). In the first study to assess the predictive power of haplotypes for drug response, Drysdale et al. (7) analyzed the entire 5′-UTR and ORF (1.6 kb) of *ADBR2* for the 13 known SNPs in four different racial cohorts. Before this study, these SNPs had not been assigned to a particular local haplotype. Using both molecular haplotyping (PCR amplification, cloning, and sequencing) and statistics (an extension of Clark's algorithm), these researchers determined the extended haplotype structure of the *ADBR2* gene. Of the 8192 possible haplotypes, only 12 were observed, and five accounted for most of the observed haplotypes. In a separate asthmatic cohort, the presence of these five haplotypes predicted the ability of albuterol to elicit increases in forced expiratory volume in 1 sec (FEV_1) in 121 white patients with asthma. However, Drysdale et al. (7) found no association between the response to albuterol and any individual *ADRB2* SNP in isolation; this finding demonstrated that haplotype data could have more predictive power than genotype data. In retrospect, this predictive power should be expected, because SNPs together on the same allele should interact with each other to influence phenotype, with each SNP potentially exerting its own effect on transcription, translation, protein structure and function, or protein processing, the combination of which ultimately determines the phenotype (Fig. 3). Only haplotype data can predict the combined effects contributed to a given phenotype by multiple SNPs.

6. HAPLOTYPES COMPOSED OF RARE SNPs

As much as 10% of human genetic variation is the result of rare SNPs (i.e., those that occur with a frequency of less than 10%) that are the result of recent mutation events. These SNPs often occur with different frequencies in different racial groups and can even be confined to only one racial group. Because the human HapMap project will not haplotype these SNPs, generating population data that can be used for statistical haplotype analyses will be more difficult and less cost-effective. Thus, it is likely that molecular methods that can unambiguously determine molecular haplotype structure, such as cloning and sequencing, will be necessary for many pharmacogenomic association studies and for pretreatment patient screening. Unless arrays are designed to quickly and cost-effectively screen for such rare SNPs, haplotyping rare SNPs will require considerable time, effort, and expense. However, as we will demonstrate for *TPMT* alleles, failure to

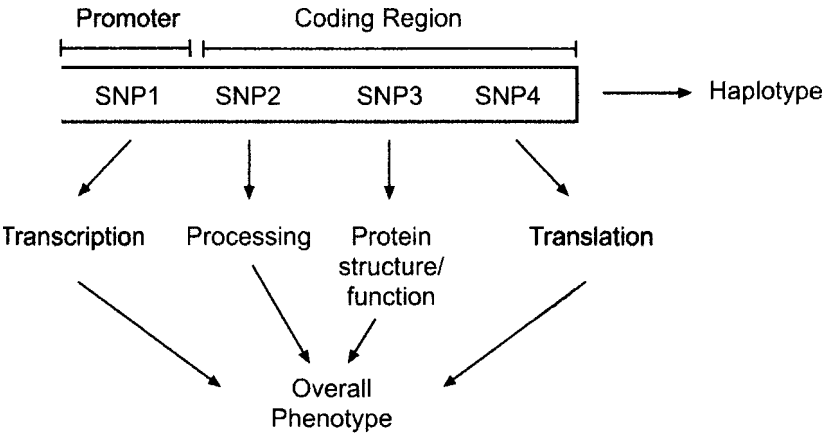


Figure 3 Many SNPs, with different effects on the expression or function of a protein, have cumulative effects on phenotype. Therefore, determining haplotypes can have greater predictive power than analyzing single SNPs.

haplotype certain rare SNPs could have important consequences for individualizing drug therapy.

7. METHODS OF HAPLOTYPING RARE SNPs

As mentioned above, large-scale genotyping followed by statistical algorithms will not be available for most rare SNPs because of the lack of population data and because SNPs will not be included on most haplotyping platforms. In the absence of sufficient population data, ambiguity arises with statistical analysis, making haplotype determination problematic (9). Thus, haplotype determination in these cases will require molecular methods that are not based on LD and statistical analysis. Currently, haplotype determination is straightforward for SNPs that are positioned closely together. However, rare SNPs that are distantly spaced present more of a challenge for haplotype determination, as illustrated below.

The PCR-based techniques are the simplest means of haplotyping closely spaced rare SNPs. Allele-specific amplification (ASA) in isolation or in combination with restriction fragment length polymorphism (RFLP) analysis can quickly haplotype multiple SNPs within a region 1–2 kb in length. The analysis is limited to this distance because the proofreading enzymes that are needed to amplify longer genomic regions will correct the 3′ primer-template mismatch that is necessary for ASA. In addition, although as many as half of all SNPs either create or eliminate a restriction site (17), the probability that a restriction site used for RFLP analysis will be found between SNPs increases with distance; thus, the usefulness of RFLP

in haplotyping distant SNPs is limited. Techniques other than PCR, such as single-strand conformation analysis (SSCP), heteroduplex analysis, and allele-specific oligonucleotide (ASO) ligation and hybridization assays, are also limited in their ability to haplotype distantly placed SNPs (17). Despite these limitations, PCR-based techniques are very useful for closely spaced rare SNPs; therefore, high-throughput assays based on ASA are being designed (18).

Haplotyping rare SNPs located great distances apart (e.g., 5–30 kb) has also been accomplished, again mainly with PCR-based techniques. Long-range PCR (LR-PCR) uses a polymerase with proofreading activity, such as *Pfu* polymerase, in combination with *Taq* polymerase to amplify as much as 30–40 kb of genomic DNA. Usually, a "hot start" (19) or a variation of the hot start (17) is required for the LR-PCR reaction to work. Alleles can be separated by cloning PCR products into a cloning vector that can accommodate such large fragments. This separation is followed by sequencing, which allows determination of the exact haplotype.

We recently developed an alternative method of haplotyping rare SNPs that is not as expensive or laborious as cloning and sequencing (17). Our concept was to create a method that could bring distantly located SNPs into close proximity with each other so that we could use techniques that are commonly reserved for closely spaced SNPs (ASA, RFLP, etc.). Indeed, LR-PCR, with primers placed 5′ to distantly spaced SNPs, followed by intramolecular ligation (circularization) to bring the SNPs into closer proximity allowed us to use techniques applicable to small genomic regions (e.g., ASA, RFLP, ASO, etc.) on these synthetic circular molecules. This approach should also be useful in global haplotype determination, because it would be ideal for haplotype analysis of distal SNPs that are located on adjacent haplotype blocks and that may or may not be in phase (Figs. 1 and 2). Other methods based on diluting genomic DNA to a hemizygous concentration or on sperm haplotyping may also be used to haplotype distal SNPs simply by genotyping hemizygous DNA samples. Regardless of the method used, the most economical approach to haplotyping rare SNPs is first to screen for them in isolation with high-throughput genotyping analysis and then to use molecular haplotyping methods to analyze SNPs that have been detected by screening.

8. CLINICAL EXAMPLE OF HAPLOTYPES COMPOSED OF RARE SNPs

Thiopurines (mercaptopurine, thioguanine, and azathioprine) are important anticancer and anti-inflammatory medications. The use of thiopurines in combination chemotherapy has increased the survival rates of children with acute lymphoblastic leukemia from less than 10% 40 years ago to more than 80% today (20). However, interindividual variation in toleration of

thiopurine medications has been a longstanding obstacle to the effective use of these medications to treat a small percentage of patients who experience severe hematopoietic toxic effects, a condition previously referred to as "idiosyncratic intolerance." It is now recognized that inheritance causes the accumulation of high concentrations of the active thiopurine nucleotides in approximately 10% of patients who are treated with doses that are generally well tolerated by most patients. Furthermore, 0.3% of patients accumulate extremely high levels of the thiopurine nucleotides, and these accumulations cause hematopoietic toxicity that can be lethal (21,22). Thus, there is much interest in elucidating the genetic basis of these phenotypic variations.

The genetic basis of the variations in thiopurine metabolism is polymorphisms within the coding region of the *TPMT* gene. The TPMT is a cytosolic enzyme that catalyzes the transfer of methyl groups from S-adenosyl methionine to the sulfur atom of thioguanine bases and nucleotides, the primary inactivation pathway in hematopoietic tissues (20). Most patients who exhibit deficiencies in TPMT activity have inherited either the *TPMT*3C* allele, which has an A–G transition in exon 10 (position 719 of *TPMT* cDNA), or the *TPMT*3A* allele, which has the A719G polymorphism and a G–A transition in exon 4 (at position 460 of *TPMT* cDNA). These alleles appear at different frequencies in different racial groups (23). Patients heterozygous for these alleles demonstrate intermediate TPMT activity and require intermediate doses of thiopurines. Compared to patients with the wild-type gene (*TPMT*1*), patients with these polymorphisms exhibit enhanced proteolysis of TPMT protein and a decrease in TPMT catalytic activity (24,25). The TMPT activity is very low in patients who are homozygous for these alleles and in those with compound heterozygosity (with both alleles), and these patients require markedly lower doses of thiopurine medications (i.e., approximately 10% of the usual dose). Of the 10% of patients with intermediate TPMT activity, most have inherited one non-functional variant *TPMT* allele and one wild-type allele, whereas the 0.3% of patients who are at danger of lethal hematopoietic toxicity when given standard doses of thiopurines is those who have inherited two non-functional *TPMT* alleles (20). The use of PCR-based genetic testing for these mutant *TPMT* alleles to guide the optimal dosing of thiopurine is one of the most well developed examples of how the bench-to-bedside ideal of pharmacogenetics can be fulfilled (26).

In addition to the two most common *TPMT* alleles discussed above, several other rare SNPs have been discovered in the promoter and coding regions of the *TPMT* gene (20). One of these *TPMT* alleles, *TPMT*3B*, contains the G460A transition in isolation. The proteolysis that this allele undergoes, like that of *TPMT*3A* and *TPMT*3C*, is enhanced in comparison to that undergone by *TPMT*1*. For this reason, patients heterozygous for *TPMT*3B* are expected to exhibit intermediate levels of TPMT activity, whereas patients homozygous for *TPMT*3B* are expected to exhibit low

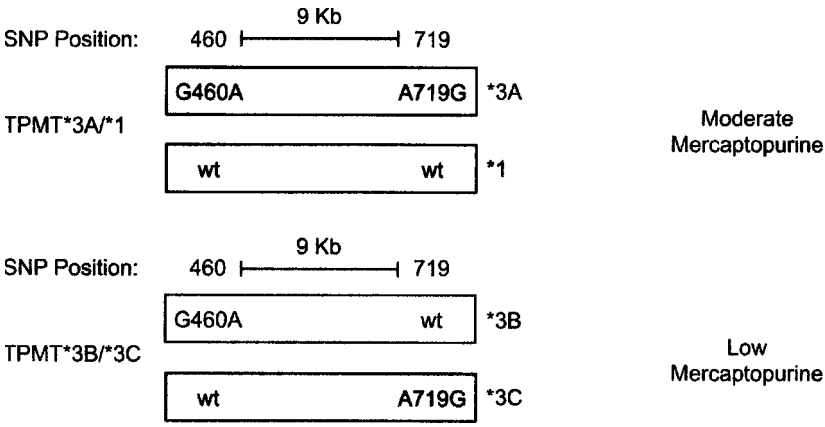


Figure 4 Determination of the *TPMT* haplotype to guide thiopurine (e.g., mercaptopurine) dosing for individual patients. The *TPMT*3A/*1* haplotype results in intermediate levels of TPMT protein activity, whereas the *TPMT*3B/*3C* haplotype results in low levels of TPMT protein activity. Thus, the dosage of mercaptopurine must be adjusted to compensate for these differences in protein activity.

levels of TPMT activity. However, patients who are heterozygous for *TPMT*3C* and *TPMT*3B* (i.e., compound heterozygotes) are also expected to have low TPMT activity, because both of these variant alleles undergo rapid proteolysis (Fig. 4). This expectation poses a problem for genetic testing that can be solved only by determining the molecular haplotype, because PCR-based genotyping assays cannot distinguish between the *TPMT*3A/*1* haplotype and the *TPMT*3B/*3C* haplotype (27). This inability to distinguish the haplotypes is due to the lack of statistically valid population data upon which to base interpretation when both of these nucleotide positions are heterozygous in the same person. The G460A and the A719G positions are separated by approximately 9 kb of genomic DNA, thereby precluding haplotype analysis with methods designed for closely spaced SNPs. Thus, methods such as LR-PCR and cloning or LR-PCR and intramolecular ligation must be used if an absolutely accurate haplotyping method is needed to distinguish between these two phenotypes (intermediate and low TPMT activity), i.e., to determine whether G460A and A719G are on the same allele or on opposite alleles.

9. CONCLUDING REMARKS

Our knowledge about how haplotype analysis may be applied to pharmacogenomics, as well as our understanding of human haplotype structure in general, is in its infancy. In fact, only recently has it been discovered

that haplotype data may have greater predictive power for disease and pharmacogenomic association studies than genotype data. This increased predictive power is due to the fact that the multiple SNPs on the same allele can exert cumulative effects on phenotype, ranging from transcription of the gene at the level of genomic DNA to stability or function of the encoded protein. Because this finding has been only recently recognized in pharmacogenomics, much work remains before haplotype analysis becomes a routine method of elucidating the genetic basis of drug metabolism and disposition. Such work will probably bring together haplotype geneticists and investigators interested in pharmacokinetics and pharmacodynamics, and these collaborations should pave the way for rapid progress.

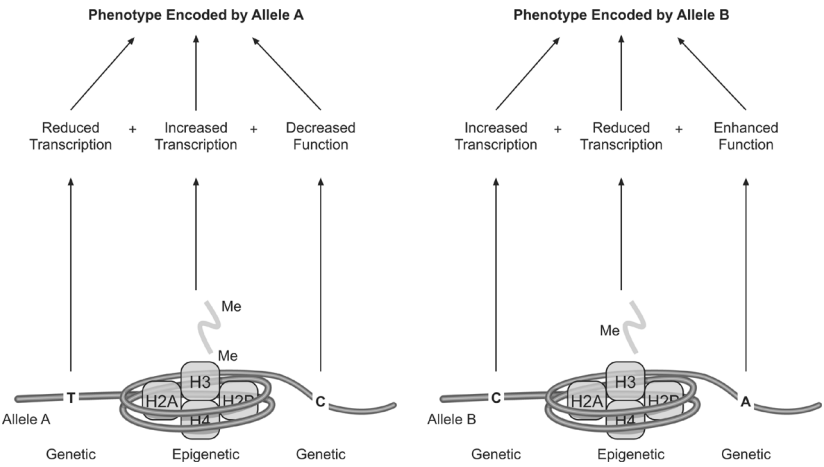


Figure 5 Consideration of hypothetical epigenetic polymorphisms in addition to SNP haplotypes could increase the power of phenotype prediction. In this example, genomic DNA (*black*) with SNPs is wrapped around histone proteins (*dark gray*) with their *N*-terminal tail regions (*light gray*) protruding from the nucleosomal surface. The black Me are methyl groups that are stably bonded to amino acids within histone tails, to cytosine–guanine nucleotides within DNA, or to both. Different combinations of these methyl modifications could in essence exhibit a function similar to that of SNPs, in that different combinations on opposite alleles (haplotypes) could have different effects on phenotype. In this example, the effect is transcription efficiency: consideration of the haplotype alone predicts that allele B will demonstrate greater transcription and expression, and greater protein function than allele A. However, consideration of the epigenetic effects and the haplotype predicts that the two alleles will demonstrate equal transcription, because the effects of the epigenetic polymorphisms on transcription compensate for the effects of the SNPs. This example depicts an epigenetic polymorphism within one individual, but such polymorphisms could also occur between individuals.

Determination of the local haplotype of one or a few genes is currently the most feasible application of haplotype data to pharmacogenomic association studies, as demonstrated by the examples of the β_2-adrenergic receptor and *TPMT*. However, even studying single genes will probably present unforeseen challenges. For example, during the last few years it has become apparent that monoallelic epigenetic modifications can exist at some genes (e.g., genomic imprinting): the maternal allele may contain DNA and histone methylation patterns that are different from those on the paternal allele. These modifications cause these alleles to exhibit different expression patterns (28). Indeed, in the future it may be important to refine our concept of haplotypes beyond DNA sequence variations to include other information such as allelic epigenetic factors, which are inherited through mitosis and meiosis with the DNA itself and which serve to extend the information content of the human genome (29).

Can epigenetic polymorphisms exist on the same haplotype as DNA polymorphisms? Should they also be considered part of the haplotype? If so, will consideration of these epigenetic polymorphisms further increase the predictive power of haplotype data? A schematic illustrating these possibilities appears in Figure 5. These speculations are meant to demonstrate that full understanding and accurate prediction of pharmacological phenotypes encoded at the level of genomic DNA will require the creative application of resources and concepts from a variety of fields, just as did the original idea of applying haplotype data in addition to genotype data to predict phenotypic responses to drug therapy. Finally, most pharmacogenomic traits will be polygenic; therefore, the completion of the human HapMap project and the development of high-throughput haplotyping methods should allow us to use association studies to elucidate the genetic basis of many traits. Such approaches may well be required if we are to fully elucidate the genomic determinants of drug response and to maximize the power of pharmacogenomics for optimizing drug therapy.

REFERENCES

1. Evans WE, McLeod HL. Pharmacogenomics—drug disposition, drug targets, and side effects. N Engl J Med 2003; 348:538–549.
2. The International HapMap Consortium. The international HapMap project. Nature 2003; 426:789–796.
3. Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet 2003; 4:587–597.
4. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science 2002; 296:2225–2229.

5. Fan JB, Surti U, Taillon-Miller P, Hsie L, Kennedy GC, Hoffner L, Ryder T, Mutch DG, Kwok PY. Paternal origins of complete hydatidiform moles proven by whole genome single-nucleotide polymorphism haplotyping. Genomics 2002; 79:58–62.
6. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BTN, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 2001; 294:1719–1723.
7. Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci USA 2000; 97:10483–10488.
8. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridenos M, Livingstone S, Ganske R, Lohmussar E, Zermant J, Tonisson N, Remm M, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Durham I. A first-generation linkage disequilibrium map of human chromosome 22. Nature 2002; 418:544–548.
9. Hodge SE, Boehnke M, Spence MA. Loss of information due to ambiguous haplotyping of SNPs. Nat Genet 1999; 21:360–361.
10. Crouau-Roy B, Clayton J. Haplotypes without children: PCR applied to close loci on individual human sperm. Hum Biol 1995; 67:171–178.
11. Ruano G, Kidd KK, Stephens JC. Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. Proc Natl Acad Sci USA 1990; 87:6296–6300.
12. Mitsunaga S, Oguchi T, Moriyama S, Tokunaga K, Akaza T, Tadokoro K, Juji T. Multiplex ARMS-PCR-RFLP method for high-resolution typing of HLA-DRB1. Eur J Immunogenet 1995; 22:371–392.
13. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. Science 2004; 304:581–584.
14. Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. Blood 2002; 100:3077–3086.
15. Johnson M. The beta-adrenoceptor. Am J Respir Crit Care Med 1998; 158:S146–S153.
16. Drazen JM, Israel E, Boushey HA, Chinchilli VM, Fahy JV, Fish JE, Lazarus SC, Lemanske RF, Martin RJ, Peters SP, Sorkness C, Szefler SJ. Comparison of regularly scheduled with as-needed use of albuterol in mild asthma. Asthma Clinical Research Network. N Engl J Med 1996; 335:841–847.
17. McDonald OG, Krynetski E, Evans WE. Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation. Pharmacogenetics 2002; 12:93–99.
18. Tost J, Brandt O, Boussicault F, Derbala D, Caloustian C, Lechner D, Gut IG. Molecular haplotyping at high throughput. Nucleic Acids Res 2002; 30:e96.

19. Nuovo GJ, Gallery F, MacConnell P, Becker J, Bloch W. An improved technique for the in situ detection of DNA after polymerase chain reaction amplification. Am J Pathol 1991; 139:1239–1244.
20. Krynetski E, Evans WE. Drug methylation in cancer therapy: lessons from the TPMT polymorphism. Oncogene 2003; 22:7403–7413.
21. Otterness D, Szumanski C, Lennard L, Klemetsdal B, Aarbakke J, Park-Hah JO, Iven H, Schmiegelow K, Branum E, O'Brien J, Weinshilboum R. Human thiopurine methyltransferase pharmacogenetics: gene sequence polymorphisms. Clin Pharmacol Ther 1997; 62:60–73.
22. Yates CR, Krynetski EY, Loennechen T, Fessing MY, Tai HL, Pui CH, Relling MV, Evans WE. Molecular diagnosis of thiopurine S-methyltransferase deficiency: genetic basis for azathioprine and mercaptopurine intolerance. Ann Intern Med 1997; 126:608–614.
23. McLeod HL, Krynetski EY, Relling MV, Evans WE. Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. Leukemia 2000; 14:567–572.
24. Tai HL, Krynetski EY, Schuetz EG, Yanishevski Y, Evans WE. Enhanced proteolysis of thiopurine S-methyltransferase (TPMT) encoded by mutant alleles in humans (TPMT*3A, TPMT*2): mechanisms for the genetic polymorphism of TPMT activity. Proc Natl Acad Sci USA 1997; 94:6444–6449.
25. Tai HL, Fessing MY, Bonten EJ, Yanishevski Y, d'Azzo A, Krynetski EY, Evans WE. Enhanced proteasomal degradation of mutant human thiopurine S-methyltransferase (TPMT) in mammalian cells: mechanism for TPMT protein deficiency inherited by TPMT*2, TPMT*3A, TPMT*3B or TPMT*3C. Pharmacogenetics 1999; 9:641–650.
26. Evans WE, Relling MV. Moving toward individualized medicine with pharmacogenomics. Nature 2004; 429:464–468.
27. Schutz E, von Ahsen N, Oellerich M. Genotyping of eight thiopurine methyltransferase mutations: three-color multiplexing, "two-color/shared" anchor, and fluorescence-quenching hybridization probe assays based on thermodynamic nearest-neighbor probe design. Clin Chem 2000; 46:1728–1737.
28. Fournier C, Goto Y, Ballestar E, Delaval K, Hever AM, Esteller M, Feil R. Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. EMBO J 2002; 21:6560–6570.
29. Jenuwein T, Allis CD. Translating the histone code. Science 2001; 293: 1074–1080.

20

Pharmacoepigenetics: From Basic Epigenetics to Therapeutic Applications

James Flanagan and Arturas Petronis

The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada

Epigenetics refers to regulation of various genomic functions controlled by stable but potentially reversible changes in DNA methylation and chromatin structure (1). Traditionally, phenotypic outcomes have been "assigned" to the DNA sequence variation, while the epigenetic component of the chromosome has been ignored for quite a long time. The last 10–15 years, however, observed numerous discoveries in terms of mechanisms and the role of epigenetic machinery in the normal and abnormal functioning of a cell, as well as the development of a wide variety of experimental technologies that now place epigenetics among the frontiers of modern molecular biology (2). One of the important observations is the increasing evidence that epigenetic factors play an important role in the etiopathogenesis of human diseases, and the proportion of epigenetic diseases could be substantial. Uncovering the epigenetic risk factors opens new opportunities for diagnostic, prognostic, and therapeutic approaches in human morbid biology. All the epigenetics-related pharmacological developments can be united under the "umbrella" of *pharmacoepigenetics*. Pharmacoepigenetics is still a new and poorly described concept, and this chapter is one of the first efforts to compile the relevant knowledge. In addition to the pharmacoepigenetic issues per se, we also briefly describe the main mechanisms of epigenetic regulation and the role, both proven and putative, in rare and common human diseases.

1. GENERAL PRINCIPLES OF EPIGENETIC REGULATION

1.1. Epigenetic Mechanisms

There are two major mechanisms of epigenetic regulation, methylation of cytosines in the DNA sequence and modification of the histone proteins that the DNA is wrapped around. The co-ordination of both mechanisms results in dramatic changes in remodeling of chromatin and altered gene transcription.

1.1.1. DNA Methylation

The DNA can be methylated at the fifth position of cytosine predominantly in the Cytosine-guanine dinucleotide (CpG) dinucleotide. The CpG dinucleotides are generally under-represented in the genome. The regions with higher proportions (than would be expected by chance) of CG dinucleotides are referred to as CpG islands, which constitute 1–2% of the total genome and account for 15% of all CpG dinucleotides. The depletion of CpG dinucleotides from the genome and the resulting under-representation is believed to have arisen from spontaneous deamination of 5-methylcytosine (^{met}C) to thymine (3). This type of mutation gives rise to a relatively common C→T transition at CpG sites, which contribute up to 30% of all germline and somatic point mutations (4,5). The CpG islands are associated with 60% of human genes (6), however, are predominantly unmethylated in the human genome. The remaining 85% of CpG dinucleotides occur outside of CpG islands and are largely methylated, occurring mostly in transposable elements such as retrotransposons, *L1* elements, *Alu* sequences, endogenous retroviruses, and juxtacentromeric satellite sequences.

The establishment and maintenance of DNA methylation patterns in mammalian cells is performed by three DNA methyltransferases (DNMTs): DNMT1, DNMT3a, and DNMT3b. Initial identification and characterization of the DNMTs has been perfomed in mice (7). Dnmt1 is a large protein of 1620 amino acids with the N-terminal portion that contains domains involved in nuclear localization (8), DNA replication (9), DNA binding (10), protein–protein interactions (11), and in particular interacts with proliferating cell nuclear antigen (PCNA) (12), the transcriptional repressor DNA methyltransferase 1-associated protein 1 (DMAP1) (13), histone deacetylates 1 and 2 (13,14), the transcription factor E2F1 and the tumor suppressor gene Rb (15). The C-terminal portion of the protein, like the other DNMT proteins, contains the catalytic methyltransferase motifs. Dnmt1 is ubiquitously expressed, it localizes to the replication foci and exhibits a 10–40-fold increased preference for hemimethylated DNA, which strongly argues for its role in copying and spreading the methylation patterns after DNA replication (9,16,17). The deletion of *Dnmt1* in mice results in global demethylation and embryonic lethality highlighting the importance of this protein and DNA methylation in development (18,19).

The two other DNMTs, Dnmt3a and Dnmt3b, contain a cysteine rich zinc-like DNA binding domain and a Pro-Try-Try-Pro domain important for proteins involved in cell growth and differentiation (20). This DNMT3 family of enzyme methylate equally both hemi- and fully unmethylated DNA, which has led to their description as "de novo methyltransferases" (21). These enzymes are highly conserved from human to mouse and zebrafish, and are all required for the wave of de novo methylation that occurs in the genome following embryonic implantation (15,22). The DNA methyltransferase 2 (*Dnmt2*) was also identified, however, despite the sequence similarities with the other methyltransferases, shows no detectable methyltransfease activity and displays no phenotype when deleted in mice (21,23).

The DNMTs at least partially account for the interaction of DNA methylation with histone modification as these enzymes are able to directly recruit histone deacetylases (HDACs) and this is achieved by binding methylated C (^{met}C) binding proteins such as MeCP2, MBD1, MBD2, and MBD3 (detailed reviews in Refs. 24 and 25).

1.1.2. Methylated Cytosine Binding Proteins (MeCPs)

There are six mammalian ^{met}C binding proteins, including MeCP2, MBD1, MBD2, MBD3, MBD4, and the p120 catenin partner protein *Kaiso*. The MeCP2 was the first ^{met}C binding protein identified by its ability to bind to methylated DNA in vitro (26). It contains a ^{met}CpG binding domain (MBD) within the N-terminus, which recognizes a symmetrically methylated ^{met}C in the CpG dinucleotide through contacts in the major groove of the DNA double helix (27) and a transcriptional repression domain (TRD) which interacts with other regulatory proteins to effect transcriptional repression (28). The MeCP2 protein is concentrated in the juxtacentromeric heterochromatin, which contains a large proportion of all genomic methylcytosine (29). The MeCP2 associates with the Sin 3a HDAC complex in mice, linking the DNA methylation and histone deacetylation transcriptional repression pathways, however, transcriptional repression by MeCP2 may also occur independently of histone deacetylation via direct interactions with transcription factors such as TFIIb (30–32). The *MBD1-4* genes were identified based on their sequence homology to MeCP2 in the methyl-binding domain and apart from MBD3 can preferentially bind ^{met}CpG (33). Recent findings have also linked the methyl-binding proteins MBD1, MBD2, and MBD3 with the chomatin remodeling machinery (34,35). The MBD3 is a part of the Mi-2/NuRD ATP-dependent chromatin-remodeling complex, while MBD2 is involved in the MeCP1 HDAC complex that also contains HDAC1, HDAC2 and the retinoblastoma protein-binding protein RbAp46/48 and can repress methylated promoters and remodel methylated chromatin with high efficiency (35,36). The MBD1, while not identified as part of a repression complex, can act as a methylation-mediated transcriptional repressor (37,38). The more recently identified MBD4 is a thymine

glycosylase that can bind to the deaminated ^{met}CpG sites, is involved in DNA mismatch repair and is mutated in many cancers displaying microsatellite instability (39,40). Finally the unrelated protein Kaiso, which does not contain an MBD domain, binds in vitro to methylated ^{met}CG ^{met}CG DNA motifs via its zinc-finger domain and can effect methylation-dependent transcriptional repression (41).

While loss of the *Dnmt1* gene in mice leads to catastrophic demethylation resulting in embryonic lethality, loss of one methyl-binding protein or even double mutant loss of MBD2 and MeCP2 does not cause major detrimental developmental changes, which suggests some level of redundancy in the methyl-binding proteins. However, loss of MeCP2 in mice presents with symptoms similar to Rett syndrome in humans (see below) (42). Absence of the *MBD2* activity in mice leads to maternal behavioral abnormalities, however, loss of *MBD3* is embryonic lethal (43). Targeted deletion of the DNA repair gene *MBD4* in mice leads to increased CpG mutation frequency and increased tumorigenesis (44).

1.1.3. Histone Modifications

Histones are basic proteins that are made up by a globular domain and an N-terminal tail that protrudes from the nucleosome. Nucleosomes form the basic unit of chromatin and are made up by a complex of DNA wrapped around an octamer of histones formed by pairs of the histones H2A, H2B, H3, and H4 (45,46) (Fig. 1). Post-translational modification of the core histone tails by methylation, acetylation, phosphorylation, ubiquitination, or sumoylation can alter the structure of the nucleosomes and thus alter gene expression. These post-translational modifications determine the structure and pattern of chromatin condensation and determine the "histone code" that drives gene transcriptional regulation (47,48). Below are briefly described the factors determining the histone acetylation and methylation.

The HDACs and histone acetyltransferases (HATs) play a major role in balancing the acetylated and deacetylated states of the chromatin. Most HATs function as transcriptional co-activators, and interact with DNA via sequence specific DNA-binding proteins (49). The HATs function to acetylate the ε-NH_2 group on lysine residues within histone tails that results in transcriptional activation. The HATs also have factor-acetyltransferase activity, which acetylates cellular transcription factors to complement transcription-factor activity (50). The human HATs are grouped into subfamilies on the basis of their histone specificity and sequence similarity including the GNAT famify, MYST family, SCR family, CBP/p300 family, and the TAFII250 family (49,51).

The HDACs operate antagonistically to HATs by deacetylating lysine residues on histone tails that results in the chromatin condensation and transcriptional repression. There are three classes of HDACs: class I (HDACs

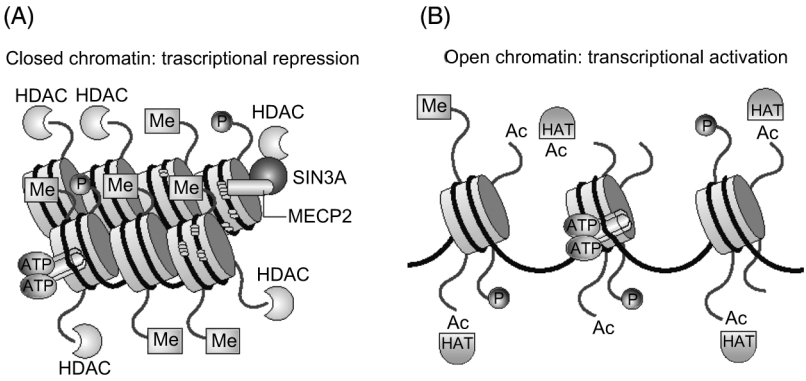


Figure 1 (**A**) Chromatin structure regulates transcriptional activity. Nucleosomes consist of DNA (*black line*) wrapped around histone octomers (*dark gray*). Post-translational modification of histone tails by methylation (Me), phosphorylation (P), or acetylation (Ac) can alter the higher-order nucleosome structure. Nucleosome structure can be regulated by ATP-dependent chromatin remodelers (*light gray cylinders*), and the opposing actions of HATs and HDACs. Methyl-binding proteins, such as the methyl-CpG-binding protein (MECP2), target methylated DNA (*light gray*) and recruit HDACs. (**A**) DNA methylation and histone deacetylation induce a closed-chromatin configuration and transcriptional repression. (**B**) Histone acetylation and demethylation of DNA relaxes chromatin, and allows transcriptional activation. *Source*: From Nature Reviews Drug Discovery (51) copyrights 2002 Macmillan Magazines Ltd.

1–3 and 8) that localize to the nucleus, class II (HDACs 4–7, 9, 10, and 11) that localize to both nucleus and cytoplasm, and class III, which are a structurally distinct class of NAD-dependant enzymes (sirtuins; SIRTs 1–7) (52). The number and diversity of HDACs and HATs in the human genome gives an indication of the large variety of cell processes they are involved in whether it is transcription, cell cycle progression, gene silencing, differentiation, hormone signaling, DNA replication, or DNA damage response (52).

The ATP-dependant remodeling complexes use the energy of ATP hydrolysis to introduce superhelical torsion into nucleosomal DNA, to form bulges or loops in the DNA (53). This allows the various chromatin proteins access to the DNA to control transcription, DNA replication, or recombination. Mammalian ATP-dependant remodeling complexes include SWI/SNF2/Brm, ISWI, and the Mi-2/NuRD complex (reviewed in Ref. 54). These complexes act by either "sliding" the DNA with respect to the nucleosome or by generating high-energy intermediate conformations that stochastically collapse to a state that allows transcriptional activation (54).

In addition to acetylation/deacetylation, there are numerous other histone modification scenarios. For example, histones can be methylated or unmethylated at lysine, serine, and arginine residues located in the

amino-terminal histone tails. Methylation of lysine 4 of histone 3 (H3-K4) is associated with active gene expression, while methylation of lysine 9 of histone 3 (H3-K9) is associated with silenced transcription (47,55). Six histone methyltransferases (HMTs) have been identified including the H3-K4 methyltransferase (56,57) and a number of H3-K9 methyltransferases, Suv39h1, and Suv39h2 (58), G9a (59), ESET/SetDB1 (60,61), and the Eu-HMTase1 (62).

1.1.4. Coordination of the Epigenetic Machinery

A number of models have been proposed to describe how DNA methylation and histone modification are both interdependant processes involved in chromatin remodeling and gene silencing (reviewed in Ref. 63). The first model revolves around DNA methylation directing chromatin remodeling where DNA methylation sites attract methyl-binding proteins, which then recruit their respective HDAC complex proteins to the methylated DNA region, which are then able to induce histone acetylation and gene silencing. The chromatin can then attract HMTs, which methylate the histones and stabilize the inactive chromatin state. The second proposed model describes the opposite scenario where histone methylation directs DNA methylation. The methylated histones recruit the heterochromatin protein 1 (HP1) which then recruits DNMTs to the silent chromatin in order to maintain the DNA methylation and stablize the inactive chromatin. The third model focusses on chromatin remodeling as the catalyst for DNA methylation. The ATP-dependant chromatin-remodeling complexes unwind the nucleosomal DNA to increase the accessiblity to the DNMTs, HDACs, and HMTs, which then modify the histones and methylate the DNA to inactivate the chromatin and silence gene expression. Regardless of the mechanism, the interplay between these epigenetic regulators is complex, and like many intracellular processes, these epigenetic processes are quite dependent on each other.

1.2. Epigenetic Functions

Epigenetic factors contribute to a large number of genomic functions, from regulation of gene activity to genome stability and segregation of chromosomes. Epigenetic regulation of genes is the most relevant to this review, and below several well-known examples dealing with genomic imprinting, X chromosome inactivation, and suppression of parasitic DNA elements are provided.

1.2.1. Genomic Imprinting

Genomic imprinting is the mechanism by, which only one of the two parental gene copies—either maternal or paternal—is expressed. Modifications in DNA and histones are the marks of genomic imprinting. There are about 50 known imprinted genes in the human genome (64). For example, the

imprinted genes *H19* and *Rasgrf1* are densely methylated on the paternal allele, while other genes such as *Igfr2* and *Snprn* gain their methylation imprint on the maternal allele (63). While in somatic cells the imprinting depends on the parental origin of the alleles, germline cells have to reprogram imprinting patterns depending on the sex of the embryo. Such epigenetic reprogramming starts with the demethylation of imprinted loci in the primordial germ cells, which erases parental imprinting marks at around embryonic day 11.5, and eventually new imprinting patterns, are established (65). Under normal circumstances, the imprinting marks are stable during the development, and such marks remain intact during the active and passive demethylation of paternal and maternal genomes, respectively, during early embryogenesis (66,67).

1.2.2. X Inactivation

X inactivation is a process that silences one of the two X chromosomes during early female embryogenesis, by converting an X chromosome from active euchromatin into transcriptionally silent condensed heterochromatin. The epigenetic mechanisms involved in X inactivation share many similarities with genomic imprinting. This process occurs shortly after implantation and is accompanied by a marked increase in the expression and accumulation of *Xist* RNA on the inactive X chromosome (68). The expression of *Xist* is negatively regulated on the active X chromosome by the antisense non-coding transcript *Tsix* initiated downstream of *Xist* (69). The mechanism by which *Xist* RNA coating the chromosome causes the inactivation of the chromosome is largely unknown, however recent studies have suggested that the *Xist* RNA recruits both H3-K9 methyltransferases and HDACs to the inactive chromosome and then spreads the inactivation along the chromosome by the formation of heterochromatin (70,71). The maintenance of the inactive X requires the synergistic action of *Xist* RNA, DNA methylation and histone deacetylation to maintain the hypoacetylated inactive X chromosome and hyperacetylated active X chromosome (72).

1.2.3. Genome Defense Against Parasitic Sequences

The majority of the CpG dinucleotides in the human genome is hypermethylated and resides, not in CpG islands of gene promoters, but in "parasitic" DNA retroelements. These elements (e.g., *L1* or *Alu* elements) account for almost 40% of the human genome, and it has been generally accepted that DNA methylation of these elements prevents gene mutagenesis due to unconstrained transposition by limiting their spread through the genome and prevents the transcription of such retroelements which would not be commensurate with normal gene expression programs (73,74). It has also been suggested that the epigenetic silencing of repetitive DNA is also a defensive mechanism to ensure genomic stability during meiosis to protect the genome from recombination between non-allelic repeats, which can

cause chromosome rearrangements and translocations (75,76). It is not clear, however, how epigenetic mechanisms identify retroelement DNA.

All the above cases of epigenetic regulation are classical examples of the epigenetic control on gene activity that completely inactivate a gene or a group of genes but spare their homologues. Only a small proportion of human genes are controlled through the "black or white" option, while epigenetic regulation of the overwhelming majority of genes is more similar to the "many shades of gray" scenario. Selection for the analysis of genes with major, rather than subtle, epigenetic a difference is understandable as the former are easier to detect in comparison to the latter. To a large extent such biases have been determined by the limitations of the available technologies for epigenetic studies. Until recently the methods used to detect methylation have been based on the quantification of global methylation levels in genomic DNA and more sequence specific methods utilizing methylation sensitive restriction enzymes. The recent renewed interest in the DNA methylation field has lead to the development of more sensitive means of detection of DNA methylation on a small sequence specific scale using the bisulfite modification approach and the large-scale global approaches using microarray technology (reviewed in Refs. 2 and 77).

Unlike DNA sequence, epigenetic regulation is a dynamic process, and epigenetic status in a somatic cell is subject to change under the influence of pre- and post-natal developmental programs as well as intra- and extracellular environmental effects. In addition, quite significant epigenetic changes can occur even in the absence of obvious environmental factors, i.e., due to stochastic reasons (78,79). After mitotic cell division, the daughter chromosomes carrying identical DNA sequences do not necessarily carry identical epigenetic patterns in comparison to the parental chromosomes. Substantial epigenetic differences may be accumulated across the cells of the same cell line or the same tissue (ibid.). The partial stability of epigenetic marks is also demonstrated during the maturation of the germline. It has been generally assumed that during gametogenesis epigenetic signals are erased and a new epigenetic profile is established (63), however, there is an increasing body of experimental evidence suggesting that not all epigenetic factors are erased in the sperm cells and oocytes and the varied epigenetic features can be transmitted from one generation to another (80). This observation blurs the demarcation between epigenetic- and DNA sequence-determined traits and also opens new avenues in the research of germline epimutations (81,82).

2. EPIGENETICS AND HUMAN DISEASE

The epigenetics research tradition in human diseases has been rather weak thus far. Epigenetic factors have been investigated only when their role was evident from epidemological, cytogenetic, or other type of studies. The best example is the rare pediatric diseases that are often caused by

imprinting defects, and such include Beckwith-Weidemann and Prader-Willi/Angelman syndrome (15). Another category of epigenetic diseases includes the diseases that are primarily caused by DNA sequence mutations in the genes encoding protein factors that play a critical role in establishing and/or maintaining of epigenetic patterns. Among these are ICF (for immunodeficiency, centromere instability, and facial anomalies) syndrome, a rare autosomal recessive disease that presents with immunodeficiency as a result of low or absent expression of various immunoglobulins and centromere instability. Clinically these patients show delayed developmental milestones, mental retardation, and facial abnormalities. The ICF syndrome is caused by mutations in one of DNMTs, namely *DNMT3b* (22,83,84) that results in the overall decrease in global methylation as well as hypomethylation of many repetitive sequences in the juxta-centromeric regions (85). Another example is Rett syndrome, a severe neurological autosomal dominant disorder that occurs nearly exclusively in females and is caused by mutations in the gene encoding MeCP2 (28,86). The fragile X syndrome presents with mental retardation caused by a combination of DNA sequence mutation in the *FMR1* gene plus the hypermethylation of the CpG island as well as histone deacetylation that all results in transcriptional silencing of *FMR1* (87). This protein is abundant in normal neurons and it has been suggested that the lack of FMR1 in neurons during development gives rise to the neurological defect (15).

In addition to these rare syndromes epigenetic factors have been investigated in cancer (88). In nearly all cancers there is some degree of epigenetic misregulation, which is presented by the main observations of global genome-wide hypomethylation and local hypermethylation of promoter CpG Island that results in transcriptional silencing of tumor suppressor genes. A large spectrum of genes are aberrantly methylated in various tumors including signal transduction genes (RASSF1, LKB1, and APC), DNA repair genes (MGMT, hMLH1, and BRCA1), cell cycle regulators (Rb, $p16^{INK4a}$, $p15^{INK4b}$, and $p14^{ARF}$), apoptosis-related proteins (DAPK and CASP8), genes involved in carcinogen metabolism (GSTP1), hormone response genes (ER, PGR, AR, and RARβ), genes involved in angiogenesis (THBS1) and genes suspected to be involved in metastasis (TIMP3 and CDH1) (reviewed in Ref. 89). Since most cancer epigenetic studies have been performed on malignant tissues, the questions of cause-effect (or primary–secondary pathological events) have been legitimately raised (90). It is likely that some of the detected epigenetic changes represent non-specific events resulting from a wide variety of earlier genetic and epigenetic changes. However, there are several studies indicating that epimutations can precede the event of carcinogenesis and therefore could be of diagnostic, prognostic, and therapeutic value. The first example is the so-called loss of imprinting (LOI), which indicates failure of the parent of origin specific monoallelic expression, which was detected for *IGF2* in the normal colonic mucosa in

up to 30% of colorectal cancer cases compared to only 10% in the general population (82). This phenomenon can also be detected in the peripheral blood lymphocytes, which argues that the *IGF2* LOI is inherited or acquired before the major differentiation of tissues. Germline epimutations have now also been detected in two colorectal cancer patients at the DNA damage repair gene *MLH1* (81). These patients did not have any detectable genetic mutations in this gene, but did show extensive methylation of one allele of the *MLH1* gene detectable in the germline DNA.

Apart from DNA methylation studies, the second aspect of epigenetic misregulation involves disruption of histone acetylation and chromatin structure. Functional disruption of some HATs genes (such as the CBP and p300) is detected in some cancers, most notably, acute pro-myelocytic leukemia (51). This has lead to the hope that HDAC inhibitors will be useful for the treatment of this disease and a number of inhibitors are currently being assessed in clinical trials (see below).

In addition to carcinogenesis, there is a very good reason to believe that epigenetic factors play a critical role in a wide variety of common non-Mendelian disorders such as schizophrenia, asthma, diabetes, multiple sclerosis, psoriasis, and rheumatoid arthritis, among numerous others. The epigenetic theory of complex disease is based on two important features of epigenetic inheritance: (i) is an absolutely necessary requirement for a normal functioning of a cell, and any epigenetic dysfunction may have serious consequences and (ii) epigenetic inheritance exhibits only partial stability, and therefore represents the dynamic component of a chromosome (while DNA sequences are static) (91). The partial stability of epigenetic regulation provides a new opportunity for the explanation of a number of non-Mendelian features in complex diseases that traditional DNA sequence-based genetics could not. Monozygotic twins, although carrying identical (or nearly identical) DNA sequences, may exhibit numerous epigenetic differences caused by environmental and stochastic influences (92). An epigenetic defect, or epimutation (93), may progress after separation of twins and reach the ''threshold'' of clinical symptoms in only one twin, while the co-twin may remain below such a ''threshold'' without any symptoms at all. Such a twin pair would be treated as discordant, with incomplete penetrance of the disease gene. In a similar way the epigenetic dynamics is consistent with other non-Mendelian features of complex diseases; relative age of onset and sometimes coincidence of the peaks of incidence with major hormonal changes in the organism (e.g., major psychosis and multiple sclerosis), sex effects (e.g., lupus and autism), parental origin effects, major fluctuations in the clinical course—from severe invalidating relapses to relatively good remissions, sometimes recovery (attention deficit, hyperactivity disorder, and asthma). In conclusion, it can be stated that the dynamic epigenetic inheritance system is more consistent with various non-Mendelian features of complex diseases than the DNA sequence-based factors. Epigenetic factors can be a common ''denominator'' for the wide variety of key

elements of complex traits, and the aberrations in gene expression (in terms of its degree, time, and location) may be an important etiological factor in complex diseases. The expectation that epigenetic changes, or epimutations, are common in human diseases provides the basis for pharmacoepigenetic approaches. In addition, epigenetic agents have already been proven to be of therapeutic value in clinical applications despite full understanding of their mechanisms of action.

3. PHARMACOEPIGENETICS: THE FIRST STEPS

3.1. Prediction of Drug Response

Pharmacoepigenetics is a new concept in biomedical research and the territory that this new development will cover is still to be mapped. In parallel to pharmacogenetics, one of the main goals of pharmacoepigenetics will be to predict drug response and adverse events based on the epigenetic individuality of an organism. Epigenetic variation across individuals is much richer in comparison to DNA sequence variation, and identical DNA sequences in unrelated individuals (94) and also monozygotic twins (92) exhibit significant epigenetic variation. Such epigenetic polymorphism may have an impact on gene expression that translates into differential density of receptors or varied numbers of molecules of an enzyme—the factors that contribute to the pharmacodynamic and pharmacokinetic features of an organism. If the DNA sequence variation, especially the one in the coding sequence, determines qualitative aspects of the protein (enzymes of high or low activity, e.g., CYP enzymes Ref. 95; receptors with high or low affinity, e.g., adrenergic receptors, Ref. 96), epigenetic factors account for the "quantitative" side of the biochemical individuality, and the overall result depends on the interaction of the DNA sequence-based and epigenetic factors. One can imagine some scenarios when DNA sequence-based pharmacogenetic idiosyncrasies are overridden by epigenetic factors: a high activity genetic allele is epigenetically suppressed that results in the low overall activity of a metabolizing enzyme, and vice versa, an over-expressed allele encoding low affinity receptor may functionally be equal to that of high affinity allele. It is the expectation that the prediction of treatment could be much more precise if epigenetic factors are taken into account and investigated together with the DNA sequence variation.

Prediction of drug response is intimately related to the issue of molecular heterogeneity of clinically indistinguishable disease phenotypes. Different molecular subtypes of a disease despite their similarities in the clinical presentation may require very different therapeutic strategies. The classical example is the syndrome of anemia that in the beginning of the 20th century was treated as a single nosological entity. Our current understanding of the very different molecular etiologies of anemias allowed

for specific and efficient therapeutic intervention in the case of at least some syndromes. The first steps in epigenetic description and characterization of different disease molecular phenotypes have already been made, although all of the studies thus far have concentrated on various cancers, endometrial (97), gastric (98), breast (99) lung (100), esophagus (101), colorectal (102), and Wilms tumors (103) and the key objective of such studies was to prognosticate the outcome of the disease. It is only recently that the concept of using aberrant CpG island methylation has been explored in an investigation to predict the responsiveness of ovarian cancer cell lines to cisplatin (104).

3.2. Epigenetic Treatment

In addition to the prediction of therapeutic response, the elucidation of epigenetically misregulated genes opens up new ways for new therapeutic strategies. Epigenetics provides an exciting new avenue to reprogram misregulated gene expression and has been the subject of many recent reviews (51,52,89,105–107). Ironically, some medications have been introduced in the clinical practice decades before their epigenetic mechanism of action has been identified. Valproate, a medication used for treatment of epilepsy, mania, and schizophrenia, was recently shown to exhibit HDAC inhibitory effects (108). Application of epigenetic agents in treatment, especially in cancer, is becoming very popular, and such developments are described below.

3.3. The HDAC Inhibitors

The HDAC inhibitors cause the accumulation of acetylated histones in nucleosomes, which results in the expression of genes that can lead to cell cycle arrest, differentiation, or apoptosis and have therefore been propelled forward to clinical trials as candidate therapeutic agents for the treatment of leukemia and other cancers (reviewed in Refs. 51 and 52). DNA methylation inhibiting agents are also currently being investigated (107) as are combinations of both HDAC and DNMT inhibitors. Aberrant transcription of genes genome-wide as a result of altered expression or abrogation of genes that encode HDACs, HATs, and other histone regulatory proteins is a critical step in the onset and progression of cancer and other disorders.

Inhibition of HDACs is one key mechanism to reactivate the expression of these misregulated genes. The astounding tumor specificity of many HDAC inhibitors relays the potential for many of these new compounds for the treatment of cancer and perhaps other disorders. There are five classes of HDAC inhibitors (reviewed in Refs. 51 and 52) including: (i) short-chain fatty acids such as sodium-*n*-butyrate; (ii) hydroxyamic acids, such as trichostatin A (TSA), suberoylanilide hydroxamic acid (SAHA), *m*-carboxycinnamic acid bishydroxamic acid (CBHA), azelaic bishydroxamic acid (ABHA), and

Oxamflatin; (iii) cyclic tetrapeptides containing a 2-amino-8-oxo-9,10-epoxy-decanoyl (AOE) moiety such as trapoxin; (iv) cyclic tetrapeptides without an AOE moiety such as apicidin and FR901228 (depsipeptide); (v) benzamides (Fig. 2 and Table 1).

There are two basic mechanisms that describe how HDAC inhibitors may function in cancer therapy, by either inhibition of cell proliferation or induction of apoptosis (105). The inhibition of cell proliferation and differentiation occurs by supporting nuclear receptor response driving terminal cell differentiation, reversal of repression by fusion transcription factors or over-expressed repressors, induction of p21, G1 arrest, and cellular differentiation, reactivation of silenced tumor suppressor genes, and suppression

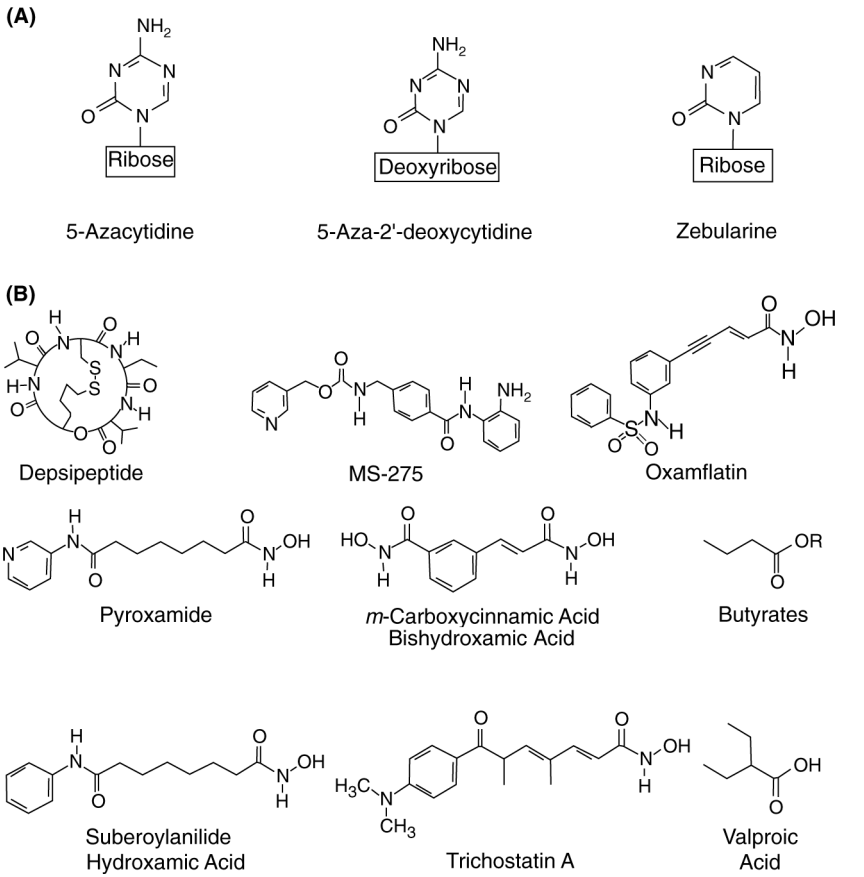


Figure 2 (**A**) DNA methyltransferase inhibitors. (**B**) Representative histone deacetylase inhibitors. *Source*: From Journal of the National Cancer Institute (89) copyright Oxford University Press, 2003.

Table 1 The In Vitro Effects of HDAC and DNMT Inhibitors

Class of inhibitor	Name	In vitro effect
HDAC inhibitors		
Short chain fatty acid	Butyrates	Apoptosis, differentiation, and cell cycle arrest
Short chain fatty acid	Valproic acid	Apoptosis and differentiation
Hydroxyamic acids	Suberoylanilide hydroxamic acid (SAHA)	Apoptosis, differentiation, and cell cycle arrest
Hydroxyamic acids	*m*-Carboxy cinnamic acid bishydroxamic acid (CBHA)	Apoptosis, differentiation, and cell cycle arrest
Hydroxyamic acids	Azelaic bishydroxamic acid (ABHA)	Apoptosis and cell cycle arrest
Hydroxyamic acids	Oxamflatin	Apoptosis and cell cycle arrest
Hydroxyamic acids	Trichostatin A	Apoptosis, differentiation, and cell cycle arrest
Hydroxyamic acids/ cyclic tetrapeptide	Cyclic hydroxamic acid-containing peptide (CHAP)	Cell cycle arrest
Cyclic tetrapeptide with AOE	Trapoxin A	Irreversible cell cycle arrest
Cyclic tetrapeptide without AOE	Apicidin	Apoptosis and cell cycle arrest
Cyclic tetrapeptide without AOE	FR901228 (depsipeptide)	Apoptosis and cell cycle arrest
Epoxide	Depudecin	Differentiaion and cell cycle arrest
Benzamides	MS-27–275	Cell cycle arrest
DNMT inhibitors		
Nucleoside analog inhibitor	5-Azacytidine	Differentiation and cell cycle arrest
Nucleoside analog inhibitor	5-Aza-2′Deoxycytidine (DAC, decitabine)	Differentiation and cell cycle arrest
DNMT1 antisense oligonucleotide	MG98	Differentiation and cell cycle arrest

of telomerase gene expression. The second mechanism is by induction of apoptosis by either activation of calpain/mitochondria-dependent apoptosis, activation and sensitization of death-receptor mediated killing, mitotic dysfunction, aberrant chromosomal segregation, and DNA damage or induction of topisomerase II, altering sensitivity to DNA-damaging agents. Other mechanisms include alteration of angiogenic signaling, alteration of microtubule

function, induction of MHC antigens on the cell surface to augment immune responses or suppression of IL-2-mediated gene expression (51).

There are currently over 40 clinical trials of HDAC inhibitors currently underway in the United States and Canada (Table 2). The majority of these trials are focussed on the safe use and efficacy of phenylbutyrate or the cyclic tetrapeptide FR901228 (depsipeptide) in the treatment of various cancers. Early results have suggested that depsipeptide can be used with minimal side effects in the treatment of lymphoma and other cancers with some patients showing limited responses to the drug (109–111). Phenylbutyrate is also well tolerated and can be used at the concentration that has been shown to have biological activity in vitro, though while no real individual responses were observed to this drug alone it may be quite useful in combination with other drugs (112). The other main focus of the many clinical trials is the use of valproic acid in the treatment of mental disorders such as bipolar disorder I and epilepsy.

3.4. The DNMTs Inhibitors

In many human cancers aberrant DNA methylation is a key feature (113,114), including the increased methylation and inactivation of many tumor suppressor genes (e.g., p.16, Ref. 115). Overall DNA methylation in cancers has frequently been found to be decreased or hypomethylated, while individual CpG islands associated with specific genes are found to be hypermethylated (116). This hypomethylation in cancers can also lead to the reactivation of genes that are normally silenced by DNA methylation (e.g., the cancer/testis antigen genes CTA, Ref. 117). Many genes that are hypermethylated in cancers can be reactivated with DNMT inhibitors (118).

There are a number of different groups of DNMT inhibitors. The first group is the nucleoside analog inhibitors of DNMT1. These include 5-azacytidine and 5-aza-2′-deoxycytidine (DAC). These drugs act as inhibitors of DNMTs only after they are incorporated into the DNA. The DAC, which appears to be more effective than 5-azacytidine, binds DNMT1 and traps it in the replication fork, which then proceeds to replicate DNA without the DNMT thereby passively demethylating the DNA (119). The major concern with the use of DAC is its potential to trap other proteins in the replication fork resulting in non-specific effects unrelated to the effects of DNMT1. These non-specific effects are also a concern with the use of 5-azacytidine as it also has non-specific genetic effects including mitotic recombination and point mutations (120). The main problem with these inhibitors is the wide-ranging toxicity seen in many clinical trials in the 1970s (121–124). The DAC also acts only on replicating cells, which is a limitation for their use in neurological disorders where epigenetic misregulation in non-dividing neurons cannot be targetted by these drugs. A recent addition to this group

Table 2 Current Clinical Trials of HDAC and DNMT Inhibitors

	Therapeutic agents	Phase	Target disease	Sponsor of trial/ institution
1	Arginine butyrate + gancyclovir	I	EBV-induced malignancies/ lymphoprolinferative disorders	NCI/BUSM
2	Phenylbutyrate + retinoic acid	I	AML, CML, myelopdysplastic syndrome	NCI/JHOC
3	Phenylbutyrate + azacytidine	I	Refractory advanced solid tumors	NCI/JHOC
4	Phenylbutyrate + azacytidine	I	Recurrent or refractory untreated AML, or MDS	NCI/JHOC
5	Phenylbutyrate (continuous infusion)	I	Recurrent prostate cancer and other cancers	NCI
6	Phenylbutyrate	I	Pediatric refractory malignancies	NCI/TCCC
7	Phenylbutyrate + genistein	I	Cystic fibrosis	NCRR
8	Phenylbutyrate + fluorouracil + indomethacin	I/II	Advanced colorectal adenocarcinoma	NCI/MTS
9	Phenylbutyrate	II	Pediatric progressive or recurrent CNS malignancy	NCI/TCCC
9	Depsipeptide	I	CLL, SLL, AML, ALL	NCI/OSU
10	Depsipeptide	I	Radioiodine-refractory non-medullary thyroid cancer	NCI
11	Depsipeptide	I	Pediatric patients with refractory or recurrent solid tumors or leukemia	NCI/COG
12	Depsipeptide	I	Unresectable pulmonary, eosophageal or pleural malignancies	NCI
13	Depsipeptide + decitabine	I	SCLC, NSCLC, mesothelioma	NCI
14	Depsipeptide +	I/II	Relapsed or refractory	NCI/MSGCC

(*Continued*)

Table 2 Current Clinical Trials of HDAC and DNMT Inhibitors (*Continued*)

	Therapeutic agents	Phase	Target disease	Sponsor of trial/ institution
	rituximab + fludarabine		low-grade B-Cell non-Hodgkin's lymphoma	
15	Depsipeptide	II	Cutaneous T-cell lymphoma, relapsed peripheral T-cell lymphoma	NCI
16	Depsipeptide	II	MDS, AML, intermediate-grade or follicular non-Hodgkin's lymphoma	NCI/MSKCC
17	Depsipeptide	II	Metastatic RCC progressive after immunotherapy	NCI and Industry/ JCCC
18	Depsipeptide	II	Relapsed or refractory AML	NCI/UCCRC
19	Depsipeptide	II	Relapsed or refractory multiple myeloma	NCI/ NYWCCC
20	Depsipeptide	II	Refractory mantle cell or diffuse large cell non-Hodgkin's lymphoma	NCI/MAYO
21	Depsipeptide	II	Advanced colorectal adenocarcinoma	NCI/SWOG
22	Valproate + DAC	I	Relapsed/refractory AML or previously treated CLL or SLL	NCI/OSU
23	Valproate + DAC	I/II	Relapsed/refractory Leukemia or MDS	SuperGen/ MDACC
24	Valproate	I/II	Kaposi Sarcoma	NCI/AMC
25	Valproate (+olanzapine or risperidone)	II	Schizophrenia	Abbott Laboratories
26	Valproate + fluoxetine	II	Intermittent explosive disorder + D62	NIMH
27	Valproate	III	Partial seizures in children	Abbott Laboratories
28	Valproate	III	Mania/bipolar disorder in children and adolescents	Abbott Laboratories
29	Valproate	III	Mania/bipolar disorder in adults	Abbott Laboratories

(*Continued*)

Table 2 Current Clinical Trials of HDAC and DNMT Inhibitors (*Continued*)

	Therapeutic agents	Phase	Target disease	Sponsor of trial/ institution
30	Valproate	III	Dementia	NIA/ADRC
31	Valproate	III	Aggressive autistic adolescents	NICHD/ NIMH
32	Valproate + quetiapine fumarate	III	Bipolar I disorder	AstraZeneca
33	Valproate	III	Early age mania/ bipolar disorder	NIMH
34	Valproate + lithium/+ lamotrigine	III	Bipolar disorder	NIMH
35	Valproate + olanzapine	IV	Bipolar disorder	Abbott Laboratories
36	Valproate	IV	Bipolar disorder	NIMH
37	SAHA	I	Advanced—leukemia, lymphoma, prostate cancer, multiple myeloma, eye cancer	NCI/MSKCC
38	SAHA	I	Advanced solid tumors or hematological malignancies	NCI/MSKCC
39	MS-27–275	I	Advanced solid tumors or lymphomas	NCI
40	MS-27–275	I	Poor risk hematological malignancies	NCI/GCC
41	Azacytidine	III	Myelodysplastic syndromes	Pharmion
42	Azacytidine	III	Myelodysplastic syndromes	SuperGen
43	DAC (decitabine)	I	Esophageal cancer, lung neoplasms, mesothelioma	NCI
44	DAC (decitabine)+ doxorubicin + cyclophosphamide	I	Pediatric relapsed/ refractory solid tumors or neuroblastoma	NCI
45	DAC (decitabine)	I	Advanced solid tumors, bladder cancer, breast cancer, male breast cancer, melanoma	NCI/USC
46	DAC (decitabine)	I	MDS or AML	NCI/PMH
47	DAC (decitabine)	I	Relapsed/Refractory	NCI/COG

(*Continued*)

Table 2 Current Clinical Trials of HDAC and DNMT Inhibitors (*Continued*)

	Therapeutic agents	Phase	Target disease	Sponsor of trial/ institution
			AML or ALL	
48	DAC (decitabine)	II	Chronic myelogenous leukemia-accelerated, blast and chronic phases	SuperGen
49	DAC (decitabine)	II	Myelodysplastic syndrome, chronic myelomonocytic leukemia	SuperGen
50	DAC (decitabine)	II	MDS, chromic myelomonocytic leukemia	SuperGen/ MDACC
51	DAC (decitabine)	II	Chronic myelogenous leukemia	NCI/MDACC
52	DAC (decitabine)	II	MDS	NCI/MSKCC
53	DAC (decitabine)	III	Advanced stage MDS	SuperGen
54	DAC (decitabine)	III	Older patients (>60 years) with MDS	EORTC LCG / GMSSG
55	MG-98	I	Advanced solid tumors	NCIC/ SKCCC/ ORCC

All information was obtained from the NCI clinical trials website http://www.cancer.gov/clinicaltrials/; and the National Institute of Health clinical trials website http://clinicaltrials.gov/.
Abbreviations: ADRC, Alzheimer's Disease Research Center; ALL, Acute Lymphocytic Leukemia; AMC, AIDS Associated Malignancies Clinical Trials Consortium; AML, Acute Myeloid Leukemia; BUSM, Boston University School of Medicine; CLL, Chronic Lymphocytic Leukemia; CML, Chronic Myeloid Leukemia; COG, Childrens Oncology Group; EBV, Epstein-Barr Virus; EORTC LCG, EORTC Leukemia Group; GCC, Greenebaum Cancer Center; GMSSG, German Myelodysplastic Syndrome Study Group; JCCC, Jonsson Comprehensive Cancer Center, UCLA; JHOC, John Hopkins Oncology Center; MAYO, Mayo Clinic Cancer Center; MDACC, M.D. Anderson Cancer Center; MDS, Myelodysplastic Syndrome; MSGCC, Marlene and Stewart Greenebaum Cancer Center; MSKCC, Memorial Sloan-Kettering Cancer Center; MTS, Mount Sinai Hospital; NCI, National Cancer Institute ; NCIC, National Cancer Institute of Canada; NCRR, National Center for Research Resources; NIA, National Institute on Ageing; NIMH, National Institute of Mental Health; NSCLC, Non-small cell lung cancer; NYWCCC, New York Weill Cornell Cancer Center; ORCC, Ottawa Regional Cancer Center; OSU, Ohio State University; PMH, Princess Margaret Hospital, Toronto; RCC, Renal Cell Carcinoma; SCLC, Small-cell Lung Cancer; SKCCC, Sidney Kimmel Comprehensive Cancer Center; SLL, Small Lymphocytic Leukemia; SWOG, Southwest Oncology Group; TCCC, Texas Children's Cancer Center; UCCRC, University of Chicago Cancer Research Center; USC, University of Southern California.
Source: Modified and updated from Ref. 51.

of nucleoside analog drugs is the zebularine [1-(β-D-ribofuranosyl)-1,2-dihybropyrimidin-2-one], which has been shown to inhibit DNA methylation in mice and result in a decrease in tumorigenesis with decreased toxicity compared to the other nucleoside drugs (125).

The next group of DNMT inhibitors, the non-nucleoside inhibitors of DNMTs, include the 4-amino-benzoic acid derivatives procaine and procainamide, which are antiarrythmic drugs that can induce global hypomethylation and demethylation of specific methylated CpG islands. Procainamide has been used clinically for many years, however, its usefulness as an anticancer therapeutic is yet to be investigated (126). Inhibition of DNMT activity can also be achieved with the use of antisense oligonucleotides to *DNMT1*, which have been shown to induce replication arrest with very limited global demethylation (127). The MBD2 is another potential target for anti-cancer therapeutic agents as targeted knockdown of this gene inhibits tumorigenesis in mice without altering cell growth or cell cycle kinetics, however currently there are no known inhibitors of this protein (126).

There are only a few effective agents that modulate DNA methylation that have been clinically evaluated: 5-azacytidine, 5-aza-2′-deoxycytidine (DAC), 1-β-D-arabinofuranosyl-5-azacytosine (fazarabine or Ara-C), and dihydro-5-azacytidine. While 5-azacytidine and DAC both showed some efficacy in treating acute myeloid leukemia (AML) the latter two-showed little efficacy and their trials were discontinued (128). The MG98 antisense DNMT1 oligonucleotide is also currently being assessed in clinical trials (129) (Table 2). Combinations of phenylburtyrates with azacytidine are also being trialed in patients with various solid tumors, leukemias, or myelodysplastic syndromes (Table 2).

3.5. The CpG Oligonucleotides and Immune Response

The innate or humoral immune system can be activated by exposure to pathogen associated molecular patterns (PAMPs) from various infectious microorganisms, and this activation is mediated by members of the Toll-like family of receptors (TLRs) (130,131). Synthetic oligodeoxynucleotides (ODNs) that contain CpG motifs, similarly to PAMPs, can stimulate an immune cascade through this interaction with TLRs (132). The CpG ODNs are rapidly endocytosed by immune cells and interact with the endocytic vesicle receptor TLR9 (133). This interaction leads to the swelling and acidification of the endocytic vesicle, and can initiate the immunostimulatory signaling cascade that results in upregulation of cytokines and chemokines that effect the T-helper 1 (Th1) related immune response (134,135). The CpG ODNs have a number of therapeutic applications including an immunoprotective effect, as a vaccine adjuvant and in prevention and treatment of allergies such as asthma. The ability of ODNs to stimulate a Th1-type immune response is not only important in eliciting the production of

IFN-gamma and IL-12 to reduce the inflammatory disease asthma, it also suggest that they may act as vaccine adjuvants (132). This adjuvant effect is mediated by promoting the immunogenicity of co-administered antigens and relies on the CpG-induced enhancement of antigen presenting cell function, induction of chemokine/cytokines that support antigen-specific immunity and increased antigen uptake by DNA binding receptors on antigen presenting cells. While the evidence does support the use of CpG ODNs as vaccine adjuvants no human clinical trials have yet significantly improved vaccine efficacy, although over a dozen phases I and II clinical trials are underway to evaluate the safety and efficacy of CpG ODNs in the treatment of cancer, allergies, asthma, or as vaccine adjuvants (132).

3.6. Designer Transcription Factors

Designer transcription factors may overcome a significant limitation of the rather non-specific approaches described above (e.g., HDAC inhibitors) and therefore to dramatically reduce the risk for adverse events (136). ''Designer transcription factors'' are chimeric proteins that contain a functional HAT or HDAC recruitment domain fused to hairpin polyamides or zinc-finger proteins that can localize the protein to specific DNA sequences. The practical engineering of these proteins involves the Cys2-His2 zinc finger (137,138). This particular DNA-binding domain is most useful for this purpose due to its great adaptability to diverse range of DNA sequence specificities. In addition, they can act as monomers and are therefore not restricted to palindromic sequences. In binding DNA these zinc finger proteins use multiple tandem zinc fingers to interact with the adjacent subsites in the major groove of the DNA. This provides a level of variability in the recognition of each subsequent finger to its cognate sequence and allows the designer to optimize base recognition one finger at a time (139). These compound proteins have been used to turn specific genes on or off in vitro (140–142). The full therapeutic usefulness of these molecules is yet to be established but these in vitro studies provide proof-of-principle that therapy with these designer proteins is possible. Pharmacological intervention via small molecule compounds against diseases such as cancer with such aberrant transcription is therapeutically effective in human patients (136,139).

While no epigenetic regulators have yet been exploited by this new technique a number of human endogenous genes have been regulated by the use of designed zinc-finger transcription factors. Some recent examples include the regulation of *erbB-2* and *erb-B3*, either activation or repression depending on the type of functional domain tethered to the zinc-finger used (143,144). Other genes that have been regulated by exogenous designer-transcription factors include the activation of erythropoietin (EPO) (145) and vascular endothelial growth factor A (VEGF-A) (141), activation and repression of the imprinted genes *IGF2* and *H19* (142) as well as the

epression of the *multidrug resistance gene 1* (*MDR1*) promoter by the use of a designer zinc-finger protein that prevents SP1 transcription factor binding to the promoter of the gene (146). These studies highlight the feasibility of regulation of endogenous mammalian genes with designer zinc-finger proteins and the future application of this technology to protein with epigenetic activity provides an interesting new avenue for therapeutic pharmacology.

4. CONCLUSIONS AND FUTURE CHALLENGES

In conclusion, the last decade has shown that epigenetic factors play a very important role in rare human diseases and there is a good reason to believe that such factors also contribute to the etiopathogenesis of common complex disorders. A number of questions, however, still need to be addressed, and such include: can epigenetic profiling be used to predict response to certain drugs in a clinical setting? Can gene-specific designer transcription factors be constructed once a mis-regulated gene is identified in a patient? What are the long-term effects of global HDAC inhibitors or DNMT inhibitors? There is little doubt that pharmacoepigenetics, while still a relatively new area, can provide the basis for exploitation of the epigenetic variabilities in human disease and may be the key to opening new opportunities for diagnostic, prognostic, and therapeutic approaches in human morbid biology.

ABBREVIATIONS LIST (EXCLUDING GENE SYMBOLS)

ABHA	Azelaic bishydroxamic acid
AOE	2-Amino-8-oxo-9,10-epoxy-decanoyl
CBHA	*m*-Carboxycinnamic acid bishydroxamic acid
CpG	Cytosine-guanine dinucleotide
DAC	5-aza-2′-deoxycytidine
DNMT	DNA methyltransferase
H3-K4	Lysine 4 of histone 3
H3-K9	Lysine 9 of histone 3
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
ICF	Immunodeficiency-centromere instability-Facial syndrome
LOI	Loss of imprinting
MBD	Methylated cytosine binding domain
MeCP	Methylated cytosine binding protein
^{met}C	Methylated cytosine
ODNs	Oligodeoxynucleotides
PAMPs	Pathogen associated molecular patterns
SAHA	Suberoylanilide hydroxamic acid
TSA	Trichostatin A

REFERENCES

1. Henikoff S, Matzke MA. Exploring and explaining epigenetic effects. Trends Genet 1997; 13:293–295.
2. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 2003; 3:253–266.
3. Jiricny J. Cancer surveys: genetic instability in cancer. Oxford: Oxford University Press, 1996:47–68.
4. Cooper DN, Youssoufian H. The CpG dinucleotide and human genetic disease. Hum Genet 1988; 78:151–155.
5. Greenblatt MS, Bennett WP, Hollstein M, Harris CC. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res 1994; 54:4855–4878.
6. Larsen F, Gundersen G, Lopez R, Prydz H. CpG islands as gene markers in the human genome. Genomics 1992; 13:1095–1107.
7. Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol 1988; 203:971–983.
8. Bestor TH, Verdine GL. DNA methyltransferases. Curr Opin Cell Biol 1994; 6:380–389.
9. Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 1992; 71:865–873.
10. Bestor TH. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. EMBO J 1992; 11:2611–2617.
11. Liu Y, Oakeley EJ, Sun L, Jost JP. Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. Nucleic Acids Res 1998; 26:1038–1045.
12. Chuang LS, Ng HH, Chia JN, Li BF. Characterisation of independent DNA and multiple Zn-binding domains at the N terminus of human DNA-(cytosine-5) methyltransferase: modulating the property of a DNA-binding domain by contiguous Zn-binding motifs. J Mol Biol 1996; 257:935–948.
13. Rountree MR, Bachman KE, Baylin SB. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet 2000; 25:269–277.
14. Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet 2000; 24:88–91.
15. Robertson KD, Wolffe AP. DNA methylation in health and disease. Nat Rev Genet 2000; 1:11–19.
16. Pradhan S, Bacolla A, Wells RD, Roberts RJ. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem 1999; 274:33002–33010.
17. Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, Jones PA. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res 1999; 27:2291–2298.

18. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 1992; 69:915–926.
19. Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 1996; 122:3195–3205.
20. Qiu C, Sawada K, Zhang X, Cheng X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol 2002; 9:217–224.
21. Okano M, Xie S, Li E. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res 1998; 26:2536–2540.
22. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999; 99:247–257.
23. Yoder JA, Bestor TH. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet 1998; 7:279–284.
24. Attwood JT, Yung RL, Richardson BC. DNA methylation and the regulation of gene transcription. Cell Mol Life Sci 2002; 59:241–257.
25. Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chem Biochem 2002; 3:274–293.
26. Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 1992; 69:905–914.
27. Wakefield RI, Smith BO, Nan X, Free A, Soteriou A, Uhrin D, Bird AP, Barlow PN. The solution structure of the domain from MeCP2 that binds to methylated DNA. J Mol Biol 1999; 291:1055–1065.
28. Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell 1997; 88:471–481.
29. Nan X, Tate P, Li E, Bird A. DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol 1996; 16:414–421.
30. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 1998; 393:386–389.
31. Yu F, Thiesen J, Stratling WH. Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2. Nucleic Acids Res 2000; 28:2201–2226.
32. Kaludov NK, Wolffe AP. MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. Nucleic Acids Res 2000; 28:1921–1928.
33. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002; 16:6–21.
34. Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet 1999; 23:62–66.
35. Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A. MBD2 is a transcriptional

repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet 1999; 23:58–61.
36. Feng Q, Zhang Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev 2001; 15:827–832.
37. Fujita N, Takebayashi S, Okumura K, Kudo S, Chiba T, Saya H, Nakao M. Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol 1999; 19:6415–6426.
38. Ng HH, Jeppesen P, Bird A. Active repression of methylated genes by the chromosomal protein MBD1. Mol Cell Biol 2000; 20:1394–1406.
39. Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 1999; 401:301–304.
40. Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R, Masciullo V, Genuardi M, Paravatou-Petsotas M, Bassi DE, Ruggeri BA, Klein-Szanto AJ, Testa JR, Neri G, Bellacosa A. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. Nat Genet 1999; 23:266–268.
41. Prokhortchouk A, Hendrich B, Jorgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev 2001; 15:1613–1618.
42. Guy J, Hendrich B, Holmes M, Martin JE, Bird A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet 2001; 27:322–326.
43. Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev 2001; 15:710–723.
44. Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science 2002; 297:403–405.
45. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 1997; 389:251–260.
46. Luger K, Richmond TJ. DNA binding within the nucleosome core. Curr Opin Struct Biol 1998; 8:33–40.
47. Jenuwein T, Allis CD. Translating the histone code. Science 2001; 293:1074–1080.
48. Turner BM. Cellular memory and the histone code. Cell 2002; 111:285–291.
49. Marmorstein R, Roth SY. Histone acetyltransferases: function, structure, and catalysis. Curr Opin Genet Dev 2001; 11:155–161.
50. Sterner DE, Berger SL. Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev 2000; 64:435–459.
51. Johnstone RW. Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. Nat Rev Drug Discov 2002; 1:287–299.
52. Thiagalingam S, Cheng KH, Lee HJ, Mineva N, Thiagalingam A, Ponte JF. Histone deacetylases: unique players in shaping the epigenetic histone code. Ann N Y Acad Sci 2003; 983:84–100.

53. Havas K, Flaus A, Phelan M, Kingston R, Wade PA, Lilley DM, Owen-Hughes T. Generation of superhelical torsion by ATP-dependent chromatin remodeling activities. Cell 2000; 103:1133–1142.
54. Narlikar GJ, Fan HY, Kingston RE. Cooperation between complexes that regulate chromatin structure and transcription. Cell 2002; 108:475–487.
55. Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev 2001; 15:2343–2360.
56. Wang H, Cao R, Xia L, Erdjument-Bromage H, Borchers C, Tempst P, Zhang Y. Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Mol Cell 2001; 8:1207–1217.
57. Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P, Reinberg D. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modificatios required for heterochromatin formation. Genes Dev 2002; 16:479–489.
58. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 2000; 406: 593–599.
59. Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H, Shinkai Y. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 2002; 16:1779–1791.
60. Yang L, Xia L, Wu DY, Wang H, Chansky HA, Schubach WH, Hickstein DD, Zhang Y. Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. Oncogene 2002; 21:148–152.
61. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB Zinc-finger proteins. Genes Dev 2002; 16:919–932.
62. Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. Science 2002; 296:1132–1136.
63. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 2002; 3:662–673.
64. Ferguson-Smith A. All the controls: genomic imprinting and the epigenetic regulation of gene expression. In: Beck S, Olek A, eds. The Epigenome: Molecular Hide and Seek. Weinheim: Wiley-VCH, 2003:81–99.
65. Lee J, Inoue K, Ono R, Ogonuki N, Kohda T, Kaneko-Ishino T, Ogura A, Ishino F. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development 2002; 129:1807–1817.
66. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science 2001; 293:1089–1093.
67. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. Nature 2000; 403:501–502.

68. Avner P, Heard E. X-chromosome inactivation: counting, choice and initiation. Nat Rev Genet 2001; 2:59–67.
69. Cohen DE, Lee JT. X-chromosome inactivation and the search for chromosome-wide silencers. Curr Opin Genet Dev 2002; 12:219–224.
70. Boggs BA, Cheung P, Heard E, Spector DL, Chinault AC, Allis CD. Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat Genet 2002; 30:73–76.
71. Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell 2001; 107:727–738.
72. Csankovszki G, Nagy A, Jaenisch R. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J Cell Biol 2001; 153:773–784.
73. Bird AP. Gene number, noise reduction and biological complexity. Trends Genet 1995; 11:94–100.
74. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 1997; 13:335–340.
75. Kazazian HH Jr, Moran JV. The impact of L1 retrotransposons on the human genome. Nat Genet 1998; 19:19–24.
76. Montagna M, Santacatterina M, Torri A, Menin C, Zullato D, Chieco-Bianchi L, D'Andrea E. Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families. Oncogene 1999; 18:4160–4165.
77. Fraga MF, Esteller M. DNA methylation: a profile of methods and applications. Biotechniques 2002; 33:632, 634, 636–649.
78. Pfeifer GP, Steigerwald SD, Hansen RS, Gartler SM, Riggs AD. Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. Proc Natl Acad Sci USA 1990; 87:8252–8256.
79. Riggs AD, Xiong Z, Wang L, LeBon JM. Methylation dynamics, epigenetic fidelity and X chromosome sturcutre. Novartis found symp 1998; 214:214– discussio 225–232.
80. Rakyan VK, Preis J, Morgan HD, Whitelaw E. The marks, mechanisms and memory of epigenetic states in mammals. Biochem J 2001; 356:1–10.
81. Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 2004.
82. Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S, Wu Y, He X, Powe NR, Feinberg AP. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. Science 2003; 299:1753–1755.
83. Xu GL, Bestor TH, Bourch'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 1999; 402:187–191.
84. Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl acad Sci USA 1999; 96: 14412–14417.

85. Ehrlich M, Buchanan KL, Tsien F, Jiang G, Sun B, Uicker W, Weemaes CM, Smeets D, Sperling K, Belohradsky BH, Tommerup N, Misek DE, Rouillard JM, Kuick R, Hanash SM. DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. Hum Mol Genet 2001; 10:2917–2931.
86. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 1999; 23:185–188.
87. Jin P, Warren ST. Understanding the molecular basis of fragile X syndrome. Hum Mol Genet 2000; 9:901–908.
88. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet 1999; 21: 163–167.
89. Kopelovich, L, Crowell, JA, Fay, JR. The epigenome as a target for cancer chemoprevention. J Natl Cancer Inst 2003; 95:1747–1757.
90. Baylin S, Bestor TH. Altered methylation patterns in cancer cell genomes: cause or consequence? Cancer Cell 2002; 1:299–305.
91. Petronis A. Human morbid genetics revisited: relevance of epigenetics. Trends Genet 2001; 17:142–146.
92. Petronis A, Gottesman II, Kan P, Kennedy JL, Basile VS, Paterson AD, Popendikyte V. Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? Schizophr Bull 2003; 29:169–178.
93. Holliday R. The inheritance of epigenetic defects. Science 1987; 238:163–170.
94. Popendikyte V, Laurinavicius A, Paterson AD, Macciardi F, Kennedy JL, Petronis A. DNA methylation at the putative promoter region of the human dopamine D2 receptor gene. Neuroreport 1999; 10:1249–1255.
95. Kim RB. Pharmacogenetics of CYP enzymes and drug transporters: remarkable recent advances. Adv Drug Deliv Rev 2002; 54:1241–1242.
96. Liggett SB. Polymorphisms of adrenergic receptors: variations on a theme. Assay Drug Dev Technol 2003; 1:317–326.
97. Risinger JI, Maxwell GL, Berchuck A, Barrett JC. Promoter hypermethylation as an epigenetic component in Type I and Type II endometrial cancers. Ann NY Acad Sci 2003; 983:208–212.
98. Wu MS, Lee CW, Shun CT, Wang HP, Lee WJ, Chang MC, Sheu JC, Lin JT. Distinct clinicopathologic and genetic profiles in sporadic gastric cancer with different mutator phenotypes. Genes Chromosomes Cancer 2000; 27:403–411.
99. Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E, Marth C, Widschwendter M. DNA methylation in serum of breast cancer patients: an independent prognostic marker. Cancer Res 2003; 63:7641–7645.
100. Fruhwald MC. DNA methylation patterns in cancer: novel prognostic indicators?. Am J Pharmacogenomics 2003; 3:245–260.
101. Brock MV Gou M Akiyama Y Muller A Wu TT Montgomery E Deasel M Germonpre P Rubinson L Heitmiller RF Yang SC Forastiere AA Baylin SB Herman JG. Prognostic importance of promoter hypermethylation of multiple genes in esophageal adenocarcinoma. Clin Cancer Res 2003; 9: 2912–2919.
102. Kondo Y, Issa JP. Epigenetic changes in colorectal cancer. Cancer Metastasis Rev 2004; 23:29–39.

103. Ravenel JD, Broman KW, Perlman EJ, Niemitz EL, Jayawardena TM, Bell DW, Haber DA, Uejima H, Feinberg AP. Loss of imprinting of insulin-like growth factor-II (IGF2) gene in distinguishing specific biologic subtypes of Wilms tumor. J Natl Cancer Inst 2001; 93:1698–1703.
104. Wei SH, Brown R, Huang TH. Aberrant DNA methylation in ovarian cancer: is there an epigenetic predisposition to drug response?. Ann NY Acad Sci 2003; 983:243–250.
105. Johnstone RW, Licht JD. Histone deacetylase inhibitors in cancer therapy: is transcription the primary target? Cancer Cell 2003; 4:13–18.
106. Maio M, Coral S, Fratta E, Altomonte M, Sigalotti L. Epigenetic targets for immune intervention in human malignancies. Oncogene 2003; 22:6484–6488.
107. Kalebic T. Epigenetic changes: potential therapeutic targets. Ann NY Acad Sci 2003; 983:278–285.
108. Gottlicher M, Minucci S, Zhu P, Kramer OH, Schimpf A, Giavara S, Sleeman JP, Lo Coco F, Nervi C, Pelicci PG, Heinzel T. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. EMBO J 2001; 20:6969–6978.
109. Piekarz RL, Robey R, Sandor V, Bakke S, Wilson WH, Dahmoush L, Kingma DM, Turner ML, Altemus R, Bates SE. Inhibitor of histone deacetylation, depsipeptide (FR901228), in the treatment of peripheral and cutaneous T-cell lymphoma: a case report. Blood 2001; 98:2865–2868.
110. Marshall JL, Rizvi N, Kauh J, Dahut W, Figuera M, Kang MH, Figg WD, Wainer I, Chaissang C, Li MZ, Hawkins MJ. A phase I trial of depsipeptide (FR901228) in patients with advanced cancer. J Exp Ther Oncol 2002; 2: 325–332.
111. Sandor V, Bakke S, Robey RW, Kang MH, Blagosklonny MV, Bender J, Brooks R, Piekarz RL, Tucker E, Figg WD, Chan KK, Goldspiel B, Fojo AT, Balcerzak SP, Bates SE. Phase I trial of the histone deacetylase inhibitor, depsipeptide (FR901228, NSC 630176), in patients with refractory neoplasms. Clin Cancer Res 2002; 8:718–728.
112. Gilbert J, Baker SD, Bowling MK, Grochow L, Figg WD, Zabelina Y, Donehower RC, Carducci MA. A phase I dose escalation and bioavailability study of oral sodium phenylbutyrate in patients with refractory solid tumor malignancies. Clin Cancer Res 2001; 7:2292–2300.
113. Goodman JI, Watson RE. Altered DNA methylation: a secondary mechanism involved in carcinogenesis. Annu Rev Pharmacol Toxicol 2002; 42:501–525.
114. Esteller M, Fraga MF, Paz MF, Campo E, Colomer D, Novo FJ, Calasanz MJ, Galm O, Guo M, Benitez J, Herman JG. Cancer epigenetics and methylation. Science 2002; 297:1807–1808; discussion 1807–1808.
115. Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D. 5′ CpG island methylation is associated with transcriptional silencing of the tumor suppressor p16/CDKN2/MTS1 in human cancers. Nat Med 1995; 1:686–692.
116. Costello JF, Plass C. Methylation matters. J Med Genet 2001; 38:285–303.
117. De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. Proc Natl Acad Sci USA 1996; 93:7149–7153.

118. Christman JK. 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 2002; 21:5483–5495.
119. Jones PA, Taylor SM. Hemimethylated duplex DNAs prepared from 5-azacytidine-treated cells. Nucleic Acids Res 1981; 9:2933–2947.
120. Zimmermann FK, Scheel I. Genetic effects of 5-azacytidine in Saccharomyces cerevisiae. Mutat Res 1984; 139:21–24.
121. Lomen PL, Baker LH, Neil GL, Samson MK. Phase I study of 5-azacytidine (NSC-102816) using 24-hour continuous infusion for 5 days. Cancer Chemother Rep 1975; 59:1123–1126.
122. Shnider BI, Baig M, Colsky J. A phase I study of 5-azacytidine (NSC-102816). J Clin Pharmacol 1976; 16:205–212.
123. Velez-Garcia E, Vogler WR, Bartolucci AA, Arkun SN. Twice weekly 5-azacytidine infusion in disseminated metastatic cancer: a phase II study. Cancer Treat Rep 1977; 61:1675–1677.
124. Weiss AJ, Metter GE, Nealon TF, Keanan JP, Ramirez G, Swaiminathan A, Fletcher WS, Moss SE, Manthei RW. Phase II study of 5-azacytidine in solid tumors. Cancer Treat Rep 1977; 61:55–58.
125. Cheng JC, Matsen CB, Gonzales FA, Ye W, Greer S, Marquez VE, Jones PA, Selker EU. Inhibition of DNA methylation and reactivation of silenced genes by zebularine. J Natl Cancer Inst 2003; 95:399–409.
126. Szyf M. DNA methylation and cancer therapy. Drug Resist Updat 2003; 6:341–353.
127. Milutinovic S, Zhuang Q, Niveleau A, Szyf M. Epigenomic stress response. Knockdown of DNA methyltransferase 1 triggers an intra-S-phase arrest of DNA replication and induction of stress response genes. J Biol Chem 2003; 278:14985–14995.
128. Goffin J, Eisenhauer E. DNA methyltransferase inhibitors-state of the art. Ann Oncol 2002; 13:1699–1716.
129. Fournel M, Sapieha P, Beaulieu N, Besterman JM, MacLeod AR. Down-regulation of human DNA-(cytosine-5) methyltransferase induces cell cycle regulators p16(ink4A) and p21(WAF/Cip1) by distinct mechanisms. J Biol Chem 1999; 274:24250–24256.
130. Medzhitov R, Preston-Hurlburt P, Kopp E, Stadlen A, Chen C, Ghosh S. Janeway CA. Jr. MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. Mol Cell 1998; 2:253–258.
131. Underhill DM, Ozinsky A. Toll-like receptors: key mediators of microbe detection. Curr Opin Immunol 2002; 14:103–110.
132. Klinman DM. Immunotherapeutic uses of CpG oligodeoxynucleotides. Nat Rev Immunol 2004; 4:249–258.
133. Hemmi H, Takeuchi O, Kawai T, Kaisho T, Sato S, Sanjo H, Matsumoto M, Hoshino K, Wagner H, Takeda K, Akira S. A Toll-like receptor recognizes bacterial DNA. Nature 2000; 408:740–745.
134. Takeshita F, Leifer CA, Gursel I, Ishii KJ, Takeshita S, Gursel M, Klinman DM. Cutting edge: Role of Toll-like receptor 9 in CpG DNA-induced activation of human cells. J Immunol 2001; 167:3555–3558.

135. Aderem A, Ulevitch RJ. Toll-like receptors in the induction of the innate immune response. Nature 2000; 406:782–787.
136. Reik A, Gregory PD, Urnov FD. Biotechnologies and therapeutics: chromatin as a target. Curr Opin Genet Dev 2002; 12:233–242.
137. Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys2His2 zinc finger proteins. Annu Rev Biochem 2001; 70:313–340.
138. Chin JW, Schepartz A. Concerted evolution of structure and function in a miniature protein. J Am Chem Soc 2001; 123:2929–2930.
139. Urnov FD, Rebar EJ. Designed transcription factors as tools for therapeutics and functional genomics. Biochem Pharmacol 2002; 64:919–923.
140. Mapp AK, Ansari AZ, Ptashne M, Dervan PB. Activation of gene expression by small molecule transcription factors. Proc Natl Acad Sci USA 2000; 97:3930–3935.
141. Liu PQ, Reber EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, Zhong X, Lee YL, Eisenberg SP, Spratt SK, Case CC, Wolffe AP. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. J Biol Chem 2001; 276:11323–11334.
142. Jouvenot Y, Ginjala V, Zhang L, Liu PQ, Oshimura M, Feinberg AP, Wolffe AP, Ohlsson R, Gregory PD. Targeted regulation of imprinted genes by synthetic zinc-finger transcription factors. Gene Ther 2003; 10:513–522.
143. Beerli, R.R., Dreier, B. & Barbas, C.F., 3rd. Positive and negative regulation of endogenous genes by designed transcription factors. Proc Natl Acad Sci USA 2000; 97:1495–1500.
144. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. & Barbas C.F., 3rd. Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem 2001; 276:29466–29478.
145. Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP, Cose CC. Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. J Biol Chem 2000; 275:33850–33860.
146. Bartsevich VV, Juliano RL. Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries. Mol Pharmacol 2000; 58:1–10.

21

WWW Bioinformatics Resources

S. W. Tang
University of Hong Kong and University of California, Irvine, California, U.S.A.
D. M. Helmeste
University of California, Irvine, California, U.S.A.

1. INTRODUCTION

The Internet represents an ideal platform for the retrieval, exchange, and computing of the voluminous, heterogeneous, and constantly evolving and expanding body of genetic information. The genome research community was indeed an early adopter of the Web and computer technology. While it was the complexity, heterogeneity, and size of the data, which led to the extensive use of the computer and the WWW by the genome research community, bioinformatics itself has grown rapidly into a distinct field of science. Biologists, computer scientists, mathematicians, and statisticians are finding a previously non-existing common ground of interaction and areas of mutual interest. It is also because of this interesting common ground of biology, computers, and mathematics that many newcomers to the field of genomic research are baffled when they encounter many of the unfamiliar methods and terminology. This chapter serves as an introduction to the rich WWW resources on bioinformatics for those new to the field.

2. WORLD WIDE WEB (WWW) RESOURCES ON BIOINFORMATICS

Bioinformatics involves two elements: databases and computing tools to operate on the databases. Thus, bioinformatics could be seen as an application of computing tools to collect, organize, and analyze the biological databases gathered or created. The rise of bioinformatics as a specialty was closely tied to the human genome project when huge data sets were generated rapidly through multiple laboratories around the world. As stated by the Department of Energy Human Genome Project (http://www.ornl.gov/sci/techresources/Human_Genome/research/informatics.shtml), there is an ongoing need to improve the content and utility of databases. Another goal is to develop robust software to handle the areas of: generation, capture, annotation, comprehensive functional studies, and for representing and analyzing sequence similarity and variation.

Bioinformatics resources are richly represented on the Web. The Weizmann GeneCard site rightly referred to the vast amount of data sources and their high degree of heterogeneity as an "information labyrinth" or "data labyrinth." Indeed, much new information and many new features have been added to the bioinformatics sites we referenced in the 2001 first edition. To a newcomer, bioinformatics information in the Web is hopelessly scattered. The situation is not different from what you would encounter in a library, where a topic of interest may be scattered among the many books in different areas. While many of the books seem useful and interesting, it is impractical to search through every one. There would be redundancy and unnecessary information scattered among the needed data. Modern Internet search engines are like indexing systems in a library except that they are much more powerful. To illustrate the magnitude of the WWW resources accumulated, a search using the key word bioinformatics returned 2,120,000 sites from Yahoo and 2,990,000 from Google in mid-2004. As Web resources are not static, the numbers will be different but probably higher again when readers try out their own searches after reading this chapter.

The numerous bioinformatics sites can be grouped in many different ways. For example, these sites could be grouped according to the type of organization sponsoring the site (e.g., governmental, university, or commercial), the subject matter emphasized (e.g., proteins, nucleotides, or diseases), or their functional nature (e.g., sequence search, structure database search, analysis, or education). Some sites concentrate on special interest topics while others are multipurpose and comprehensive. There are university sites that focus on a particular topic or purpose. There are also sites put together by individual professors or departments. Irrespective of the heterogeneity of the Web resources on bioinformatics, most major sites nowadays include

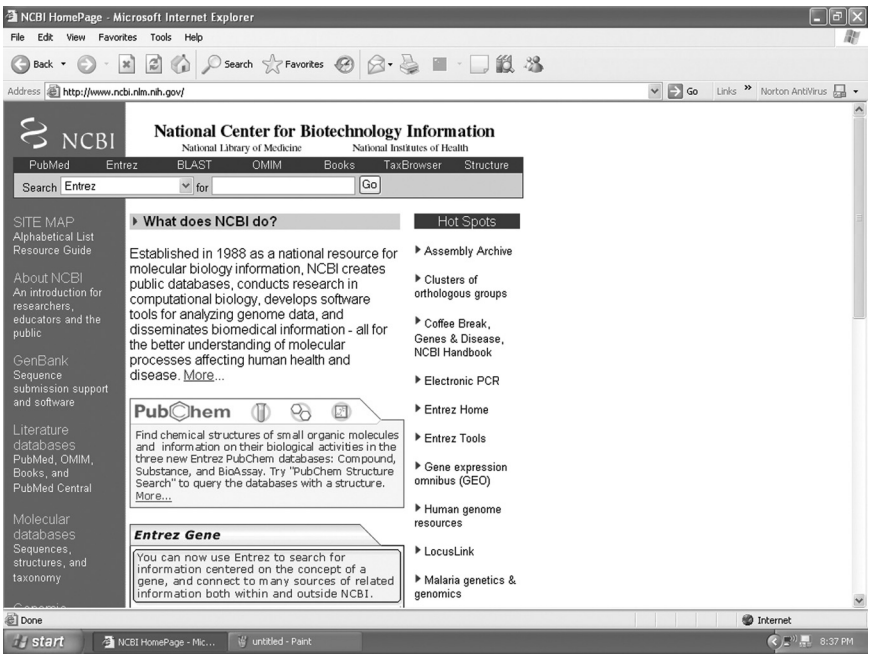
common databases links, software tools, educational materials, courses, and cross-links to other sites.

3. MAJOR SITES

Similar to Web portals in other areas, due to the large amount of information generated through the human, animal, and plant genome projects in recent years and the tremendous financial and human resources required in managing such information, public funded sites have now become huge and dominate the field. At the NCBI site, one could find pointers to almost any genome, molecular biology, or bioinformatics topic and locate any major software tool. In becoming a mega site itself, the NCBI site in certain ways has also become an ''information labyrinth'' and difficult to navigate for a new user.

3.1. National Center for Biotechnology Information (NCBI) Web Site (http://www.ncbi.nlm.nih.gov)

The NCBI's homepage gives the following description: ''Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical

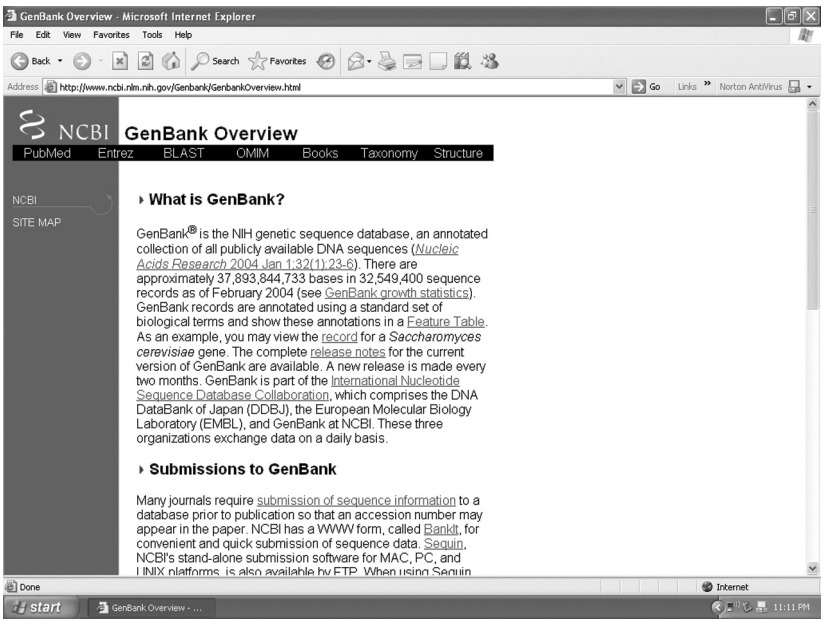


information.'' The NCBI site is probably the most comprehensive bioinformatics site at present.

The NCBI Home page provides links to a large number of bioinformatics databases and tools, under titles such as ''GenBank,'' ''Molecular databases,'' and ''Genomic Biology and Software.'' Much information could also be retrieved through these major links as well as links listed in other pages such as the ''hotspots'' on the NCBI homepage. The hotspots topics include databases and tools such as ''clusters of orthologous groups,'' ''gene expression omnibus,'' ''Human genome resources,'' ''Map viewers,'' ''mouse and rat genome resources,'' ''dbMHC,'' ''retrovirus resources,'' etc. It is almost impossible for the determined researcher not to be able to find their area of interest using the NCBI portal. The problem for a newcomer is that they might not know where to begin their search, or how to handle the duplicative information retrieved.

3.1.1. GenBank

GenBank is the NIH genetic sequence database, and an annotated collection of all publicly available DNA sequences. It is part of the International Nucleotide Sequence Database Collaboration, which comprises the GenBank at NCBI, DNA DataBank of Japan (DDBJ), and the European Molecular Biology Laboratory (EMBL). These three organizations exchange data on a daily basis.



To illustrate the magnitude of growth of the GenBank database, GenBank has published a graph and table showing the statistics from 1982 to 2002. These could be viewed at http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html. After an essentially flat growth from 1982 to 1990, it took off exponentially. In 1982, there were 680,338 bp and 606 sequences in the database. In 1990, this grew to 49,179,285 bp and 39,533 sequences. The numbers increased to 2,008,761,784 and 2,837,897 in 1998, 3,841,163,011 and 4,864,570 in 1999, and 28,507,990,166 and 22,318,883 in 2002. There are approximately 37,893,844,733 bp and 32,549,400 sequence records in the GenBank, as of February 2004.

3.1.2. Data Submission

Data submission is an important function of the GenBank site. The most important sources of new data for GenBank® are direct submissions from scientists.

3.1.3. Search

The NCBI's Entrez is a powerful database search engine. This integrated search engine provides a menu (http://www.ncbi.nih.gov/Entrez/index.html) offering the user selections to search biomedical literature (PubMed), databases on OMIM, nucleotide sequence, protein sequence, whole genome sequences, 3D macromolecular structures, taxonomy (organisms in GenBank), SNP,

Eukaryotic homology, gene-orientated clusters of transcript sequences, conserved protein domain, markers and mapping, population study (PopSet), expression and molecular abundance profiles (GEO), and GEO data sets and cancer chromosomes.

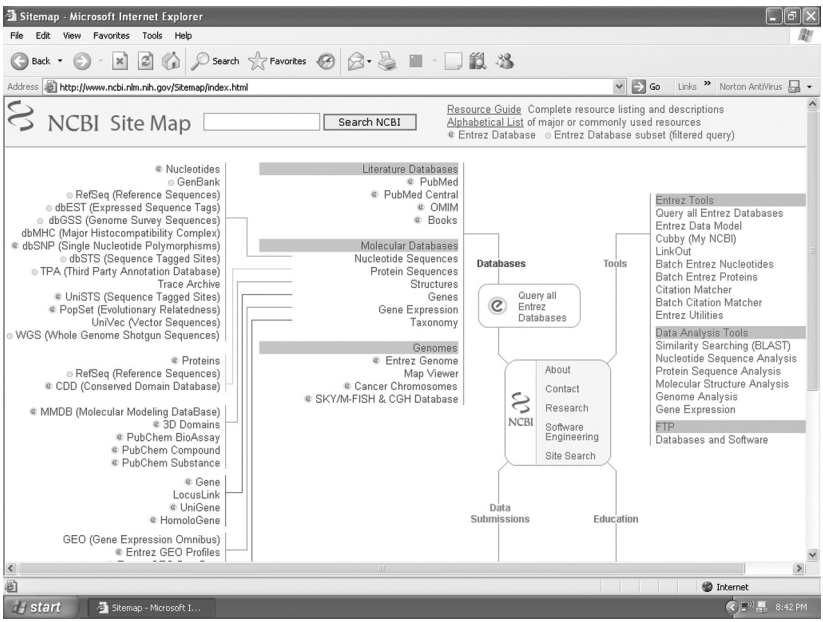
3.1.4. Resources

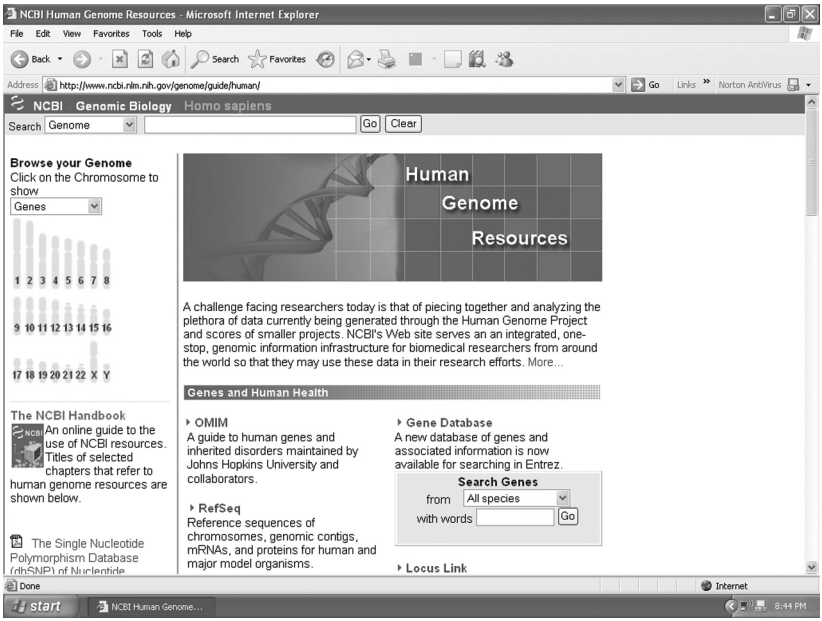
The NCBI genome and bioinformatics resources are also listed at the following pages:

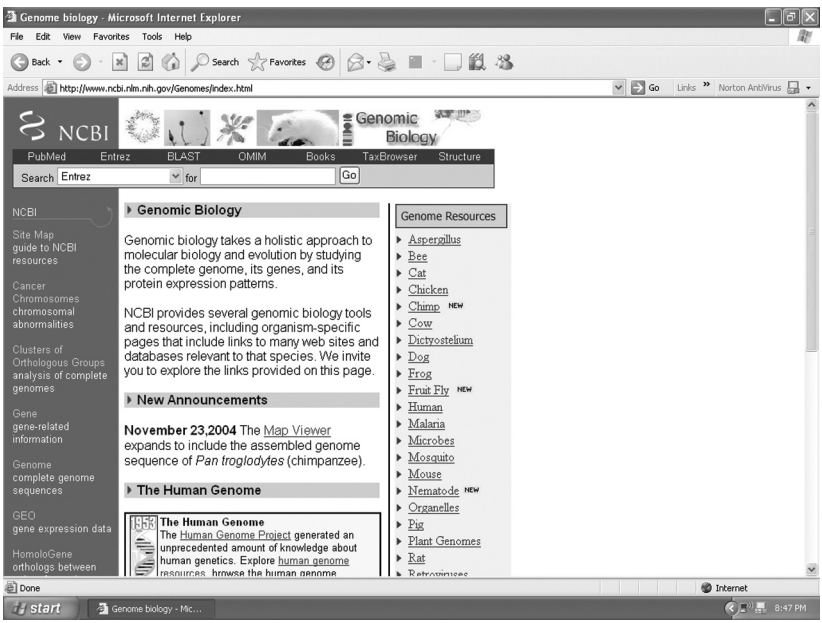
http://www.ncbi.nih.gov/Sitemap/index.html
http://www.ncbi.nlm.nih.gov/genome/guide/human
http://www.ncbi.nlm.nih.gov/Genomes/index.html

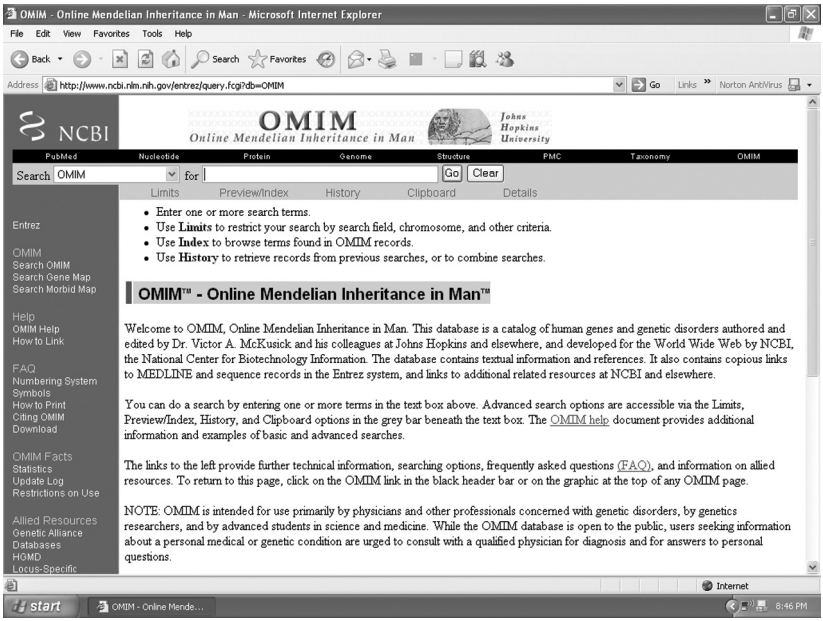
These pages provide links to larger resource categories such as GenBank, Molecular databases, Genomes and Maps, Tools, Software engineering, and FTP sites. There is also a table with a comprehensive alphabetical index of Web resources.

Although the NCBI site is the most comprehensive site of all, indexing of the diverse topics in and molecular biology and bioinformatics is still far from ideal. Unless you are a frequent visitor, newcomers can easily be overwhelmed by the tremendous amount of pointers, or links to various topics, which may not be clearly related to each other. For example, you could access human chromosomal information via the page on OMIM—









Online Mendelian Inheritance in Man (http://www.ncbi.nih.gov/entrez/query.fcgi?db=OMIM) or the page on human genome resources (http://www.ncbi.nlm.nih.gov/genome/guide/human), whereas the address with only small difference (http://www.ncbi.nlm.nih.gov/Genomes/index.html) would take you to a page with information on other animals.

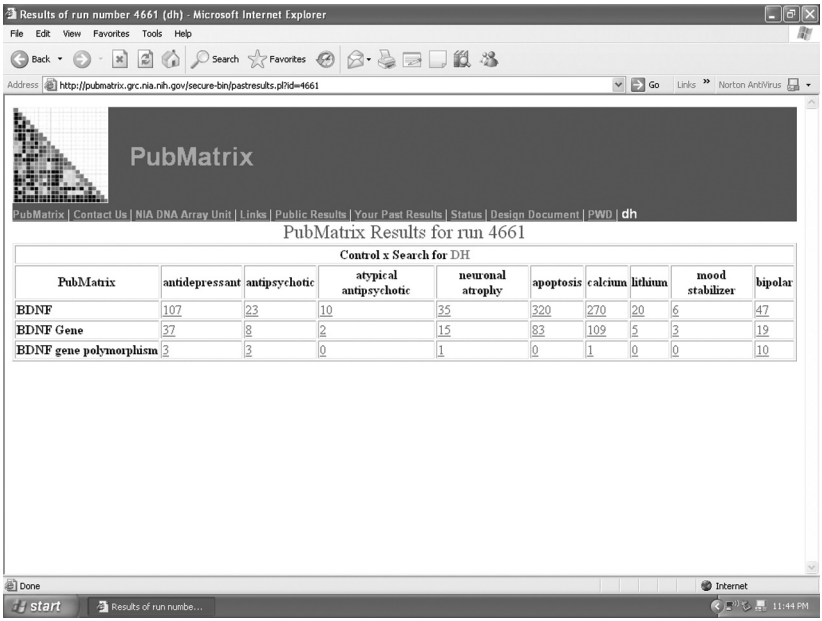
3.1.5. PubMatrix: (http://pubmatrix.grc.nia.nih.gov)

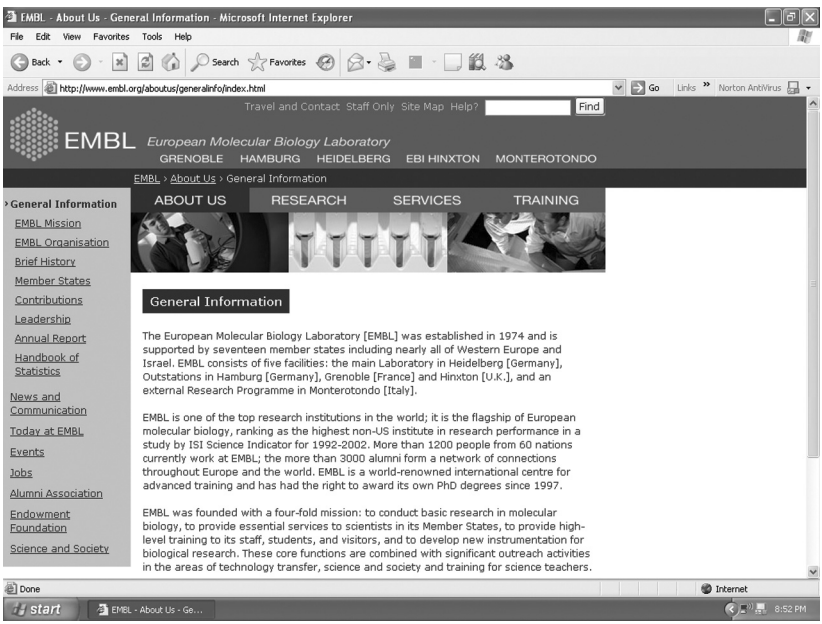
PubMatrix is a simple way to rapidly and systematically compare any list of terms against any other list of terms in PubMed. It reports back the frequency of co-occurrence between all pairwise comparisons between the two lists as a matrix table. It allows the user to build up tables of words in the context of experiments. This is useful for analyzing combinatorial datasets such as cDNA microarrays, genomic, proteomic, or other multiplex comparisons.

A simple example searching for BDNF, BDNF Gene, and BDNF gene polymorphism against the words antidepressant, antipsychotic, atypical antipsychotic, neuronal atrophy, apoptosis, calcium, lithium, mood stabilizer, and bipolar is shown in the figure as a demonstration.

3.2. European Molecular Biology Laboratory

The European Molecular Biology Laboratory (EMBL) (http://www.embl.org) is publicly funded by 17 member states, including most of the EU,



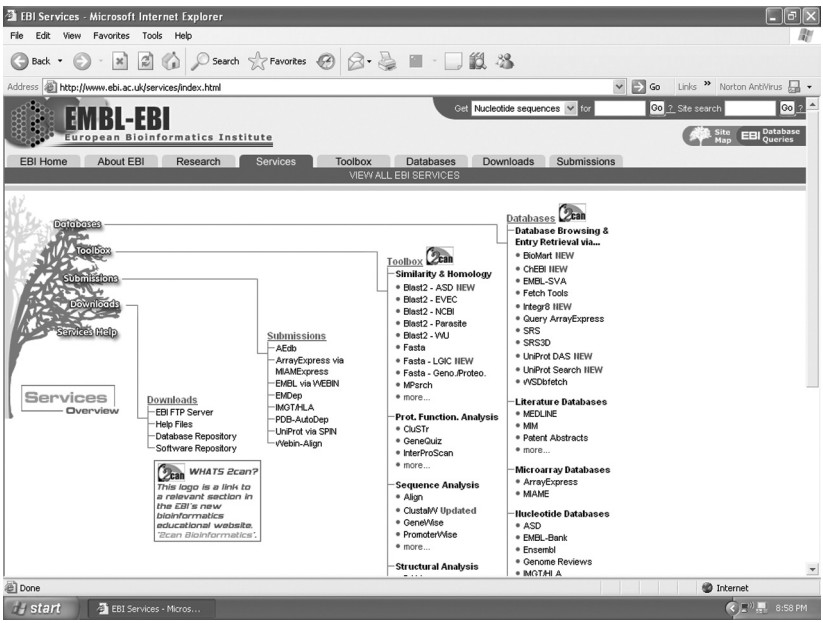


Switzerland, and Israel. It is the flagship of European molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg (Germany), Outstations in Hinxton (U.K.) (the European Bioinformatics Institute), Grenoble (France), Hamburg (Germany), and Monterotondo near Rome (Italy).

3.2.1. European Bioinformatics Institute

Part of the EMBL, the European Bioinformatics Institute (EBI) is a centre for research and services in bioinformatics. The mission of the EBI is to ensure that information from molecular biology and genome research is placed in the public domain and is accessible freely to all facets of the scientific community. The Institute manages databases of biological data including nucleic acid, protein sequences, and macromolecular structures.

In contrast to the complex NCBI site, the EBI offers a more clearly organized home page. A comprehensive array of services is listed nicely on the page http://www.ebi.ac.uk/services/index.html, under four headings: databases, toolbox, submissions, and downloads. Software tools are very well organized under the toolbox category and include all the common tools such as Blast and Fasta for homology and similarity analysis, GeneQuiz for protein functional analysis, Align and GeneWise for sequence analysis, and DALI and MSD for structural analysis. Databases are nicely listed with their respective search tools: Database browsing and retrieval



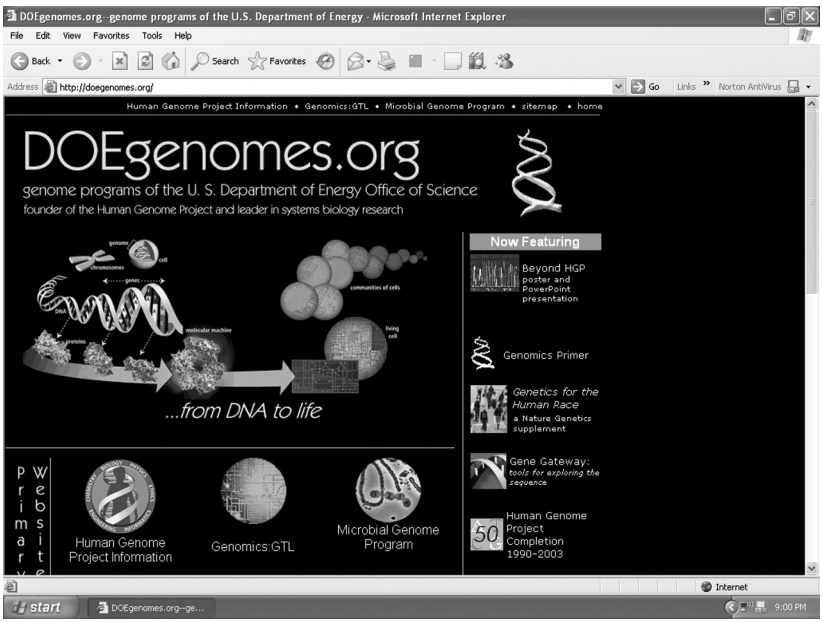
(via SRS, SRS3D, and others), literature databases (including MEDLINE, MIM, and others), microarray databases (via ArrayExpress and MIAME), nucleotide databases, protein databases, and structural databases.

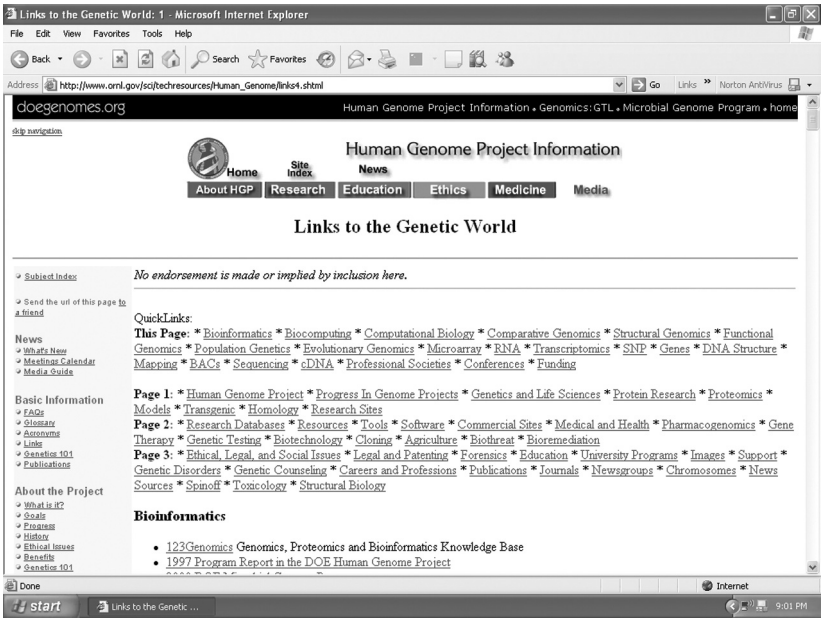
3.3. Genome Programs of the Department of Energy Office of Science (http://doegenomes.org)

The Department of Energy's website is interesting. Under the address: http://www.ornl.gov/sci/techresources/Human_Genome/links4.shtml its main topics include Bioinformatics, Biocomputing, Computational Biology, Comparative Genomics, Structural Genomics, Functional Genomics, Population Genetics, Evolutionary Genomics, Microarray, RNA, Transcriptomics, SNP, Genes, DNA Structure, Mapping, BACs Sequencing, cDNA, and their subtopics. It gives an exhaustive list of useful sites leading to a huge menu of bioinformatics and genome research. Obviously, it is more of an information site and is not as orientated towards researchers, in contrast to the NCBI and EMBL–EBI sites.

4. SOME OTHER GOOD SITES

Compared to the comprehensive NCBI and EBI sites, many other bioinformatics sites have become more "local" in nature, emphasizing their research topics, or software they developed, with some links to common tools and





databases. The disparity in resources between the smaller organizations and bigger ones like the NCBI and EBI limits these other sites from becoming the main portals to bioinformatics.

4.1. U.K. HGMP Resource Center—MRC Rosalind Franklin Center for Genomics Research Bioinformatics (http://www.hgmp.mrc.ac.uk/Bioinformatics/)

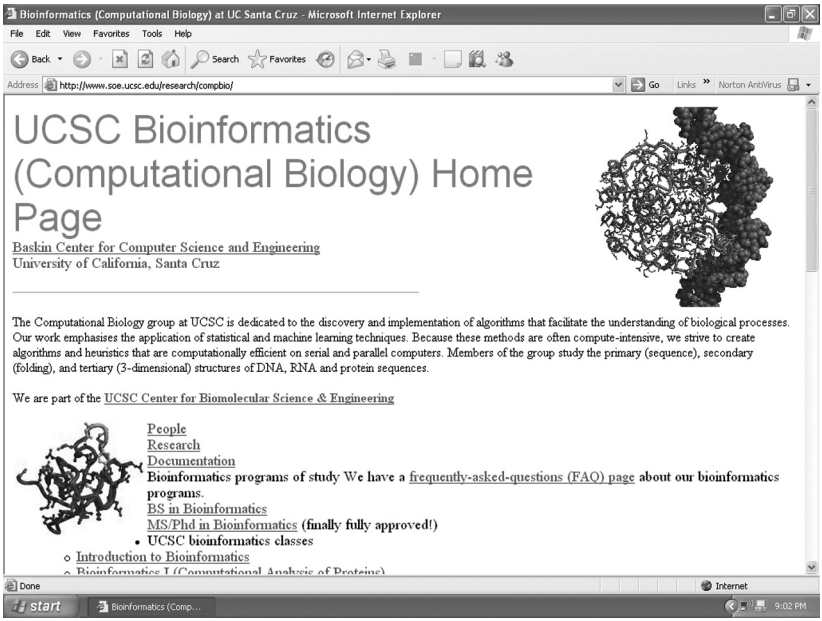
This site gives registered users access to a collection of programs and databases to aid genomic and proteomic research.

4.2. Bioinformatics at UC Santa Cruz (http://www.soe.ucsc.edu/research/compbio/)

This site has an extensive coverage of bioinformatics and computational biology sites.

4.3. Weizmann Institute of Science Crown Human Genome Center (http://bioinformatics.weizmann.ac.il/genome_center/)

This site gives good information regarding many of its research projects and education materials.



4.4. Whitehead/MIT Center for Genome Research (http://www.broad.mit.edu/resources.html)

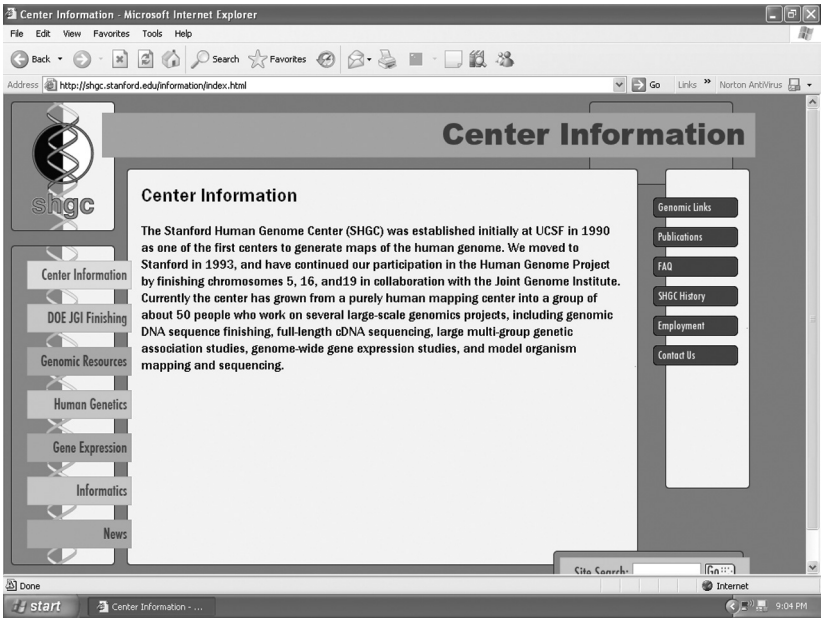
This site has good databases on human and mouse SNP, Archaebacterium, fungal and bacterium sequences, and a good collection of software and links for mapping projects.

4.5. Washington University Genome Sequencing Center (http://genome.wustl.edu)

The Genome Sequencing Center focuses on the large-scale generation and analysis of DNA sequences. They have played an important role in the Human Genome Project. They also sequenced the genome of other species such as the chicken, C. elegans, planarian, chimpanzee, and drosophila.

4.6. Stanford Human Genome Center (SHGC) (http://shgc.stanford.edu)

The Center was established initially at UCSF in 1990 as one of the first centers to generate maps of the human genome. This small but interesting site provides good information about their many research projects on various aspects of human genetics, including data analysis and search tools. It also offers interesting information and tools for genetic studies on non-human organisms such as Zebrafish and Xenopus.



4.7. University of Manchester Bioinformatics Education and Research (UMBER) Site (http://www.umber.embnet.org/resources)

The site mainly offers useful software for sequence alignment such as ALIGN. It also offers some other useful tools such as PRINTS (Protein motif fingerprint data) and MAXD for microarray data storage and analysis.

4.8. The Institute for Genomic Research (http://www.tigr.org/tdb/index.shtml)

The Insitute for Genomic Research (TIGR) provides a collection of curated databases containing DNA and protein sequences, gene expressions, cellular roles, protein families, and taxonomic data for microbes, plants, and humans. Anonymous FTP access to sequence data is also provided.

5. BIOINFORMATIC AND BIOTECHNOLOGY LAW RESOURCES

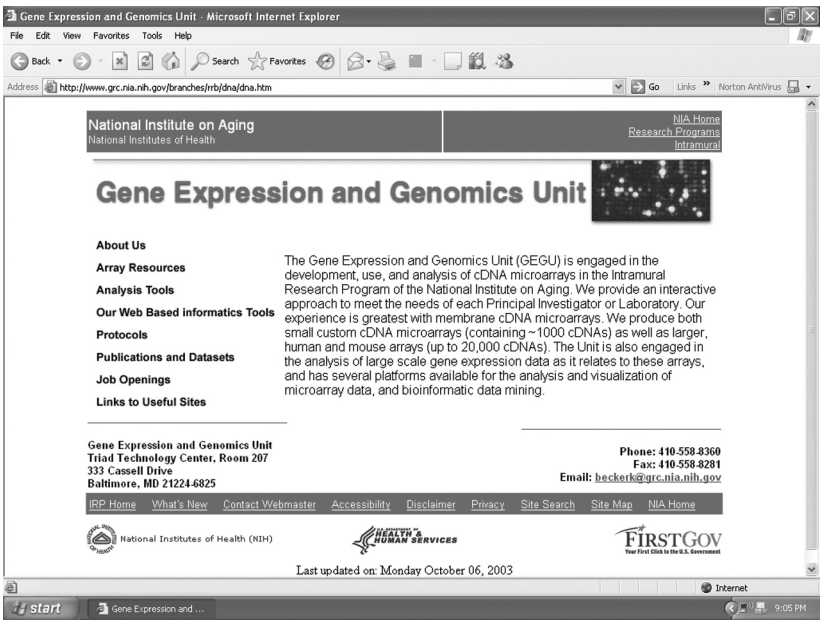
Bioinformatics and biotechnology law resources have been compiled to deal with this other side of research in some useful sites, e.g., University of Washington School of Law site (http://lib.law.washington.edu/ref/biotech.htm) and a commercial site (http://www.denniskennedy.com/bioinformaticslaw.htm).

6. MICROARRAY RELATED SITES

Large-scale gene expression studies, especially microarray technologies, have rapidly become a significant part of bioinformatics. There are many excellent sites focusing on microarray databases and tools for data analysis. The following sites are simply some examples.

6.1. Microarray Gene Expression Data (MGED) Society (http://www.mged.org)

The Society is an international organization of biologists, computer scientists, and data analysts. The purpose is to facilitate the sharing of microarray data from functional genomics and proteomics experiments. The MGED was founded as a grass roots movement in 1999 by many of the major microarray users and developers. The MGED became a non-profit organization in 2002. The current focus of MGED is on establishing standards for microarray data annotation and exchange, facilitating the creation of microarray databases and related software implementing these standards, and promoting the sharing of high quality, well-annotated data. The MIAME (Minimum Information About a Microarray Experiment) stands for the minimum information about a microarray experiment that is needed to enable the interpretation of the results of the experiment unambiguously and potentially reproduce it. The public repositories ArrayExpress at the EBI (http://www.ebi.ac.uk/



arrayexpress) and GEO (Gene Expression Omnibus) (http://www.ncbi.nlm.nih.gov/geo) at NCBI are designed to accept, hold, and distribute MIAME compliant microarray data. A similar database CIBEX is being developed in Japan.

6.2. Microarrays.org (http://www.microarrays.org)

Microarrays.org is a public source for microarray protocols and software. This site is built and maintained by the Department of Biochemistry and Biophysics, University of California at San Francisco.

6.3. GRID IT Resources for Microarray Technology (http://www.bsi.vt.edu/ralscher/gridit)

It is a collaborative project led by Virginia Tech and the Forest Biotechnology Group at North Carolina State University. It provides resources for current microarray users and newcomers alike. It includes links to database searching, microarray analysis tools, sequence analysis tools, and analysis sites that include data.

6.4. National Cancer Institute mAdb (Mad Bee) NCI/Center for Cancer Research (CCR)μArray Center Gateway (http://nciarray.nci.nih.gov/)

It was developed by the Center for Information Technology (CIT) in collaboration with the NCI's Advanced Technology Center (ATC) Array Facility to provide a set of web-based analysis tools and a data management system for storing and mining gene expression data, the system allows registered users to upload and then analyze microarray data from National Cancer Institute (NCI), National Institute of Allergy and Infectious Diseases (NIAID), Food and Drug Administration (FDA) produced, or other microarrays.

6.5. Stanford Microarray Database (SMD; http://genome-www5.stanford.edu)

It stores raw and normalized data from microarray experiments, as well as corresponding image files. In addition, Stanford Microarray Database (SMD) provides interfaces for data retrieval, analysis, and visualization. Access to non-public data is limited to registered Stanford researchers and their collaborators. The site also provides links to other microarray resources (http://genome-www5.stanford.edu/resources.html)

6.6. TREX Program in Genomic Applications

It provides microarray expression profiling of rodent models of human disease and cDNA microarray assays to elucidate patterns of gene expression in heart, lung, blood, and sleep disorders (http://pga.tigr.org)

6.7. Children's National Medical Center Microarray Center

The HOPGENE Program for Genomic Applications provides expression-profiling data for many human diseases and animal models of human disease. (http://microarray.cnmcresearch.org/pgadatatable.asp)

6.8. CardioGenomics

It is one of eleven Programs for Genomic Applications (PGAs) funded by the National Heart, Lung and Blood Institute (NHLBI) of the NIH. The CardioGenomics PGA Public Data (http://cardiogenomics.med.harvard.edu/public-data.tcl) provides microarray data on a number of cardiac-related disease topics.

6.9. Human Gene Expression Index (HuGE Index) (http://132.183.243.28/hio/databases/index.html)

It aims to provide a comprehensive database to understand the expression of human genes in normal human tissues.

6.10. Whitehead Institute Center for Genome Research

It provides microarray data for a number of cancer-related projects (http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi).

7. MICROARRAY SOFTWARE

7.1. Eisen Lab (http://rana.lbl.gov/EisenSoftware.htm)

All software produced by Eisen lab may be downloaded and used free of charge by academic and other non-profit researchers. Software is available for microarray image analysis, cluster analysis and visualization and combined expression data and sequence analysis.

7.2. The Institute for Genomic Research (http://www.tigr.org/software)

The TIGR has software systems available for free download. These include software for gene finding/annotation, alignment, sequencing/finishing, and microarray analysis. All of them are OSI (Open Source Initiative) Certified Open Source Software.

8. EDUCATION SITES IN BIOINFORMATICS

As this field continues to expand, so do the courses offering education in bioinformatics. The two basic models for instruction have not changed much. In the first model, students with either a biology or computer science background enter the program to train in computational biology. Depending on the program, one can be a computer scientist who specializes in biology, or a biologist who specializes in computer science. This allows individuals with either background to enter into a given program and emerge with a relatively even level of expertise in both biology and computer science, or a special emphasis in one. In the second model, students enter into the bioinformatics/computational biology program without a background in either biology or computer science. This may be at the undergraduate or graduate school training levels.

In many of these programs, the molecular biology training includes genetics and biochemistry as they relate to protein structure and function. Specific bioinformatics topics include computational methods in molecular biology. Types of analyses that the students learn include finding genes and gene control regions in a sequence, phylogenetic trees, alignment of sequences, secondary and tertiary structure predictions, and finding features and domains in proteins. In some of these programs, prerequisites may include programming skills (Perl, C++, and Unix) and mathematics of an applied nature (matrix mathematics, linear algebra, and calculus).

At the time of writing (spring, 2004), it has become widespread for educational institutions to offer degree-based programs. Numerous web-based as well as traditional classroom style courses are available. Some examples are the bioinformatics programs at the University of Manchester, U.K. (http://www.umber.embnet.org/education/), Stanford University (http://scpd.stanford.edu/scpd/pdf/bioCerts.pdf), Waterloo University (http://www.cs.uwaterloo.ca/undergrad/programs/plans/bcs_bio.shtml), the Pasteur Institute, France (http://www.pasteur.fr/formation/infobio/infobio-en.html), Oxford University, U.K. (http://www.molbiol.ox.ac.uk/), the Weizmann Institute of Science, Israel (http://bioinformatics.weizmann.ac.il/courses/), and many others. Virtually every major university has bioinformatics courses available. To attempt a complete inventory of such courses and programs would be hopeless since they are expanding and changing every day. The reader is instead referred to the Web-based lists of courses that have been compiled by organizations such as the International Society for Computational Biology (http://www.iscb.org/univ_programs/program_board.php).

9. FUTURE DIRECTIONS

Clinical application of bioinformatics information is expected to increase dramatically with microarrays expected to play a large role. Better efforts

are still needed in assessing, understanding and minimizing variability in DNA microarray results. After this, methods to deal with the complexity of results are very much needed. While it is simple enough to analyze gene expression in tissue "X" after drug "Y" administration, the clinical reality involves dissecting multi-gene expression following multi-drug exposures, which have different dosage and exposure histories in each patient under study. How to deal with large data sets involving very variable gene expression profiles is important for predicting clinical status and response to treatments. This is an area we can expect to see developed in the near future.

GLOSSARY OF TECHNICAL TERMS

Much has changed in the information technology field since we wrote the chapter for the first edition. Many previously unfamiliar terms relating to computer and information technology have become part of daily language. Many clinicians who were late adopters of the computer are now regular users of e-mails, searching the MEDLINE, downloading articles from the Internet and giving lectures with PowerPoint instead of carrying a carousal of slides. Dial-ups have been replaced by broadband connections. Computer CPUs are rarely below the GHz range. It no longer makes sense to give a glossary of the basic terms here. The following are simply a compilation of those common terms in computer sciences and IT, which may not be familiar to researchers coming from a mainly biological background.

802.1 X: is a port access protocol for protecting networks via authentication, extremely useful in the wireless environment due to the nature of the medium. When a wireless device is authenticated via 802.1x for network access, a virtual port is opened on the access point allowing for communication.

802.11b: the Institute of Electrical and Electronic Engineers (IEEE) sets standards for communication through Local Area Networks (LANs). The 802.11b is the most widely deployed wireless LAN standard known as "WiFi," 802.11b utilizes the 2.4 GHz spectrum with a bandwidth of 11 Mbps. A different standard, 802.11a, uses the 5 GHz spectrum and has a capacity of 54 Mbps. The 802.11g also uses 2.4 GHz.

Bandwidth: the measure of telecommunication link capacity. The broader the bandwidth, the more information can be transmitted within a time unit. The industry is looking toward low-cost broad bandwidth connections. University offices and laboratories generally have fast connections via a T1 or greater. For many families, the final connection to the computer via phone lines is generally the limiting factor for fast information transmission. Cable and DSL connection are changing the scenario for some homes. New apartment buildings in some countries now have built-in fiber-optic connections offering broad bandwidth Internet connections.

Bluetooth: the need for low-cost short-range wireless radio networking among different electronic devices such as mobile PCs, cell phones, and

other portable devices has led to the development of the bluetooth protocol by a consortium of wireless product companies. Bluetooth can also be utilized to interconnect two or more devices having such capability. The radio unit is small for low power consumption and allows it to become an integral part of the mobile equipment. Interoperability was another important required feature, so that different electronic components and devices manufactured by different companies could communicate.

Broadband: referring to high-speed data transmission in which the connection can carry a large amount of data. Connections using cable (which use the same connection as cable TV), DSL technologies, large leased lines, or optic fiber can carry large amounts of data at high speed.

Browsers: are software tools designed for browsing Internet and web information. Popular browsers include Internet Explorer and Netscape; both are free and downloadable from the Web. Browsers carry with them different features or capabilities. As many Web sites now require advanced features only found in the later editions of the browsers, it is advisable to update the browser periodically.

Centrino: refers to the mobile wireless and power management technology by Intel. Centrino allows one to connect to the Internet or a corporate network without wires or an add-on adapter card. Wireless LAN (WLAN) uses radio waves to wirelessly connect computers to each other, with ability to connect to 802.11b, 802.11a, and 802.11g based infrastructures. It supports a wide range of industry wireless-LAN security standards.

CGI (Common Gateway Interface)

The CGI is a protocol, not a language. It is the common way browsers communicate with web servers.

CGI Scripts

These are scripts written in some form of language (usually, but not necessarily Perl, PHP, or Python) for server communication.

Client: the user's computer, when visiting a server hosting information, is referred to as the client.

Compilers: a computer does not read human language. Programming languages have been invented by humans to instruct the computer to carry out jobs. A compiler translates codes of a computer programming language into machine language before execution. Some computer languages, such as C and C++, need to be compiled before execution. Some are directly interpreted without compiling. Java is a compiled and interpreted language while Perl is an interpreted language.

.com: indicates commercial in Internet addresses.

.org: indicates organization in Internet addresses.

.edu: indicates a University affiliated Internet address.

C and C++ programming: C is a very popular programming language. It is easy to read, efficient, and flexible. C++ is a programming language that was built off the C language. With a syntax nearly identical to C, it has added object-oriented features.

CD-RW: compact disk rewritable.

Dial-up: refers to establishing of a network connection over the telephone line.

DVD: "Digital Video Disc," a high-capacity optical disc commonly used for storing movies or other video records.

Google: is one of the most popular search engines for searching data and information in the Internet.

HTML: The HTML stands for Hypertext Markup Language. It is the native language of the Web. The HTML is like a text/typing instruction set; documents written in HTML are interpretable by browsers. Web pages written in HTML when browsed through browsers will take their final form as the pages seen on screen. Hypertext, a form of electronic cross-reference, also makes information interactive and links Web documents. The reader may jump from topic to topic unhindered by geography. Browsers do not support all HTML codes the same way, and some browsers do not support certain HTML codes.

http://: refers to the Hyper-Text Transfer Protocol (HTTP), which is the method used to transfer Web pages to your computer.

Java Applets and Applications: Java applets are small programs written in Java, which can be deployed over the Internet to the users' computers and viewed within a HTML page. Java applications are formal computer programs written in Java, but are stand-alone programs not requiring a browser to run.

Intranet: in contrast to Internet, an Intranet is a private network used strictly within the confines of a company, university, or organization. "Inter" means "between or among," hence the difference between the Internet and an Intranet.

IT: stands for "Information Technology." It refers to the whole computing technology and industries related to it, such as networking, hardware, software, the Internet, or the people that work with these technologies.

Perl: Perl is an acronym for Practical Extraction and Reporting Language. It is an interpreted language developed by Larry Wall in 1986. It is considered a combination of various UNIX tools, shell scripting, C, and occasional OOP features. Perl is a powerful text (string, arrays, and lists) and pattern manipulation tool. It is cross-platform and portable. These attributes make it a suitable language for data manipulation and genetic research. There are many uses for WWW interactivity (processing forms, guestbooks, and counting visits), and Perl has been generally accepted as the standard for writing CGI scripts and Web server maintenance. Some Perl terminology encountered in bioinformatics includes scalars ($, a piece

of data, constant or variable), arrays (@, collection of scalars), and hashes (%, collection of scalar in pairs).

Programming Languages: computer languages are instructions written according to a specification to direct the computer to perform certain tasks. Common modern programming languages include C, C++, and Java. Perl in particular is the common computer language for bioinformatics applications. Markup languages such as HTML are not traditionally referred to as programming languages.

Programming refers to a set of coded instructions (using various programming languages) that enable a computer or other machine to perform a desired sequence of operations.

UNIX: a popular computer operating system commonly used for servers.

www: abbreviation for "World Wide Web," part of the Internet. The Web consists of information in pages accessible through a Web browser, whereas the Internet refers to the network of networks where all the information resides. Tools like e-mail, Telnet, FTP, and Internet Relay Chat (IRC) are all part of the Internet, but are not part of the World Wide Web.

22

Pharmacogenomics: Applied Bioinformatics Chapter

Lyle D. Burgoon

Department of Pharmacology and Toxicology, Center for Integrative Toxicology, National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan, U.S.A.

Paul C. Boutros

Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada

Timothy R. Zacharewski

Department of Biochemistry and Molecular Biology, Center for Integrative Toxicology, National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan, U.S.A.

1. INTRODUCTION TO APPLIED BIOINFORMATICS

The statistics associated with drug development are astonishing. Historically, only one in 5000–10,000 chemicals screened successfully reach the market, only 30% of marketed drugs produce sufficient revenue to recover research and development costs, new drug development requires 12 to 15 years and a $500–800 million investment, and 30–50% of drug candidates fail due to toxicity. Each of these factors significantly contributes to the time and cost of drug development. Even incremental increases in the success rate would have favorable economic impacts for all stakeholders including patients, health provider organizations, and drug manufacturers. With ever increasing demands for cheaper drugs with predictable toxicities, pharmaceutical,

and biotechnology companies are exploring alternative paradigms for the development of safer, more cost effective drugs.

One emerging paradigm suggests that to fully assess the potential adverse health effects of chronic and subchronic exposure to synthetic and natural chemicals, and their interactions, a more comprehensive understanding of the molecular, cellular and tissue level effects is required within the context of the whole organism, its genome, proteome, and metabonome. This strategy, referred to as toxicogenomics, incorporates genomics, proteomics, and metabonomics into traditional toxicological practices, to further elucidate mechanisms of toxicity and the etiology of adverse effects elicited by exposure to drugs, chemicals, and natural products as well as their interactions. The application of toxicogenomics into mechanistic research, predictive toxicology, and preclinical safety assessment has resulted in the accumulation of a torrent of data which must be accurately and efficiently indexed and archived to facilitate data analysis and the extraction of decision supportive information important to human health and drug development. Borne from this need, applied bioinformatics is the application of information and statistical sciences to the management and analysis of complex biological data sets to aid in the discovery of new information, the extraction of knowledge, and the development of solutions that enhance human health. This chapter reviews available bioinformatic resources relevant to pharmacogenomic studies elucidating mechanisms of toxicity and preclinical drug development.

The chapter is divided into four complementary sections. Sections 2 to 4 describe the process of sequence analysis, data management and reporting standards, and the design and analysis of microarray experiments. To demonstrate the described computational tools and concepts covered in these sections, a practical example is illustrated in Section 5.

2. SEQUENCE ANALYSIS

Sequence analysis is a core area of bioinformatics research. There are four basic levels of biological structure (Table 1), termed primary, secondary, tertiary, and quaternary structure. Primary structure refers to the representation of a linear, hetero-polymeric macromolecule as a string of monomeric units. For example, the primary structure of DNA is represented as a string of nucleotides (G, C, A, T). Secondary structure refers to the local three-dimensional shape in subsections of macromolecules. For example, the alpha- and beta-sheets in protein structures are examples of secondary structure. Tertiary structure refers to the overall three-dimensional shape of a macromolecule, as in the crystal structure of an entire protein. Finally, quaternary structure represents macromolecule interactions, such as the way different peptide chains dimerize into a single functional protein.

In general, analysis increases in complexity as one attempts to predict higher levels of structure. Primary structure is highly amenable to *in silico*

Table 1 Levels of Biological Structure

Level of structure	Definition	Examples	Tools
Primary (1°)	Heteropolymer sequence	Raw DNA sequence	Sequence-alignment and motif-searching
Secondary (2°)	Local 3D structure (fold)	α-helices, β-sheets	RNA and protein fold prediction
Tertiary (3°)	Overall 3D structure (fold)	Protein crystal structure	Protein structure prediction
Quaternary (4°)	Spatial relationship of 3D structures	Structure of a protein complex	Minimal

analysis while quaternary structure is virtually intractable computationally. In this section the major methods of analysis at each level of structure will be reviewed. However, due to rapidly evolving computational capabilities, discussion will be limited to the underlying motivations for each type of analysis with a brief outline of the major software implementations.

2.1. Primary Structure

Primary structure refers to raw ordering of amino acids in a protein or nucleotides in a DNA or RNA molecule. The techniques for analyzing primary structure are largely derived from standard computer science algorithms for finding specific words within a long document. Primary structure analysis can be broken up into three phases: finding and formatting sequences, measuring the similarity between sequences, and detecting functional ''regions'', or domains, in a sequence.

2.1.1. Sequence Alignment

A sequence alignment is a way of determining the similarity between two strings. This is a classical question in computer science, and has an exact solution referred to as the Smith/Waterman alignment. Unfortunately, this exact solution can be slow when analyzing large sequences, and therefore, approximate methods, such as Basic Local Alignment Search Tool (BLAST), have been developed to identify very similar sequences.

There are two basic types of sequence alignment. Single alignments compare pairs of sequences to find the best match and would be used to locate a DNA sequence on the genome. In contrast, multiple alignments compare sets of sequences with one another to identify overall similarities, and have been used to compare sequence similarity within a superfamily of proteins such as the cytochrome P450 superfamily (Fig. 1).

```
tagctgagagagaaaaaag
||||||||||||||||||
tagctgagagagaaaaaag
```

(A) Single Alignment of DNA Sequences

(B) Multiple Alignment of Protein Sequences

Figure 1 Single versus Multiple alignment. A single alignment (**A**) compares two sequences, showing gaps and matches. A multiple-alignment tries to find the best similarity across a larger range of sequences and can indicate both the exact identify of sequences within a column as well as the conservation of broad chemical similarity (**B**).

In addition, sequence alignments can conduct global alignments which attempt to match two sequences against each other over their entire length when sequences are believed to be related, or locally in order to find small (local) regions of similarity. Although local alignments are more popular, the following discussion is also applicable to global alignments (1). All alignments—single or multiple; local or global—are predicated on some definition of what makes two strings similar, and are mathematically defined by a scoring system.

2.1.1.1. Scoring systems: The scoring system, the core of sequence alignment, assesses similarity by positively scoring matches and penalizing mismatches. In addition, any gaps between two sequences also receive a negative score (Fig. 2). By tuning the scoring system, specific types of matches can be preferentially selected. For instance, some scoring systems are optimized for evolutionarily distant proteins while others are specialized for membrane proteins (2).

The score/penalty assigned to a match/mismatch is dependent on the likelihood that this match/mismatch occurred by chance alone. If an event occurs randomly with a high frequency the score will be small, while very rare events will receive very large scores. For example, a conservative amino-acid substitution will receive a very small penalty, while the introduction of a stop codon will be penalized heavily. A variety of scoring schemes have been generated based on factors such as the bio-physical character, evolutionary distance, and the subcellular localization of the sequences being compared.

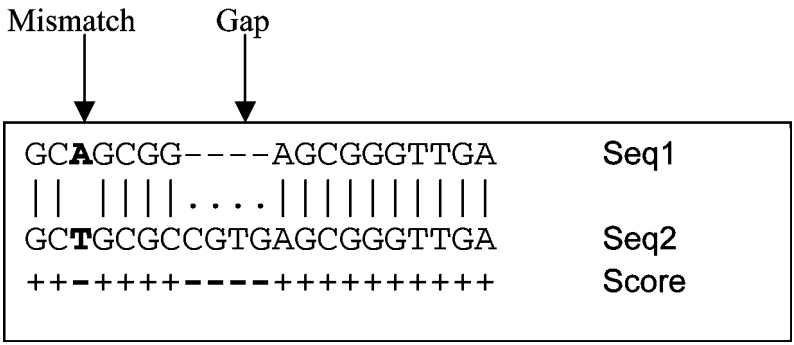


Figure 2 Overview of alignment scoring system. Sequence alignments are scored by attributing a positive score to each match (+ in the score line) and a negative score to either mismatches (the A→T transversion) or gaps.

The presence of gaps in a sequence is a very important topic in sequence alignment, especially in phylogenetic applications. In general a large penalty is assigned for introducing a gap, and a smaller penalty is assigned based on the length of the gap. This gap-penalty often takes the form of a linear equation, and is called an "affine gap penalty" (3).

2.1.1.2. Single alignment: A single alignment compares two sequences to determine the location and the quality of the best match. Single alignments can be used with nucleotides to identify unknown sequences within a database by searching for the most similar or to map novel sequences onto the genome and with peptide sequences in the identification of novel proteins, for phylogenetic comparisons, and to identify orthologous proteins.

2.1.1.2.1. Smith/Waterman. Smith/Waterman is a slow, computationally intensive method that finds the optimal alignment between a pair of sequences. It is based on the classic dynamic programming algorithm with run times proportional to the length of the sequence squared (n^2). For large databases or long sequences dynamic programming can become impractically slow, even on powerful systems (4).

Nevertheless Smith/Waterman alignment still has significant utility in practical applications when the need for sensitivity is paramount and the dataset to be analyzed is relatively small. Many evolutionary studies focusing on individual proteins use Smith/Waterman alignment. Several public servers providing access to Smith/Waterman alignment exist (Table 2).

2.1.1.2.2. BLAST. To overcome the computational limitations of Smith/Waterman alignments, statistical approximations were developed by computational biologists in the early 1990s. These approximations do not *guarantee* that the optimal alignment will be found, but are overwhelmingly likely to identify the best alignments. More importantly, programs

Table 2 Smith–Waterman Web Resources

Resource name	Notes	URL
USC SAS	Very flexible	www-hto.usc.edu/software/seqaln/
Blitz	E-mail-based server	www.ebi.ac.uk/searches/blitz_input.html
Hybrid sequence aligner	Uses a Smith-Waterman variant with more tractable exact statistics	bioinfo.ucsd.edu/hybridparameters/
GeneStream	E-mail based server	www2.igh.cnrs.fr/bin/ssearch-guess.cgi

implementing these approximations generally run a full order of magnitude faster than full Smith/Waterman alignment (5).

The predominant statistical algorithm for sequence alignment is BLAST. The BLAST algorithm does not perform an exhaustive search of two sequences for similarity, but looks for short, exact matches called seeds or words, which are then extended on either side to find the overall best alignment. Short seeds will often be found by random chance, but longer seeds can miss strong-but-inexact similarities. Therefore, choosing a seed-length is an important factor, as it involves a trade-off between sensitivity and search speed. The major versions of BLAST are listed in Table 3, along with a brief description of their primary uses (6).

2.1.1.2.3. Other alignment software. While Smith/Waterman and BLAST alignment are the most popular methods of performing single-alignments a host of other software has been developed for specialized purposes that attempt to improve execution speed by making simplifying assumptions about the types of matches that can be found (Table 4).

2.1.1.3. Multiple alignment: Multiple sequence alignment involves the simultaneous comparison of several distinct sequences. If these sequences represent a single protein across multiple species, the alignment will indicate evolutionary relatedness. Alternatively, alignment of proteins within a single

Table 3 Versions of BLAST

Program name	Sequence 1	Sequence 2
BLASTN	Nucleotide	Nucleotide
BLASTP	Protein	Protein
BLASTX	Nucleotide	Protein
TBLASTN	Translated nucleotide	Protein
TBLASTX	Translated nucleotide	Translated nucleotide

Table 4 Other Major Sequence Alignment Packages

Package name	Primary use
FASTA	BLAST-like local alignments
REPuter	Maximal repeat identification
MUMmer	Suffice-tree algorithm
SIM	Linear space requirements
SENSEI	Very fast ungapped alignments
QUASAR	Suffix-array algorithm
BALSA	Bayesian-based alignment
PatternHunter	Extremely rapid large-scale alignments

species may represent the series of gene-duplication events that led to the protein family. Multiple alignments are commonly used as the raw data input for the generation of phylogenetic trees (7). In contrast to single alignments, two basic classes of multiple alignment algorithms, progressive and iterative algorithms have been developed (Table 5). Their performance varies with the characteristics of the individual alignments.

Progressive algorithms follow a three-step path. First, pair-wise alignments are performed to rank sequences in order of similarity. The two closest sequences are then aligned together, and finally, the remaining sequences are added to the alignment one-by-one, in descending order of similarity. Therefore, this approach progresses by adding increasingly distant alignments to the analysis, and is the traditional approach to multiple sequence alignment. Common progressive multiple alignment software packages include CLUSTALX and GCG. Despite its age, CLUSTALX (and its graphical version, CLUSTALW) is the most common multiple alignment package.

In contrast, iterative algorithms use the entire set of alignments at once, but repeatedly iterate over the set until a stable alignment has been generated. Iterative approaches to multiple alignment are both newer and

Table 5 Selected Online Multiple Alignment Resources

Resource name	Resource type	URL
ClustalW	Alignment server	www.ebi.ac.uk/clustalw/
BCM search launcher	Alignment server	searchlauncher.bcm.tmc.edu/multi-align/
MultAlin	Clustering server	prodes.toulouse.inra.fr/multalin/multalin.html
Multiple alignment guide	Tutorial	www.techfak.uni-bielefeld.de/bcd/Curric

less popular than traditional progressive methods. Some common iterative multiple alignment software packages are DIALIGN and PRRP (8).

2.1.2. Motif Detection

Motif detection involves searching for one sequence within another and is based on the assumption that one sequence is a part of another. Protein motifs are generally short sequences that represent conserved domains (CDs) or characterize protein families. Nucleotide motifs can range in size from full genes within the genome to short enhancer elements.

2.1.2.1. Protein motifs: identifying protein function: Protein motifs are fragments ranging from a few to several hundred residues. Short fragments encode functions that evolved independently of their surrounding context, like glycosylation or phosphorylation sites. Moderate-sized fragments encode specialized functions, such as localization signals, while longer fragments constitute CDs with broader functionality. The CDs are usually thought to have occurred once in evolution, and were then transferred or swapped throughout multiple genomes. The total number of protein motifs is staggeringly large, with well over 15,000 motifs known, and new ones continuously being discovered (9).

Short sequences can be found by thoroughly scanning protein sequences for all possible occurrences. For example, localization signals and other targeting sequences are usually detected in this manner. Longer regions, like DNA-binding regions, enzymatic active sites, and protein–protein interaction domains are typically identified through sequence alignments, much as described above (10).

A variety of databases and online tools exist to facilitate searches for protein motifs (Table 6). The most comprehensive resource for the detection of large protein motifs is the Conserved Domain Database (CDD) provided by NCBI. The CDD includes all data present in the SMART and PFAM databases, along with some manually curated entries. All protein–protein BLAST

Table 6 Protein Motif Resources

Resource name	Tools provided	URL
SMART	Database and search tools	smart.embl-heidelberg.de
PFAM	Database and search tools	www.sanger.ac.uk/Pfam
CDD	Database and search tools	www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
SignalP	Signal peptides only	www.cbs.dtu.dk/services/SignalP
Big-PI	Sugar-anchors only	mendel.imp.univie.ac.at

searches performed through NCBIs web-service automatically perform a CDD search using Reverse Position Specific BLAST (RPS-BLAST) (11).

2.1.2.2. Nucleotide motifs: annotating the genome: Nucleotide motif finding occurs at three distinct levels: identification of coding sequences, identification of promoters, and identification of enhancers. At each level the core question is how to identify a signal from the underlying noise of genomic sequence (Fig. 3).

2.1.2.2.1. Finding genes. Gene finding is one of the most important genomic applications. Originally it was believed that the human genome sequence could be fully annotated using automated gene-prediction software. Instead, the identification of open-reading frames (ORFs) has turned out to be a challenging problem.

Even in simpler organisms, like yeast, ORF annotation has exhibited high error-rates. High-throughput proteomic and genomic studies combined with *in silico* approaches have corrected many mis-annotated ORFs in the yeast genome (12). With the greater complexity of the mammalian genome, it is not surprising that ORF prediction has had limited success.

Empirically, ORFs have been identified through the high throughput single pass sequencing of individual clones within a cDNA library to rapidly and inexpensively generate Expressed Sequence Tags (ESTs). The ESTs can

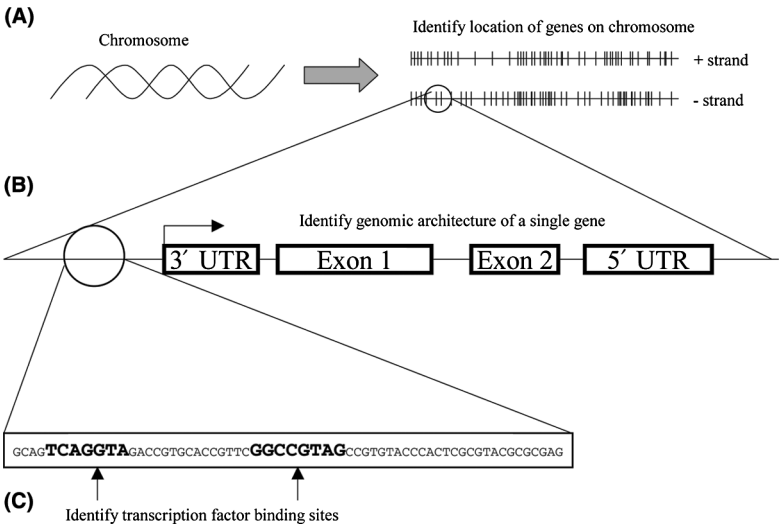


Figure 3 Relationships between the three types of nucleotide motifs. Nucleotide motifs can be considered at three levels. At the genomic level (**A**) is the gene-finding problem: where are the genes located along a chromosome? Once a gene is located, the gene-architecture problem (**B**) attempts to determine the number of exons and the overall structure of the transcribed sequence. Finally, transcription-factor binding-sites (**C**) within or near a gene regulate the gene's spatial and temporal pattern of expression.

then be mapped to the genome to identify representative genes. This approach has also been used to detect splice-variants. Unfortunately it is biased towards highly expressed genes, and requires that a particular mRNA be expressed in a given experimental sample. Further, the iterative computational assembly of multiple ESTs into "clusters" that represent putative gene-products is a challenging problem, which is error-prone (13).

Coding sequences have characteristic features such as codon frequencies, start- and stop-codons, upstream TATA-boxes, and splice-sites, which can be used to annotate an ORF (14). While each piece of evidence is not, by itself, conclusive, the combination of factors can lead to firm identification of genes. Unfortunately this method is prone to a large number of false-positives, with at least 75% of predicted genes typically being artefacts (15).

The most reliable method of identifying ORFs is by identifying genes that share high sequence homology across species, which are also known as orthologs. Conserved genomic structure with comparable function in a closely related species can be used as a strong guideline for identifying ORFs. However, this method is inherently limited to genes with high homology in closely related species. For example, annotating rat genes through comparison with their mouse orthologues will miss genes that are unique to that species. Nevertheless, this approach does not suffer from the excess of false-positives associated with gene-structure-based algorithms, nor the very high computational burden associated with EST-based predictions (16).

2.1.2.2.2. Finding promoters. The promoter is a region proximal to the transcriptional start site (TSS) that contains *cis*-acting motifs for transcription factors that regulate gene expression. Finding the promoter of a gene is usually significantly more difficult than finding the coding sequence for a gene since there are no absolutely conserved elements that define the promoter. In contrast, identifying coding sequence requires identification of conserved elements such as the start codon, splice-sites, and the stop codon. Two basic methods for identifying promoter sequences have been developed: signal-based approaches and content-based approaches: signal-based and content-based approaches.

Signal-based approaches involve searching for elements like the TATA and CCAAT boxes used by the basal transcription machinery. Often these elements occur with conserved spacing patterns which can be used by techniques like support vector machines and genetic algorithms to identify putative promoter regions (17). Content-based approaches leverage the underlying structure of DNA to identify promoters, rather than the presence of conserved elements. In contrast, content-based algorithms will look for features such as base-pair triplet frequencies, the presence of CpG islands, or a randomness characteristic to non-coding sequence (14).

Alternatively, promoters could be identified by mapping mRNA sequences directly against the genome. If the mRNA transcript sequence

is complete, the region immediately 5′ to the genomic location will be the promoter location. This approach, often used in combination with signal- or content-based algorithms has proven to be up to 80% accurate and 50% sensitive (18).

However, the identification of alternative promoters and splice-variants continues to be a challenge. A recent study showed that at least three-quarters of multi-exon human genes exhibit alternative-splicing, usually in tissue- and cell-type- specific patterns (19). No computational approaches have been developed to rapidly predict and screen potential alternative promoters for a single gene.

2.1.2.2.3. Finding enhancers. In general, the proximal promoter of a gene contains the binding sites for components of the basal transcription machinery. Upstream of this region is the distal promoter, which contains binding sites for a variety of *cis*-regulatory elements. These *cis*-acting elements act as binding sites for transcription factors. Transcription factors regulate aspects of gene-expression such as tissue-distribution, response to signalling pathways, and temporal specificity. This regulatory function is achieved in several ways, including direct recruitment of the basal transcription machinery, recruitment of complexes that modify the histones, and recruitment of co-activator proteins that specialize in enhancing transcriptional initiation (20).

Enhancer sequences are short 5 to 25 bp motifs may randomly occur many times in the genome. For example, a 10 bp motif would be expected to occur approximately once every million (4^{10}) bases by chance alone. Accordingly a major challenge in searching for enhancer elements is determining the functionality of an element. At present no current *in silico* search algorithm can definitely identify a functional *cis*-regulatory element. Statistical analysis can indicate whether a given promoter is likely to be active, but predictions must be empirically confirmed. More effective *in silico* prediction of enhancer elements may utilize comparative genomic approaches and will require a better understanding protein and DNA interactions.

The consensus binding sequence for a known transcription factor is usually encoded in a weight-matrix. This matrix indicates what base each position of the binding-site is likely to contain, and is generated from transcription factor binding specificity data. The most comprehensive listing of transcription factor binding sites and their associated weight-matrices is in the TransFac database. A less extensive free database, JASPAR, is also available (21).

Alternately, an analysis of co-regulated genes can identify conserved motifs for novel transcription factor binding sites. This involves identification of a set of co-regulated genes that have similar expression profiles in response to some stimuli. The distal promoters of these genes are then analyzed for sequences that occur more often than would be expected by chance alone. These over-represented sequences are putative binding sites, which can be compared to the consensus sequences of known transcription

factors and empirically tested using DNA foot-printing, reporter gene and band shift assays (17).

Phylogenetic foot-printing is a technique that is gaining popularity in enhancer searches. The underlying premise is that regulatory sequences will be conserved across species. In other words, it is assumed that non-coding regulatory sequence is evolutionarily conserved, and that regions of non-coding DNA that are highly homologous across species are likely to contain important regulatory motifs. A study in skeletal muscle showed that 19% of the best conserved regulatory DNA contained 74 of the 75 known transcription factor binding sites (22). Accordingly, a transcription factor binding site that is observed in a conserved position in multiple species is much more likely to be functional than one that is only present in a single species. Both the weight-matrix scans and sequence-representation searches described above can be supplemented by phylogenetic foot-printing.

2.2. Higher-Order Structure

While the primary structure of proteins and nucleic acids can be experimentally determined in a straight-forward manner, their higher-order structures are much more difficult to elucidate. In general, computational methods dealing with primary structure focus on interpretation of the structure–function, as in promoter analysis. By contrast computational methods working on higher-order structure instead focus on the prediction of structural details. Further, most techniques are limited to the prediction of RNA and protein structures—sugar-, fatty-acid-, and DNA-structural prediction methods are in their infancy.

We consider higher-order structure for both RNA and proteins. Surprisingly, the folding problem is quite different for each of these two heteropolymers at all levels of structure, with RNA folding being more tractable, both theoretically and computationally.

2.2.1. RNA Folding

The RNA secondary structure can be defined as a state where all bases are mapped into one of two states: pairing or non-pairing (23). The RNA base-pairs include both Watson–Crick associations and weaker "wobble" pairs like G:U associations (24). These pairs of associated bases are usually adjacent with and anti-parallel to other associated pairs, creating regions with complex three-dimensional structure. These well-folded regions are islands, separated by stretches of single-stranded, unstructured RNA. There are four basic types of RNA secondary structure—helices, much like those found in DNA; loops, like the hair-pins found in tRNAs; bulges and junctions. Of these four basic forms, only helices are made up primarily of Watson–Crick base-pairs. The latter three structural types are largely comprised of aberrant base-pairs, terminated by helical regions (25).

While there are several different approaches to predicting RNA structure, it should be emphasized that RNA structure prediction remains a

challenging, but tractable task. All the algorithms and software packages discussed here are capable of predicting RNA structure with 70% or greater accuracy. The four major approaches to RNA structure prediction are: thermodynamical, naïve, homology-based, and empirical.

Thermodynamical approaches assume that a folded RNA molecule will adopt its most stable conformation. These algorithms model the free energy of the molecule as the sum of the individual base-pairing and base-stacking interactions (24). The best-known implementation of this approach is the mfold web-server, which has been available for nearly a decade. The web-server provides access to a wide-range of tools, including RNA- and DNA-folding, melting-point determination, and structure-viewers. This server is being continually maintained and upgraded with new tools and enhanced implementations (26).

Naïve approaches avoid theoretical assumptions and instead focus on statistics about solved RNA structures, using these to probabilistically align new sequences with solved structures. One elegant approach to this problem has used an rRNA database to generate a novel RNA-specific substitution matrix. The advantage of this approach is that it makes the whole spectrum of primary-structure sequence-analysis tools available for secondary-structure prediction (27).

Homology-based approaches to RNA-folding are conceptually similar to the phylogenetic foot-printing methods used for transcription-factor binding-site analysis detailed above. The assumption underlying both methods is that evolutionarily conserved sequences will be functionally relevant. Phylogenetic foot-printing assumes that such conserved sequences contain transcription-factor binding-sites. Homology-based RNA-folding methods assume that these conserved sequences contain regions of secondary structure, rather than unfolded regions of primary structure. Unfortunately this is a stronger assumption for folding methods than for motif-detection methods because it is now postulating linkages between both primary structure and secondary structure, and between secondary structure and function. Despite the strength of this assumption, homology-based methods have proved useful, and some methods have even been developed that achieve $>80\%$ accuracy from aligned sequences (28).

The final method of RNA structure prediction, empirical algorithms, are also analogous to primary-structure motif detection methods. Known RNA structural motifs are extracted from structural databases, and the primary-structure patterns underlying these motifs are identified. Novel RNA sequences are then scanned for these primary-structure motifs much like a novel protein sequence might be scanned for CDs. In essence, these methods search the primary structure of sequences for conserved motifs that indicate secondary structure. One of the most flexible and powerful empirical tools is RNAMotif, which is freely available for download, but does not have an associated web-server (23).

2.2.2. Protein Folding

The process through which a linear string of amino acid residues newly synthesized at a ribosome folds into a complex, three-dimensional, biologically active protein structure remains poorly understood. Consider how protein-folding contrasts with RNA-folding. Proteins have 20 distinct monomeric units, RNA only four. The amino acids include aromatic, hydrophobic, cationic, and anionic chemical properties compared to four comparable RNA nucleosides. Moreover, secondary and tertiary structures were fundamentally inter-linked in proteins, but are essentially distinct in RNA molecules.

An enormous body of research into protein-folding has led to the development of several computational methods for predicting and understanding the lowest level of three-dimensional protein structure. While high-resolution (1 Å) crystal structures are the gold standard for elucidating three-dimensional protein-structures, this level of accuracy is not necessary for many applications. For example, virtual screening of ligands through docking simulation can be done with lower resolutions. Site-directed mutagenesis studies can be motivated by structures with resolutions of only 4 Å and three-dimensional motif-scanning can be performed at even lower resolutions (29). In short, low-resolution predictions of structure still have utility.

2.2.3. Computational Approaches

A major problem in predicting protein structure is the computational intractability. A short, 100-residue protein will contain at least 100 side-chain-to-side-chain or side-chain-to-solvent interactions. The orientation of each of these interactions will lead to cascading effects throughout the protein. Comparative modeling, threading algorithms, and de novo predictions seek to predict protein structure in reasonable execution times.

Comparative methods leverage primary-structure similarity between the sequence of interest and a protein with known structure. Sequence homology is then assumed to result in structural homology, and is used to develop an all-atom structural model (30). Of the four basic steps in homology modeling (Fig. 4), it has been shown that template-target alignment is the most critical for development of a reliable model. When the target and template have sequence identity of at least 50%, the leveraging of a homology relationship becomes very powerful, and comparative methods are the most accurate protein structure prediction tool. When target-template identity is low, heuristics and user-input are required to generate a model of any reliability. Threading algorithms, like comparison methods, leverage known structures to help identify the correct fold for a new protein sequence. Rather than employing sequence-level homology as a screening tool, threading methods attempt to fit the new sequence to every known crystal structure. These algorithms essentially ''thread'' the new sequence through each of the known structures. The energetic stability of each conformation is calculated, and taken as a

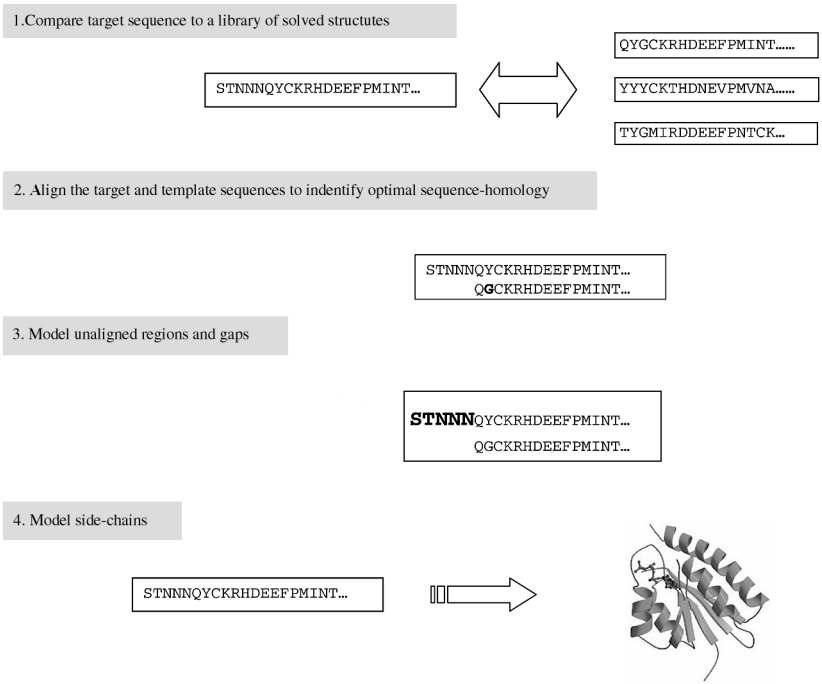


Figure 4 Steps in homology modeling of protein structure. Modeling the structure of a protein by considering its homology to known homologs is one of the most effective methods of structure-prediction. The critical steps are the identification and alignment of the target sequence to the template sequence (1) and (2).

measure of how well the known structure matches the new sequence (29). While threading methods are successful as comparative methods, they are very computationally intensive, and are inherently limited to previously characterized folds. As novel protein folds are continually discovered, this limitation is decreasing.

De novo algorithms predict three-dimensional structure directly from the primary sequence, without any leveraging of known structures. While predictions are often of poor accuracy (low-resolution), it remains the only way of determining structures for families with no known members. The sole assumption is that the native conformation is the most energetically favorable. This may not be true for proteins that undergo chaperone-mediated folding, but is a reasonable assumption (29). Most de novo algorithms walk along a sequence in words varying in length between 5 and 27 residues. For each word, the most probable set of secondary structures is predicted. Finally, the protein is allowed to vary over the possible combinations of secondary structures, with the lowest energy structure being selected as the prediction. One major implementation of this algorithm is

the ROSETTA software, for which a web-server is available (www.bioinfo.rpi.edu/~bystrc/hmmstr/server.php) (30).

Protein-structure prediction methods are routinely compared and contrasted in a public competition called the Critical Assessment of Techniques for Protein Structure Prediction (CASP). These large-scale events allow for a comparison of the state-of-the-art tools and algorithms on a variety of target sequences. The results of each event are scrutinized. There have been five such events thus far, with CASP5 being the most recent (31).

3. DATA MANAGEMENT

3.1. Introduction to Bioinformatics Data Management

Data management refers to the storage, sorting, and retrieval of data. While not traditionally an important part of bioinformatics, the explosive growth in high-throughput technologies has generated large complex data sets requiring a need to archive, process, integrate, and search data for biological meaning. Consequently, data management is becoming an increasingly important part of applied bioinformatics.

3.1.1. Why Is Data Management Needed?

The sheer magnitude of modern biological datasets is daunting. Consider the publicly accessible GenBank sequence database currently comprises 591 distinct files, with a cumulative size of dozens of gigabytes. Even the largest available hard-drives available for a personal computer can barely store this entire database (Table 7). In comparison, a standard two-color microarray experiment will generate two image files (≈30 MB each), one quantification file (≈5 MB), a normalization file (≈5 MB each), and one final gene-list (≈1 MB). A single array can generate up to 70 MB of data, and even a simple experiment can easily require 20 arrays resulting in over 1 GB for this simple experiment. This data must be stored securely so that it remains accessible after publication, and for future researchers looking to explore the data in greater detail, possibly with newly developed methodologies.

Table 7 Relative Size of Standard Bioinformatics Resources

Resource	Size	Best comparison
One Microarray Slide	~40 MB	One full length music album
One Human Chromosome	~250 MB	One full installation of Microsoft Office
All SNPs on a Chromosome	~2 GB	One full length movie
All ESTs on a Chromosome	~50 GB	Every complete Simpson's episode

Fortunately, microarray experiments generate highly ordered data, and therefore the challenge is not simply the data storage but the intrinsic relationships within these highly structured datasets, which must be preserved and enforced. Each microarray in an experiment has very specific samples hybridized to it, and each sample must be associated with a microarray. Each feature on the microarray contains a specific sequence, and has associated with it a variety of intensities, quality metrics, and ultimately expression measures. These different types of data must be integrated and co-ordinated, so that a researcher can look at a single gene, and see how that gene has behaved across a set of microarray experiments, and what annotation has been ascribed to that gene. The user-friendly integration and storage of large datasets is a long-standing challenge for bioinformatic researchers.

3.1.2. A Role for Databases

Storing and organizing large datasets is typically performed using relational databases. Tuned to optimize storage efficiency and retrieval speed, relational databases were first used by government departments and large corporations. As their broader applicability became apparent, relational databases were deployed across a broad spectrum of industries and groups, and standard office software packages now come with a rudimentary relational database (e.g., Paradox or MS-Access) included.

The term ''relational'' is derived from the way the software leverages known relationships between different types of data to optimize storage space, to speed data access, and to enforce rules that maintain the underlying coherence or integrity of the data. A typical relationship may be expressed as: every polymorphism in a coding region must be either synonymous (silent) or non-synonymous, or every microarray slide must contain one or more spots. The underlying design or schema of a relational database is the specification of a large number of data associations of this sort. A sample schema diagram is shown in Figure 5, while a listing of common database software packages is given in Table 8.

3.2. Genomic Data Standards

A great deal of interest has been generated by many for sharing genomic data (32–36). However, without data standards in place to govern protocols for sharing data, exchanging genomic data will be fraught with difficulties. Recently, the Microarray Gene Expression Data Society (www.mged.org) proposed the Minimum Information About a Microarray Experiment (MIAME) standard to provide guidance in determining the minimum amount of information required to independently replicate an experiment or reanalysis a deposited data set (32). Others have since began exploring other key issues in standards development, including the development of

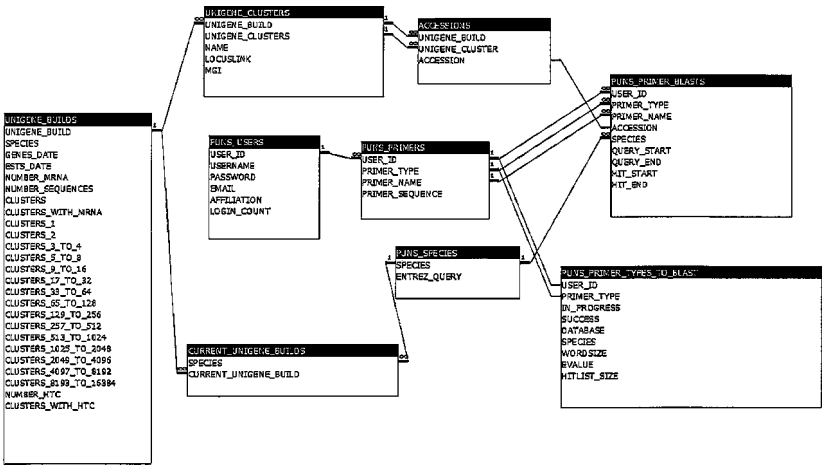


Figure 5 Sample database schema. A database schema shows the types and nature of links between different types of data. Each box represents a table within the database. The rows within that box correspond to the fields of the table. The lines connecting the boxes identify the required relationships amongst the different types of data stored in the different tables.

controlled vocabularies and ontologies, to harden terms and definitions, thereby creating a standard language for describing genomic experiments.

This section will discuss the standards for sharing and discussing microarray experiments, including the ontologies and controlled vocabularies. A non-technical description of the sharing process will be provided, as well as an introduction to the pharmacology/toxicology focused version of the MIAME standard, MIAME/Tox, which is currently under development.

3.2.1. Ontologies and Controlled Vocabularies for Genomic Data

In order for investigators to effectively share their data, they must ensure that the language they use for describing the experiment can be understood by others. The first step in sharing data is the creation of a controlled vocabulary or an ontology. Once the community is speaking the same language, they can begin sharing data.

A microarray generally consists of a glass slide with some material affixed to it, and some other labeled material hybridized to the array. When investigators use the term "probe," to which material they are referring? The term probe has been used to describe both materials in the past, and may be a source of confusion. The solution to this kind of terminology problem is for the community to agree on a common language, called a controlled vocabulary or an ontology. A controlled vocabulary is a set of

Table 8 Overview of Common Database Software

Database name	Vendor	Notes
MS-Access	Microsoft	Light-weight database bundled with Microsoft office, not generally suitable for bioinformatics applications
Paradox	Corel	Light-weight database bundled with Corel WordPerfect, not generally suitable for bioinformatics applications
PostGreSql	Open source (free)	Powerful free database with many features, but not fully available on Windows platform
MySQL	Open source (free)	Light-weight free database, popularly used as the back-end for bioinformatics web-sites
Oracle	Oracle	Extremely powerful, state-of-the-art commercial database software; highly customizable, but expensive
DB2	IBM	Extremely powerful database software, with several bioinformatics-specific features available
Sybase	Sybase	A moderately priced but powerful commercial database

descriptors that a community agrees to use when describing an experiment. Definitions are not explicitly given with the vocabulary as they are expected to be understood. One example of a controlled vocabulary would be the National Toxicology Program's Pathology Code Tables for describing pathology lesions. The terms used in the vocabulary should be definite and distinct, such that any lesion should only be identified by one term in the vocabulary list. If controlled vocabulary terms are associated with a specific definition, then an ontology is created. Generally, a true ontology should be a complete specification of a concept, in this case a microarray experiment. Currently, the MGED Ontologies Working Group is assembling a microarray experiment ontology (http://mged.sourceforge. net/ontologies/index.php).

3.2.2. MAGE-OM and MAGE-ML

The emerging method for sharing genomic data between databases is the Microarray Gene Expression Object Model (MAGE-OM) and MAGE Markup Language (MAGE-ML). The MAGE-ML is derived directly from the MAGE-OM, a data representation standard approved by the Object Management Group (OMG; http://www.omg.org). The MAGE-ML is built in the eXtensible Markup Language (XML), and is the preferred data shar-

ing method. As part of the MAGE-ML development, documents are generated that are used by computers to decode the MAGE-ML documents. The MAGE compliant software and databases use these decoding documents to properly interpret the MAGE-ML encoded data, and to populate the appropriate database tables or data structures within the application (33).

The MAGE has built within it a controlled vocabulary that is used to standardize communication between data providers. However, MAGE can also be extended to encode other types of "omic" data beyond genomics, such as proteomic data, so long as a reference to the ontology or controlled vocabulary is provided. Although a description of the extension mechanism is beyond scope of this book, practitioners must become familiar with it in order to ensure their software products and applications will be able to accept any and all annotation data that may be submitted with the genomic data.

3.2.3. The Minimum Information About a Microarray Experiment Standard

The MIAME standard defines the minimum information investigators must report for a microarray experiment to be reproduced. The MAGE standard was born partially from MIAME, and the European Bioinformatics Institute used MIAME and MAGE to guide the development of ArrayExpress, their public genomic data repository (34). Sample annotation lies at the heart of MIAME, underscoring the need to understand as completely as possible the experimental conditions that may influence the microarray data. Many journals that publish microarray data require the submission of MIAME-supportive microarray data to a public genomic data repository as a condition of publication. These typically include submission of protocols; species, strains, and sex used for in vivo studies; cell line name and culture conditions for in vitro studies, and other relevant information.

To assist investigators in meeting the requirements of the standard, the MGED Society provides a MIAME-checklist. The checklist outlines the most pertinent points of the standard to facilitate data submission compliance (www.mged.org/Workgroups/MIAME/miame.html). The MIAME standard includes a glossary of the controlled vocabulary used throughout the MIAME document to assist data for submission.

3.2.4. MIAME/Tox—A MIAME Specific for Toxicology and Pharmacology

Although the MIAME standard will suffice for many experimental paradigms, it is not complete enough for the purposes of the regulatory toxicology and pharmacology communities. The MIAME standard does not include reporting requirements for toxicological or pharmacological endpoints or measures to place the genomic data within proper context.

The MIAME/Tox standard is being developed to address these concerns. Additional reporting requirements of the MIAME/Tox extension

include (1) clinical pathology, (2) chemical mixtures and constituents, (3) chemical exposure protocol, (4) husbandry details, and (5) histopathology data. It is unclear at this time what the actual reporting burden will be, and whether or not differences will exist between academic research and industrial regulatory submissions as the standard is still under development by MGED Society's Toxicology Working Group.

4. MICROARRAY DATA ANALYSIS

4.1. Microarray Data Analysis

Analysis of microarray experiments spans many subjects including study design, quality control, normalization, data filtering, and result interpretation as depicted in Figure 6.

4.2. Microarray Experimental Design

Prior to any experiment, study design considerations must be considered in order to ensure the study addresses the question of interest and has sufficient technical and biological replication to support data analysis using appropriate statistical methods. A technical replicate is a sample from the same source that is hybridized to more than one array and is expected to have less variance. Biological replicates are samples from independent experimental units (e.g., animal). For example, total RNA from the right lateral hepatic lobe excised from individual Sprague–Dawley rats receiving the same treatment would constitute biological replicates. Whereas a technical replicate allows for better estimation of the precision for gene expression assay, biological replicates provide an estimate of assay variation and biological variation (35,36). In general, biological replicates are more informative than technical replicates, however, technical are still important to assure competence and consistency with the assay. Other factors that will influence study design and the degree of replication include the type of study (e.g., time-course, dose-response, and class discovery), available resources (e.g., number of slides and expense), and the amount of sample.

For two-channel microarrays, conventional wisdom suggests that comparisons of greatest interest should be paired on the same microarray. Table 9 shows examples of the different microarray designs, and when these scenarios may be appropriate.

4.3. Microarray Data Normalization and Transformation

Normalization attempts to remove technical variation in the data that is not attributed to biological or treatment related variation. Examples of technical variation include differences in dye incorporation, physical differences in fluorescence efficiency of incorporated dyes, print-surface irregularities, and print-

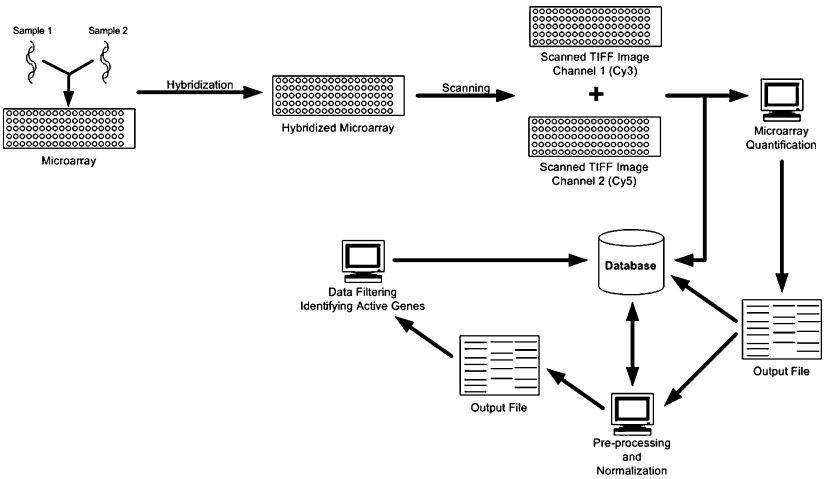
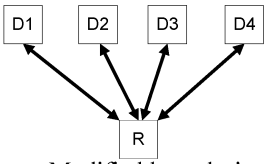
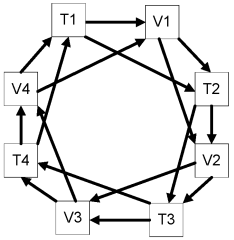
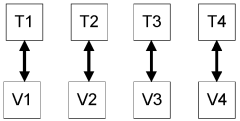


Figure 6 Microarray data flow. Microarray data exist on several different levels depending upon the analysis stage. Two-channel microarrays typically consist of two reverse transcribed labeled samples hybridized to the same microarray. Following scanning two grayscale TIFF images are generated, one for each channel, with intensities ranging from 0 to 65535 units (0 = black, 65,535 = white). Some scanners combine the two TIFF images into one TIFF file, while others generate two separate files, one for each image. These images are typically stored in a database and are the rawest form of the data. Image quantification software is then used to quantify feature intensities and background intensities. These are placed in an output file, the contents of which are generally stored in a database. This data is processed and normalized prior to further analysis. Statistical analysis then assist with the filtering process, to decrease the size of the data to a manageable size, and to control the number of false positives within the data. The output from all of these steps should be saved in the database to facilitate future re-analysis once improved methods are introduced, and to facilitate data sharing between laboratories, to repositories, or regulatory bodies.

tip effects, although there are several methods, most investigators use global or local normalization methods (Table 10) (37,38). Global methods normalize across the entire dataset whereas local methods perform the normalization across a subset of the data. Local methods have the advantage of correcting for spatial biases (38).

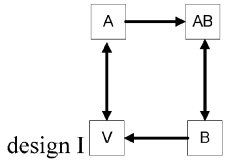
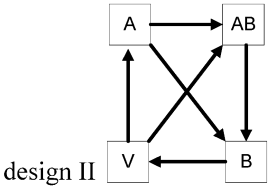
Microarray data may require transformation to stabilize the variance by applying a global logarithmic transformation prior to normalization and further analysis. This may unpredictably skew the measurement, making further analysis difficult and unreliable (39). As many of the techniques used for the analysis of microarray data are not robust to the equal variance assumption, a family of variance stabilizing transformation (VST) methods for microarray data have been proposed (40–44).

Table 9 Common Microarray Designs

Design	Typical uses	Notes
Reference design 	(1) Dose-response (depicted)	For dose-response studies the Reference can be either pooled vehicle samples or a pool of all samples. In the latter case, one of the D*n* samples would represent a vehicle sample
	(2) Population comparisons (e.g., tumors)	For a population comparisons the Reference should be a pool of all samples
Modified loop design 	(1) Time-course (depicted)	The depicted time-course experiment would be appropriate if identifying differentially expressed genes at any time-point compared to vehicle and temporal expression changes within treatment (including vehicle) compared to an adjacent time-point
	(2) Comparison of different populations (e.g., tumors)	Population comparisons can be carried out between parent tumor samples (T*n*) and metastases (V*n*), as well as between all parent and metastases
Independent reference design 	Time course	Identifies changes in expression as a result of treatment compared to time-matched vehicle controls. Does not allow for efficient comparison within treatment class

(*Continued*)

Table 9 Common Microarray Designs (*Continued*)

Design	Typical uses	Notes
2 × 2 Factorial  design I	Combination or coadministration chemical Exposures	Identifies changes in expression as a result of Drug A, Drug B, or the combination of Drug A and B (AB) compared to vehicle
2 × 2 Factorial  design II	Combination or coadministration chemical exposures	Identifies changes in expression as a result of Drug A, Drug B, or the combination of Drug A and B (AB) compared to vehicle (36) show this design yields more variance for the interaction term than Design I

D*n*, *n*-level dose; R, Reference sample; T*n*, Chemical treated sample from time *n*; V*n*, Vehicle treated sample from time *n*; A, Chemical A; B, Chemical B; AB, Coadministration of chemicals A and B. An arrow represents a microarray, with the head representing one dye (e.g., Cy5), and the tail representing a second dye (e.g., Cy3). Double headed arrows represent a dye-swap where sample is labeled with both dyes, and hybridized to two different arrays.

Table 10 Normalization Methods Commonly Used in Microarray Studies

Normalization method	Category	References
z-score	Global	(59)
Standard loess normalization	Global/local	(38,45,46) (62)
Faster cyclic loess (fastlo)	Global/local	(58)
General Linear Models (GLM)	Global/local	(60,61)
General Linear Mixed Models (GLMM)	Global/local	(63)

4.3.1. Normalization Factor Calculation Methods

One common microarray data normalization method is to calculate a normalization factor on a per array basis or across an entire experiment. The primary assumption for using a singular normalization factor is that the volume of labeled sample is comparable across the two channels. Thus, due to the large population of labeled cDNA within the uniform volume it is assumed that the same number of labeled cDNAs exist in both samples. Ideally, the overall intensity in the two channels will be the same. Furthermore, any increases in labeled cDNAs, due to increases in mRNA, must result in decreases of some other labeled cDNAs. Typical methods include mean- or median-centering, where the mean/median values are centered within the data distribution, and *z*-score normalization which adds a scaling factor to mean-centering.

The drawback of these approaches is their failure to compartmentalize sources of variation. Thus, in studies where biological and technical replicates both exist, there is the possibility that the normalization method will inappropriately remove biological variance, such as the treatment effect itself. This can lead to a reduction in the overall treatment related response. Other sources of variation include replicate effects, where the timing of hybridization and scanning, may cause persistent trends in the data.

4.3.2. Lowess Normalization

Lowess normalization methods are based on lowess (loess) scatterplot smoothing algorithms. The lowess smoother attempts to smooth contours within a dataset. Typically the lowess will be robust to genes which are active in treatment as they will be observed as outliers (45). Some normalization methods include a print-tip normalization (46), since physical location on the array and the print-tip may contribute some effect and variance beyond the biological and treatment variation.

The lowess normalization method is typically coupled with scatterplots (e.g., M vs. A) illustrating the relationship between the two dye channels, as a function of the overall intensity (i.e., geometric mean of the intensities) (47). The lowess normalization smooths the microarray data in

the scatter plot to correct for data artefacts. The normalization is performed on a per-array basis, and does not account for temporal effects, such as the date the assay was performed.

4.4. Microarray Data Ranking and Prioritization

Normalized data may be compared across the experimental variables of interest, such as dose, time, or chemical exposure. However, analysts are interested in grouping features into changed or unchanged groupings compared to some metric or control. Numerous methods abound for identifying significant changes in gene expression including (1) arbitrary fold-change cut-offs, (2) model-based *t*-test, and (3) ANOVA with post hoc test (Table 11). The fold-change method, which historically was the primary method for filtering, has generally been replaced by more robust statistical means (38,47). All of the methods discussed or presented may be coupled with an empirical Bayes method to either adjust the false discovery rate (FDR), or define the cut-off level (48).

4.4.1. Ratios of Normalized Values vs. Not Using Ratios

For two-channel microarray data, ratios are commonly used to compare genes based on their fold-changes. However, use of ratios requires the proper propagation of error between the divisor and dividend, which can become complicated if the samples or measures are not statistically independent. Furthermore, ratio values generally do not follow a normal distribution, a requirement of parametric tests. Although it is true that most populations will begin to follow a normal distribution as the population size approaches infinity (due to the Central Limit Theorem), this assumption must be tested prior to appropriately using a parametric analysis technique. Therefore, ratios or the normalized values for data filtering, as the techniques discussed below will work for both types of data provided these assumptions are met.

Note that performing multiple hypothesis tests (e.g., Student's *t*-test) may inflate the false positive rate. That is, the number of genes detected as active by chance alone will increase with the number of genes tested. For example, a microarray with 7000 features would require at least 7000 hypothesis tests per treatment comparison. Several methods have been developed to control the false positive rate, such as the conservative Bonferroni correction and the FDR control method (49).

4.4.2. Student's *t*-test

The *t*-test is a popular method for comparing measures from two samples exposed to different conditions, or from different time-points. The hypotheses that may be tested include: (1) does treatment cause a change in gene expression compared to vehicle treatment and (2) does treatment cause a change in gene expression relative to the control after H hours? The key

Table 11 Data Filtering Methods

Analysis method	Type	Pros	Cons
Student's *t*-test	Parametric	(1) Simple to calculate	(1) Data must be normally distributed
		(2) Easily interpreted	(2) Multiple testing requires false positive rate control
Model-based *t*-test	Parametric	(1) Simple to calculate	(1) Data must be normally distributed
		(2) Easily interpreted	(2) Multiple testing requires false positive rate control
Wilcoxon's rank sum test	Nonparametric	(1) Simple to calculate	(1) Less powerful than *t*-test methods
		(2) Easily interpreted	(2) Multiple testing requires false positive rate control
		(3) Robust to assumptions concerning data distribution	
Empirical Bayes methods	Nonparametric	(1) Robust to assumptions concerning data distribution	(1) Difficult to perform
		(2) May be combined with methods discussed above	(2) Not easily interpreted

to using the *t*-test is that a comparison is being made between two samples. As with all hypothesis tests an alpha value is chosen, typically thought of as a cut-off value for significance, which determines the false positive rate (the rate of finding something significantly changed due to chance alone). Alpha values for most scientific studies are set to $\alpha = 0.05$, meaning the study accepts a 5% false positive rate, regardless of the *p*-value.

The *t*-test also assumes that the data have equal variance and follow a normal distribution. If the data violate the equal variance assumption, VST

methods, or a modified *t*-test that is robust to the assumption of equal variance (present in most statistical software packages) can be applied.

Data filtering can occur using either the *p*-value or the *t*-statistic. Use of the *t*-statistic avoids complications due to arguments concerning estimation of the degrees of freedom, but the false positive rate becomes much more difficult to calculate. The *p*-value is confounded by false positive rate inflation and arguments concerning the degrees of freedom.

The Wilcoxon's Rank-Sum Test (WRST) is a non-parametric alternative. The WRST is robust to the normal distribution assumption, but not to the assumption of equal variance. Furthermore, this test requires that the two groups of data under comparison have similarly shaped distributions. Non-parametric tests typically suffer from having less statistical power than their parametric counterparts. Similar to the *t*-test, the WRST will exhibit false positive rate inflation across a microarray dataset. It is possible to use the Wilcoxon test statistic as the single filtering mechanism; however calculation of the false positive rate is challenging (48).

4.5. Pattern Classification

Following the identification of significant changes in gene expression, the formidable challenge of interpreting the biological relevance of these changes. Borrowing from other large data fields, a common practice is to subject the data to one or more visualizations in order to group similar gene expression change patterns, which may shed light on shared mechanisms of co-regulation and co-ordinated biological function that is involved in the etiology of the observed pharmacological or toxic response.

All pattern classification methods listed group items by similarity. Measurement of similarity differs depending upon the method, and therefore different methods yield different results. Table 12 describes several similarity metrics and why they produce different results for the same data set.

Ultimately, the underlying question dictates the pattern classification method. Sample classification methods are most appropriate where the goal is to classify different samples based on expression. These methods are also known as supervised methods because there is prior knowledge of which samples should cluster and are useful for generating mathematical models that can be trained and validated for the assignment of chemicals based on gene expression fingerprints. If instead the question concerns identifying patterns for individual genes then an unsupervised, or clustering, method would be more appropriate.

A common misconception about clustering methods, especially unsupervised methods, is that all genes within a cluster are co-regulated by the same mechanism. Although genes within a cluster exhibit comparable expression dynamics, their regulatory mechanisms may differ. For example, the patterns of the genes within Figure 7 are such that they all fall within the

Table 12 Distance Metrics Used in Unsupervised Pattern Recognition (i.e., Clustering)

Metric Name	Formula	Description
Minkowski metric	$L_k(a,b) = \left(\sum_{i=1}^{d} \lvert a_i - b_i \rvert^k\right)^{1/k}$	Parent distance metric where a user chooses a value for k to calculate the distance between a and b
Manhattan (city block) distance	$L_1(a,b) = \left(\sum_{i=1}^{d} \lvert a_i - b_i \rvert\right)$	Minkowski metric with $k=1$. The shortest path from a to b walking on a co-ordinate axis. Named for the streets of Manhattan, which resemble a co-ordinate axis
Euclidean distance	$L_2(a,b) = \left(\sum_{i=1}^{d} \lvert a_i - b_i \rvert^2\right)^{1/2}$	Minkowski metric with $k=2$. This metric takes into account both the shape of a pattern as well as the magnitude of the pattern
Pearson's correlation distance	$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}\right)$ σ_{xy} is the co-variance of x and y, σ_x the standard deviation x, and σ_y is the standard deviation y	The correlation co-efficient, ρ, measures the distance between two patterns. This measure is robust to the magnitude of the patterns. Assumes the center of the data is the mean, thus it is mean-centered
Cosine (uncentered) correlation distance	$\rho = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(x_i)(y_i)}{\sigma_x \sigma_y}\right)$	The cosine correlation is the same as the dot product between the vectors of the two patterns, x and y. Conceptually it can be thought of as the cosine of the angle between the two patterns. The cosine correlation assumes the center of the data is 0, thus it is zero-centered
Spearman's rank correlation distance	$\rho = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)}$, $n =$ number of ranked pairs	The Spearman's Rank Correlation is the non-parametric alternative to the Pearson's and Cosine Correlation Distances. No assumption is made concerning the data distribution or the center of the data

same cluster. The simplest conclusion is that these genes follow similar dynamics due to a comparable regulatory mechanism that governs their dynamic expression.

However, clustering methods only identify genes that are highly correlated in their expression, not by their mechanism of regulation, which can only be definitely elucidated through empirical studies.

5. PRACTICAL EXAMPLES IN PHARMACOGENOMIC ANALYSIS

Each of the tools and databases discussed previously is grounded in a significant amount of both technical and theoretical detail. To illustrate the utility of these tools, practical data-analysis examples are provided that outline how a microarray experiment can be designed and analyzed. In addition, the annotation of an uncharacterized EST is defined and mapped to the genome. These examples also demonstrate how to assign annotation to microarray analysis, when the identify of a EST represented on an array may be unknown.

5.1. Designing and Analyzing a Microarray Experiment

5.1.1. Dose-Response: Reference Design

The goal of a dose-response experiment is to identify significant changes in gene expression following exposure to different doses of some treatment. Although gene expression can be compared from dose-to-dose, it is imperative that vehicle gene expression profiles be taken into consideration. Thus, the best design would be the common reference design with dye-swaps to take into consideration dye biases, where every microarray hybridization contains at least one labeling from the pooled reference vehicle sample. Although, the reference design takes many more measurements of the relatively biologically uninteresting vehicle sample, the increased technical replication of the vehicle sample is not a weakness, but instead a strength, due to the importance of establishing vehicle effects within the model. This design should include at least three biological replicates.

5.1.2. Normalizing the Raw Data

Our normalization method (50) is based on a combination of a nonparametric loess smoother and a parametric General Linear Mixed Model (GLMM). As the goal of normalization is to ensure data are comparable across microarrays, that is a value of x represents the same biological concept across all arrays, and the underlying assumption is that most genes on the array will be unchanged, normalized microarray data will typically exhibit the same distribution. The distributions in Figure 8 illustrate that the mean and median are comparable, and all of the distributions have similarly sized tails.

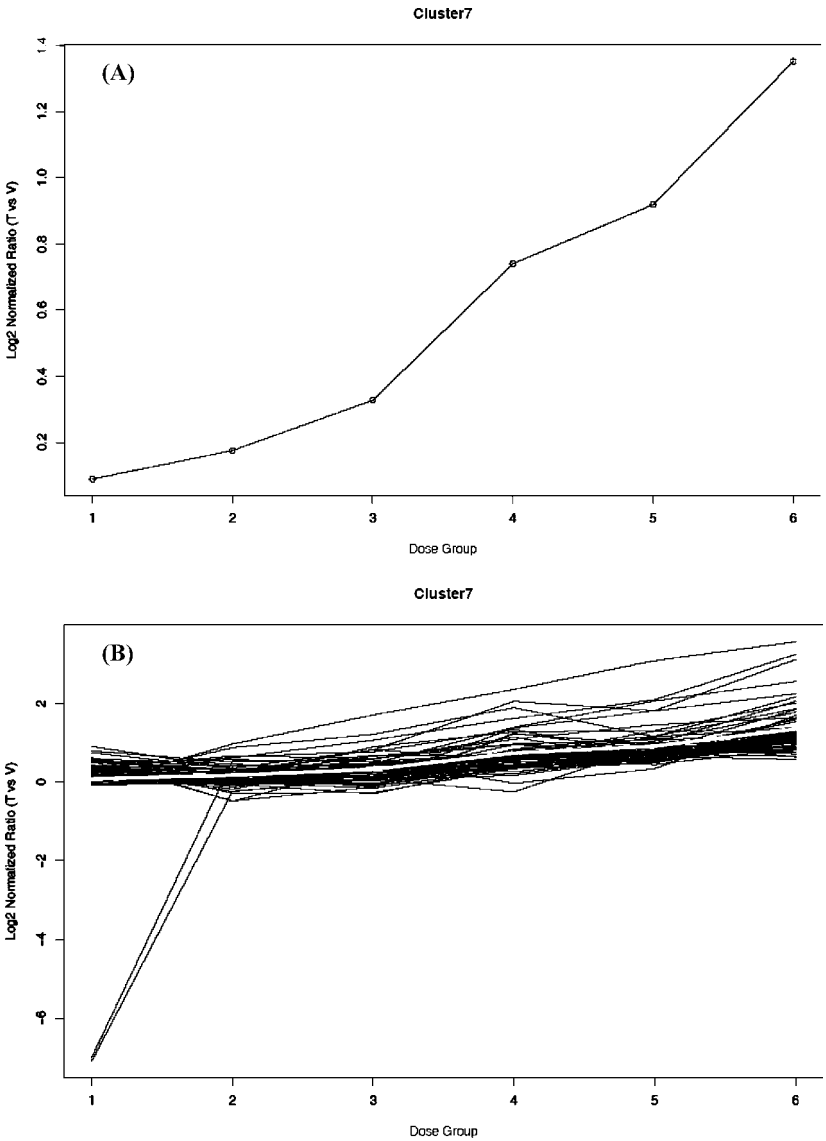


Figure 7 Clustering of genes by expression pattern by k-means clustering. The k-means algorithm clusters genes by expression, where a gene may only exist within one cluster. A correlation-based distance metric was used to determine cluster membership. **(A)** The overall pattern exhibited by this cluster illustrates a dose-response where the response increases with increasing dose. **(B)** Examining the plot of all genes within the class makes it difficult to discern the overall pattern of the cluster. However, this plot places gene expression in context with other genes within the same cluster.

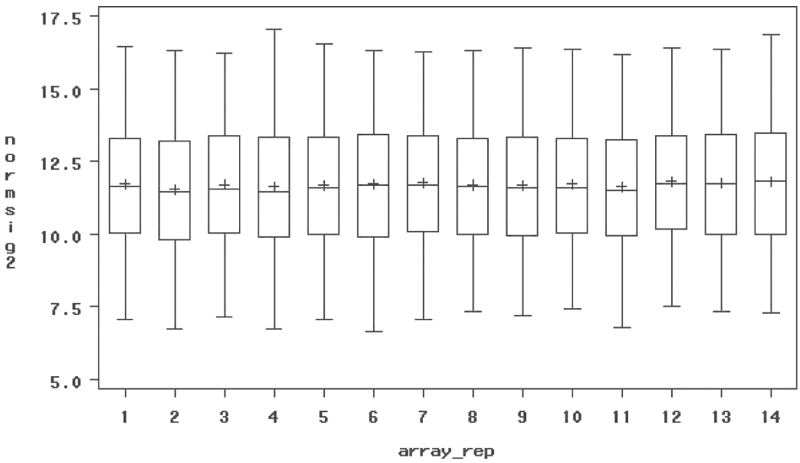


Figure 8 Boxplots of normalized intensity values across arrays. boxplots facilitate comparison of sample distributions and dynamic range across microarrays. If the microarrays follow different data distributions increased false positive and negative rates may be demonstrated. The depicted signal is the $\log_2$ of the normalized feature signal intensity.

5.1.3. Ranking and Prioritization of Data

The most robust method for identifying alterations in transcript levels is the WRST. This method does not make assumptions concerning the distribution or the variance of the data to be compared. However, many investigators use the *t*-test due to its simplicity. Alternatively, a model-based *t*-test from the GLMM could be used which performs contrast tests between two groups of interest, such as an effect at a particular dose vs. a vehicle effect, much like a *t*-test. The values for the test come from the model fit to the data, and provide a better estimate of the difference between the two groups, as other data are used to better constrain the real effect of the dose and vehicle under consideration.

When constructing the GLMM to perform the model-based *t*-test, the terms within the model include the dye (to model dye-effects), the microarray (to model any random microarray-specific effects), and the dose (including vehicle, to model dose/vehicle effects).

5.1.4. Data Interpretation and Mining

Following the identification of active genes, and their ranking and prioritization, the resulting subset is subjected to clustering to identify common patterns of expression using, for example, agglomerative hierarchical clustering. This method results in a dendrogram relating genes by their expression pattern. Alternatively, for a dose response study, a two-way agglomerative

hierarchical clustering, where the doses are also clustered based on expression, could be used to assess which doses are more alike based on gene expression, which may identify subtle changes in target specificity across dose.

Due to the difficulty in identifying the boundaries of clusters in a dendrogram, k-means or fuzzy c-means are also useful. These methods generate a predefined number of clusters where a consensus pattern may be determined and a gene can only belong to one cluster. Although the genes have been clustered, a cluster is of little biological significance. It is difficult to assign a single function to a cluster, as genes with common expression profiles do not necessarily reflect genes with similar functions. Further, clustered genes may or may not represent genes with similar regulatory mechanisms.

In order to identify common functional themes, genes may be clustered by using the Gene Ontology (GO), specifically GO Slims (http://www.geneontology.org/). The GO is an ontology, which can be thought of as a controlled vocabulary with term definitions for defining gene functions, cellular locations, and biological processes. The GO is a directed acyclic graph, where a term, and a gene, may be associated with several other ''higher level'' GO terms representing a broader category, such as ''protein kinase.'' The GO Slims are a ''slim'' version of the GO, providing slightly higher-level mappings of genes to ease clustering and functional annotation. By clustering the filtered dataset by GO entries it may be possible to identify ''over represented'' functions within the cluster. The GO clusters can be further mined for interactions; places where two GO terms share the same genes. The GO numbers with a high degree of interaction may represent functional annotations that are of more importance in the mechanism of action/toxicity.

5.2. Assigning Putative Function to an Uncharacterized EST

Consider this scenario: you have recently designed, executed, and analyzed a complex microarray experiment. With the help of a biostatistician, you have analyzed the results, and now have a list of genes believed to be significantly differentially regulated in your experimental system. However, one of the interesting candidates exhibiting an intriguing expression profile has a UniGene cluster number of Hs.482760, and is annotated as:

Transcribed sequence with weak similarity to MOST-1 Protein

What can you do with this information? Is it possible to link a function to this hypothetical protein? How much trust can you even have in this sequence? Is it really a gene, or could it just be an error on the array? Figure 8 and the following sections detail the steps that can be taken to assess the potential biological validity and functionality of this target gene.

5.2.1. What Is the Raw Sequence?

The UniGene database operated by NCBI clusters full-length mRNAs from short, single-pass sequencing reads called ESTs. While ESTs are inexpensive to sequence, their error rat is typically 3–5%, leading to a large quantity of moderate-quality data. The UniGene database is an attempt to organize ESTs into a set of putative gene products (13).

The first task is to identify a sequence with this cryptic gene name. The UniGene (www.ncbi.nlm.nih.gov/UniGene) cluster number is a unique identifier that facilitates retrieval of the record for this cluster. The record provides weak homology information to another human protein, MOST1, with a region of 63% identity between the two proteins covering 36 residues. In addition, there are no mRNAs matching this set of ESTs, and most importantly, all ESTs that have been grouped together in this cluster are listed. This can be used to select a representative sequence to assess the gene based not only on sequence length, but also the presence of a poly adenylation signal, which provides evidence that of a transcribed gene. For this clusters an EST with accession AA004311 of >500 bp was selected for further analysis.

5.2.2. Does This Sequence Match Any Known Mouse Genes?

While this appears to be a novel gene in human, it is possible that it has already been characterized in another species such as the mouse. Annotation of the mouse genome is nearly complete, and in many cases share similar biology with its human ortholog. The BLAST sequence-alignment tool is used to characterize the homology of this sequence across species by using the novel sequence as a probe and searching a database of all mouse sequences (Table 13). The *E*-value shows the "expectation" of this match occurring by chance in the database. For example, an expectation of one indicates that an alignment of this strength would be expected to occur by random chance once. Lower *E*-values indicate stronger the alignments. The sequence was found to align with moderate strength ($E = 0.015$) to two sequences on chromosome 18, making this the likely location of a mouse homologue. This potential homologue does not appear to be identified and the length of this match (26 bp) is short.

5.2.3. Where Does This Sequence Match to the Genome?

If homology assessment fails to provide insight, it is still possible to map the sequence to the human genome using the BLAST tool offered at the University of California, Santa Cruz (genome.ucsc.edu) (Table 14). The BLAST is an efficient sequence-alignment tool that specializes in locating the best match between genes and the genome, taking into accountant potential splicing. There is only one strong match, on the + strand of chromosome 5. This mapping is 97% accurate, fitting well within the 3–5% error-rate typical of EST sequences, and encompasses nearly 500 of the ~580 base sequence.

Table 13 Assessment of Murine Homology Through BLAST Analysis

Sequence name	Identities	*E*-Value
Mus musculus BAC clone RP23-344I7 from chromosome 18	26/27	0.015
Mus musculus clone RP23-106G21 from chromosome 18	26/27	0.015
Mus musculus BAC clone RP24-267I18 from chromosome 5	21/21	0.23
Mus musculus clone RP23-207F4 on chromosome 2	27/29	0.23
Mus musculus clone RP23-9P8 on chromosome 3	23/24	0.91

5.2.4. What Protein Does This Gene Encode?

To see what the genome looks like in this region, the "browser" link can be used show EST alignment to the genome in this region, thus providing evidence for a real gene-product (Fig. 9). Even more encouragingly, several full-length mRNAs aligning in this area are also shown, and gene-prediction software packages predict a full-length protein in this area that includes the initial sequence used for further analysis.While there are several protein predictions that could be used, all of which might differ by a few residues, the sensitive TwinScan prediction software package (16) identified chr5.87.001.a, and provides further information regarding the putative protein product.

5.2.5. What Is This Protein's Function?

Putative functional domains within the uncharacterized protein can be identified using the CDD search tools while performing protein–protein BLAST (BLASTP) alignments to obtain insight into possible functions of the predicted protein. The predicted protein sequence is used since amino-acid sequences are more evolutionarily conserved than nucleotide sequences, and therefore a broader search across all available sequence information from a variety of organisms can be conducted to identify potential functionality. Results from the BLASTP and CDD analysis (Fig. 10) have identified

Table 14 Assessment of Genomic Mapping Through BLAST Analysis

Score	Identity (%)	Chromosome	Strand
497	97.1	5	+
33	100	1	+
25	100	9	−
21	100	9	−

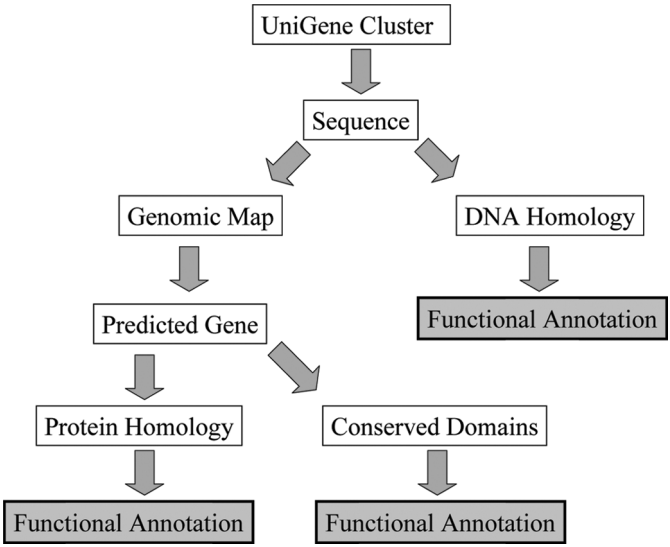


Figure 9 EST-identification pathway. Identification of gene-function based on a novel sequence starts with the knowledge of its UniGene cluster. From this cluster identification one can retrieve the corresponding sequence, which can be mapped against either genomic DNA or against databases of known mRNA transcripts. If the genomic mapping corresponds to a predicted gene, the protein-sequence for this gene can be investigated by conserved-domain analysis and homology-assessment.

three KOG3508, GTPase-activating protein domains with Expectation-Values below 10^{-10}. In addition, an SH2 domain and a RAS-like GTPase domain are also detected. Collectively, this provides strong evidence that the protein is a RAS-interacting GTPase.

The actual BLAST results provide a graphical depiction of the alignments (Fig. 11). The strongest alignments (*E*-values below 10^{-50}) are to RAS-like GTPase proteins, with homology to a protein called RAS p21 protein activator (GTPase activating protein) 1. Investigating the biology of this protein through the LocusLink indicates that it is a close neighbor of our target sequence on chromosome 5, and plays a critical role in neuronal differentiation.

5.2.6. What Have We Learned About the Gene from all This?

This example demonstrated the ability to characterize an unknown DNA sequence of questionable validity using publicly available resources. Although cross-species homology searches bore failed to identify an ortholog, further information and putative identification of the sequence was achieved by aligning it to the genome. Moreover, the genome alignment spanned a region

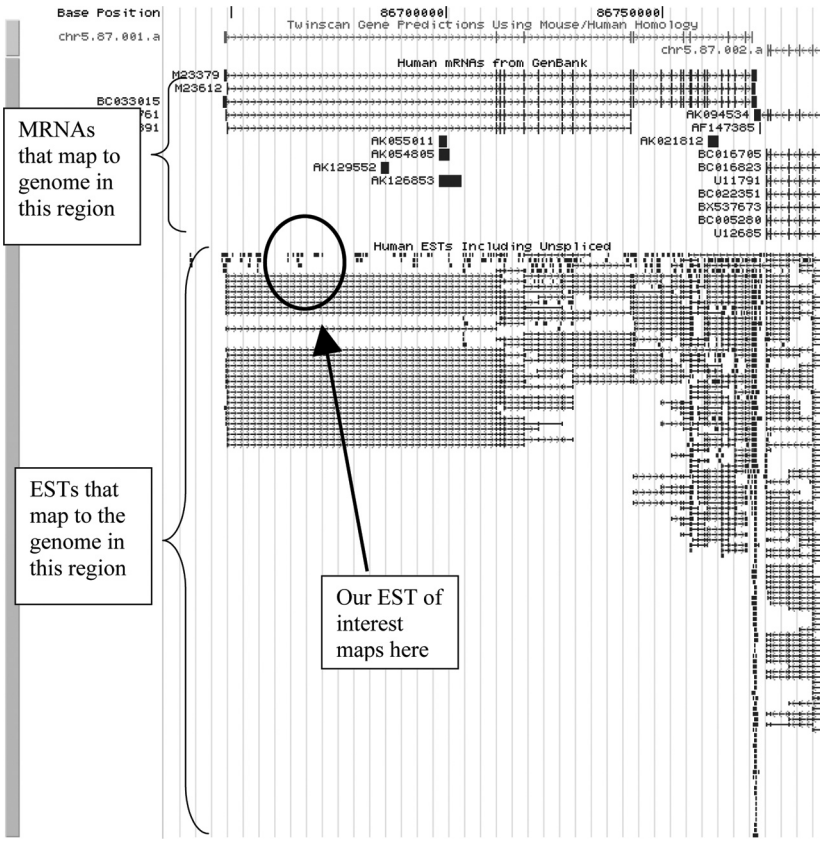


Figure 10 Considering genomic alignment: UCSC genome browser view.

covered by other ESTs, supporting the presence of an actively transcribed gene, rather than a pseudogene. Using the genomic sequence, a predicted protein was examined for functional domains identifying several protein regions with high homology to RAS-like GTPase, which was further supported by the strong protein-level homology to several human RAS-like GTPases (Fig. 12). In addition, the presence of this protein in relative proximity to its closest homologue strongly suggests a gene-duplication event. Empirical studies can now be designed to test the predicted functions, and

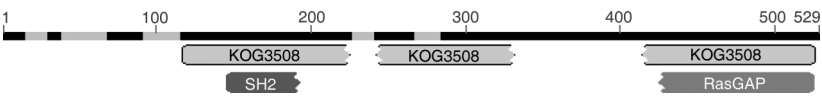


Figure 11 Conserved domain analysis of putative protein.

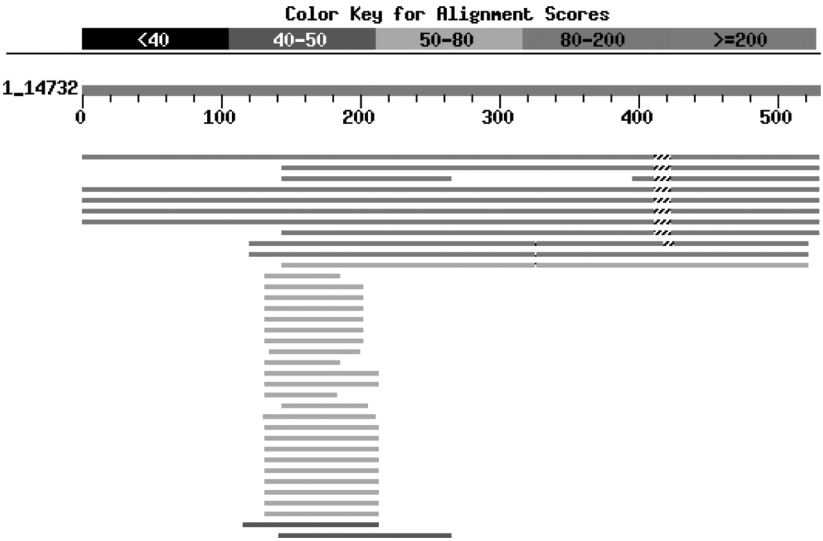


Figure 12 Homology analysis of putative protein.

elucidate the significance of its change in expression in the original microarray experiment.

6. THE COMPUTATIONAL FUTURE OF PHARMACOGENOMICS

Emerging technologies, novel computational approaches, and the availability of human, mouse, and rat genome sequences continues to provide unprecedented opportunities to investigate the mechanisms of action of drugs, toxicants, natural products, and their mixtures that can significantly contribute to safety and risk assessments. The next challenge is to translate this information into decision supportive knowledge by integrating all available disparate data into the assessment process in order to make scientifically sound decisions that minimize potential negative impacts on patients, health provider organizations and the allied health care industry. The current biomedical research "build and test" paradigm that is fueled by combinatorial chemistry, genomic mining for drugable targets, broad coverage omic technologies, and high throughput screening will need to evolve and adopt simulation strategies that have been used in other data and knowledge rich enterprises such as model simulations in the aeronautical field. Central to this objective is the growing importance of computational biology in safety and risk assessments which will not only provide a richer and more refined understanding of the elicited adverse effect through the management and analysis of large, complex data sets, the extraction of knowledge through the integration of disparate data, and the development

of solutions that enhance human health, but will also produce models that are based on empirical data and can accurately predict outcomes. In anticipation of biomedical research becoming more computationally intensive, the National Instutite of Health established the Biomedical Information Science and Technology Initiative (BISTI, http://www.nih.gov/about/director/060399.htm), which was charged with identifying potential impediments in the application of computational approaches in biomedical research. In parallel, the new field of systems biology emerged which can be defined as the iterative development of computational models that integrate disparate biological and other relevant meta data which can be used to predict adverse health outcomes (51–53). More recently, similar efforts have been initiated in pharmacology, toxicology, and risk assessment in order to more accurately predict adverse health effects following exposure to drugs, chemicals, contaminants, natural products, and their mixtures (54–57) (http://www.epa.gov/comptox/comptox_framework.html). It is widely believed, as demonstrated by the examples provided above, that applied bioinformatics and computational biology will play an ever-increasing role in pharmacogenomic research.

REFERENCES

1. Pertsemlidis A, Fondon JW 3rd. Having a BLAST with bioinformatics (and avoiding BLASTphemy). Genome Biol 2001; 2:REVIEWS2002.
2. Muller T, Rahmann S, Rehmsmeier M. Nonmmetric score matrices and the detection of homologous transmembrane proteins. Bioinformatics 2001; 17(suppl 1):S182–S189.
3. Altschul SF, Gish W. Local alignment statistics. Meth Enzymol 1996; 266: 460–480.
4. Gribskov MR, Devereux J. Sequence Analysis Primer. New York: Basingstroke, Hants, England: Stockton Press, Macmillan Publishers, 1991: xv, 279.
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403–410.
6. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389–3402.
7. Baldauf SL. Phylogeny for the faint of heart: a tutorial. Trends Genet 2003; 19:345–351.
8. Phillips A, Janies D, Wheeler W. Multiple sequence alignment in phylogenetic analysis. Mol Phylogenet Evol 2000; 16:317–330.
9. Bork P, Koonin EV. Protein sequence motifs. Curr Opin Struct Biol 1996; 6:366–376.
10. Ponting CP. Issues in predicting protein function from sequence. Brief Bioinform 2001; 2:19–29.
11. Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR,

Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 2003; 31:383–387.
12. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. Global analysis of protein expression in yeast. Nature 2003; 425:737–741.
13. Yuan J, Liu Y, Wang Y, Xie G, Blevins R. Genome analysis with gene-indexing databases. Pharmacol Ther 2001; 91:115–132.
14. Mantegna RN, Buldyrev SV, Goldberger AL, Havlin S, Peng CK, Simons M, Stanley HE. Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 1995; 52:2939–2950.
15. Scherf M, Klingenhoff A, Frech K, Quandt K, Schneider R, Grote K, Frisch M, Gailus-Durner V, Seidel A, Brack-Werner R, Werner T. First pass annotation of promoters on human chromosome 22. Genome Res 2001; 11:333–340.
16. Korf I, Flicek P, Duan D, Brent MR. Integrating genomic homology into gene structure prediction. Bioinformatics 2001; 17(suppl 1):S140–S148.
17. Qiu P. Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. Biochem Biophys Res Commun 2003; 309:495–501.
18. Liu R, States DJ. Consensus promoter identification in the human genome utilizing expressed gene markers and gene modeling. Genome Res 2002; 12:462–469.
19. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 2003; 302:2141–2144.
20. Labrador M, Corces VG. Setting the boundaries of chromatin domains and nuclear organization. Cell 2002; 111:151–154.
21. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 2004; 32:D91–D94.
22. Wasserman WW, Palumbo M, Thompson W, Fickett JW, Lawrence CE. Human–mouse genome comparisons to locate regulatory sites. Nat Genet 2000; 26:225–228.
23. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res 2001; 29:4724–4735.
24. Jaeger JA, Turner DH, Zuker M. Improved predictions of secondary structures for RNA. Proc Natl Acad Sci USA 1989; 86:7706–7710.
25. Tinoco I Jr, Bustamante C. How RNA folds. J Mol Biol 1999; 293:271–281.
26. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003; 31:3406–3415.
27. Smith AD, Lui TW, Tillier ER. Empirical models for substitution in ribosomal RNA. Mol Biol Evol 2004; 21:419–427.
28. Major F, Griffey R. Computational methods for RNA structure determination. Curr Opin Struct Biol 2001; 11:282–286.
29. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001; 294:93–96.

30. Meiler J, Baker D. Coupled prediction of protein secondary and tertiary structure. Proc Natl Acad Sci USA 2003; 100:12105–12110.
31. Tramontano A, Morea V. Assessment of homology-based predictions in CASP5. Proteins 2004; 55:782.
32. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001; 29:365–371.
33. Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D, Sherlock G, Ball C, Lepage M., Swiatek M, Marks WL, Goncalves J, Markel S, Iordan D, Shojatalab M, Pizarro A, White J, Hubley R, Deutsch E, Senger M, Aronow BJ, Robinson A, Bassett D, Stoeckert CJ Jr, Brazma A. Design and implementation of microarray gene expression markup language (MAGE-ML). Genome Biol 2002; 3:RESEARCH0046.
34. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA. ArrayExpress–a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2003; 31:68–71.
35. Simon R, Radmacher MD, Dobbin K. Design of studies using DNA microarrays. Genet Epidemiol 2002; 23:21–36.
36. Yang YH, Speed T. Design issues for cDNA microarray experiments. Nat Rev Genet 2002; 3:579–588.
37. Leung YF, Cavalieri D. Fundamentals of cDNA microarray data analysis. Trends Genet 2003; 19:649–659.
38. Quackenbush J. Microarray data normalization and transformation. Nat Genet 2002; 32(suppl):496–501.
39. Tran PH, Peiffer DA, Shin Y, Meek LM, Brody JP, Cho KW. Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. Nucleic Acids Res 2002; 30:e54.
40. Geller SC, Gregg JP, Hagerman P, Rocke DM. Transformation and normalization of oligonucleotide microarray data. Bioinformatics 2003; 19:1817–1823.
41. Durbin B, Rocke DM. Estimation of transformation parameters for microarray data. Bioinformatics 2003; 19:1360–1367.
42. Huber W, Von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002; 18(suppl 1):S96–S104.
43. Durbin BP, Hardin JS, Hawkins DM, Rocke DM. A variance-stabilizing transformation for gene-expression microarray data. Bioinformatics 2002; 18(suppl 1):S105–S110.
44. Rocke DM, Durbin B. A model for measurement error for gene expression arrays. J Comput Biol 2001; 8:557–569.
45. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002; 30:e15.

46. Yang IV, Chen E, Hasseman JP, Liang W, Frank BC, Wang S, Sharov V, Saeed AI, White J, Li J, Lee NH, Yeatman TJ, Quackenbush J. Within the fold: assessing differential expression measures and reproducibility in microarray assays. Genome Biol 2002; 3:research0062.
47. Dudoit S, Yang HY, Callow MJ, Speed T. Statistical methods for identifying differntially expressed genes in replicated cDNA microarray experiments. Stat Sin 2002; 12:111–139.
48. Efron B, Tibshirani R. Empirical bayes methods and false discovery rates for microarrays. Genet Epidemiol 2002; 23:70–86.
49. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 1995; 57:289–300.
50. Eckel JE, Gennings C, Therneau TM, Burgoon LD, Boverhof DR, Zacharewski TR. Normalization of two-channel microarray experiments: a semiparametric approach. Bioinformatics 2005; 21:1078–1083.
51. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature 1999; 402:C47–C52.
52. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2001; 2:343–372.
53. Kitano H. Systems biology: a brief overview. Science 2002; 295:1662–1664.
54. Klein TE, Altman RB. PharmGKB: the pharmacogenetics and pharmacogenomics knowledge base. Pharmacogenomics J 2004; 4:1.
55. Olden K, Wilson S. Environmental health and genomics: visions and implications. Nat Rev Genet 2000; 1:149–153.
56. Tong W, Cao X, Harris S, Sun H, Fang H, Fuscoe J, Harris A, Hong H, Xie Q, Perkins R, Shi L, Casciano D. ArrayTrack—supporting toxicogenomic research at the US Food and Drug Administration National Center for Toxicol Res. Environ Health Perspect 2003; 111:1819–1826.
57. Waters M, Boorman G, Bushel P, Cunningham M, Irwin R, Merrick A, Olden K, Paules R, Selkirk J, Stasiewicz S, Weis B, Houten BV, Walker N, Tennant R. Systems toxicology and the chemical effects in biological systems (CEBS) knowledge base. Environ Health Perspect 2003; 111:811–824.
58. Ballman K, Grill D, Oberg A, Therneau T. Faster cyclic loess: normalizing DNA arrays via linear models. Technical Report 2003; 68:1–23.
59. Cheadle C, Cho-Chung YS, Becker KG, Vawter MP. Application of z-score transformation to Affymetrix data. Appl Bioinformatics 2003; 2:209–217.
60. Kerr MK, Churchill GA. Experimental Design for Gene Expression Microarrays. Biostatistics 2001a; 2:183–201.
61. Kerr MK, Churchill GA. Statistical design and the analysis of gene expression microarray data. Genet Res 2001b; 77:123–128.
62. Smyth GK, Speed T. Normalization of cDNA microarray data. Methods 2003; 31:265–273.
63. Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 2001; 8:625–637.

23

Mapping of Disease Loci

Glenys Thomson and Ana Maria Valdes

Department of Integrative Biology, University of California, Berkeley, California, U.S.A.

1. INTRODUCTION

The complete sequencing of the 3000 million bases (Mb) of the human genome has had a tremendous impact on the research community's ability to discover the genes that underlie human genetic variation (1,2). Recent successes in the genetic mapping of Mendelian traits and diseases have been remarkable. However, progress has been exceedingly slow in elucidating the genetic components of complex diseases and traits, such as cardiovascular diseases, asthma, diabetes, rheumatoid arthritis, obesity, alcoholism, and schizophrenia. Genetic variation is likely to contribute to the risk of many complex diseases; in some cases the genetic component may be small compared to environmental factors. The fact that genes and environment interact only adds to the challenge of unraveling the etiology of these disorders. Rather than being due to specific and relatively rare mutations, complex diseases and traits may result principally from genetic variation that is relatively common in the general population. Examples include apolipoprotein E gene (*APOE*) variation and age of onset of Alzheimer's disease, the angiotension-converting enzyme gene (*ACE*) and myocardial infarction, the chemokine receptor CCR5 gene (*CMKBR5*) and the risk of infection in those exposed to HIV, and the many diseases for which immune response genes of the HLA region have been implicated.

In the following, we talk about Mendelian or complex diseases with the understanding that the statements also apply to non-disease traits. Diseases caused by a single major gene or biochemical pathway defect are referred to as Mendelian or single gene traits. Complicating factors such as incomplete penetrance—not all genetically predisposed individuals manifest the disease, as in polydactyly, and variable age of onset effects, as in Huntington disease, are often present. However, the basic single gene nature of the inheritance pattern is often evident, with Mendelian segregation patterns according to recessive, dominant, or sex chromosome X-linked inheritance (see Ref. 2 for review).

Sir Archibald Garrod's brilliant studies in 1908 of the genetics and biochemical pathways involved in "inborn errors of metabolism" laid the conceptual foundation for all later studies in human genetics (2). The single gene Mendelian nature of a large number of traits and diseases was subsequently identified. Gene mapping in humans before the 1970s was however essentially limited to the X chromosome, based on the specific pattern of inheritance. Future progress in biochemical genetics and gene mapping awaited advances in molecular biology.

A Mendelian disease can have complex clinical effects, e.g., sickle cell anemia affects many organ systems. Even with the same allelic mutation, as in individuals with sickle cell anemia, there may be considerable variation in disease severity, due to modifier genes. The majority of Mendelian traits also show allele and/or locus heterogeneity. Different alleles at one gene may give rise to disease, e.g., cystic fibrosis, which can result in variation in the disease expression. Different genes may also be responsible for disease in different families, e.g., the *BRCA1* and *BRCA2* genes and early onset familial breast cancer.

In contrast to Mendelian diseases, complex, or multifactorial diseases, are generally more common, and result from the interaction of multiple genes and environmental factors. Studies of rates of occurrence of disease in twins and other family members compared to population level rates are used to demonstrate the role of genetic and environmental factors in complex diseases (3,4).

The distinction in terminology between Mendelian and complex traits does not imply that complex diseases do not follow in general the rules of Mendelian inheritance, but rather that there is no obvious simple pattern of inheritance. Although the boundary between these two categories is not precisely defined, a large number of diseases clearly fall into each category. It is also important to remember that while we use the shorthand notation of disease gene, or disease-causing gene, we are actually talking about variation at genes involved in normal human health and development, specific variations of which may contribute to the development of a disease (5).

Genetically complex diseases may exhibit the following properties: incomplete penetrance—not all susceptible individuals are affected; the

involvement of several disease-predisposing loci—some of which may have a major effect, but many of which may have a relatively minor effect; interaction effects between these loci and with environmental factors; and heterogeneity—so that different loci and/or alleles cause disease in different groups. As a result it is difficult with complex diseases to: localize disease genes; ascertain the number and relationship of disease loci involved; understand modes of inheritance and interaction effects; determine the molecular basis of disease; and understand the mechanism(s) by which these genetic changes give rise to disease, including all genetic and environmental factors and their interactions.

For many complex diseases the fact that a large number of genes, many with relatively small effects, are involved drastically complicates efforts to identify genetic regions involved in the disease process, and makes replication of results difficult. Additionally, each complex disease can present its own set of unique problems and statistical issues in the localization of disease genes and definition of disease predisposing factors, including: a low or high disease prevalence; late-onset diseases for which parental data may be difficult to obtain; and infectious diseases where family studies are difficult and issues of exposure confound the analyses.

2. MAPPING STRATEGIES, DNA POLYMORPHISMS, AND POPULATION PARAMETERS

2.1. Mapping Genes Involved in Disease

There are three main approaches to mapping the genetic variants involved in a disease: functional cloning, the candidate gene strategy, and positional cloning (2). In functional cloning, identification of the underlying protein defect leads to localization of the responsible gene (disease–function–gene–map). Sickle cell anemia was the first human disease to be successfully understood at the molecular level and isolation of the mutant molecule led to the localization of the β-globin gene on chromosome 11. Functional cloning is, however, only useful in a subset of Mendelian traits where the biological basis is known, and in very few complex traits.

In the candidate gene approach, genes with a known or proposed function with the potential to influence the disease phenotype are investigated for a direct role in disease (2). The most successful application of the candidate gene approach to mapping complex diseases in humans has been with the HLA region on chromosome 6p21, the major histocompatibility complex of humans. Genes in the HLA region have been implicated in the etiology of over 100 diseases (6). These include: complex autoimmune diseases such as type-1 diabetes (until recently called insulin-dependent diabetes mellitus—IDDM), rheumatoid arthritis, and multiple sclerosis;

cancers, e.g., Hodgkin disease; infectious diseases such as malaria, tuberculosis, and AIDS; and also other diseases such as narcolepsy.

Positional cloning is used when the biochemical nature of a disease is unknown. Marker genes not related to disease physiology and genome-wide screens are then the starting point for mapping the genetic components of the disease. The aim is first to identify the genetic region within which a disease-predisposing gene lies and, once this is found, to localize the gene, and determine its functional and biological role in the disease (disease–map–gene–function). Most markers used nowadays are DNA based.

Landmarks in positional cloning have included genetic linkage of Huntington disease in 1983 to an RFLP (restriction fragment length polymorphism) marker (Sec. 2.2) on chromosome 4 (the gene was cloned in 1993) and the cloning of the cystic fibrosis gene on chromosome 7 in 1989. By 1990, when the human genome project began, a handful of additional successes had accrued. In 1994, the early onset familial breast/ovarian cancer (*BRCA1*) gene was cloned, and the early onset familial breast cancer (*BRCA2*) gene in 1995. By 1997 close to 100 Mendelian disease loci had been identified by positional cloning (2).

A complementary array of approaches is available to uncover the different genetic facets of both Mendelian and complex traits and diseases (7,8). Two complementary analytic methods, linkage analyses, and association [linkage disequilibrium (LD)] mapping (Secs. 3 and 4), are used to detect the specific genetic regions and genes involved in the disease process. These approaches can be applied without prior knowledge of the biological basis of the disease using genome-wide studies, together with the candidate gene approach and comparative analyses using animal models of disease.

2.2. DNA Polymorphisms

Markers such as protein and blood group loci were initially used in the analysis of genetic traits; however they were limited in utility due to low variation. Early in the 1980s these markers began to be replaced with DNA polymorphisms, initially RFLPs which are detected by the ability of a segment of DNA to be cut, or to not be cut, by a specific restriction enzyme.

Microsatellites have rapidly replaced RFLP markers in studies to map disease predisposing genes since they occur frequently and randomly across the human genome, have high levels of variation, are easy to type, and can be amplified using the polymerase chain reaction (PCR) methodology and hence only a small amount of template DNA is required. Single-nucleotide polymorphisms (SNPs) are the most recent type of genetic marker to be considered. They are the most common type of human DNA genetic variation, occurring on average 1 per 1000 bp (9). The SNPs are mostly bi-allelic and less informative than microsatellites, however they are more frequent and

mutationally more stable than microsatellites, and more amenable to automation and DNA chip technology.

While the complete sequencing of the human genome has greatly aided studies of Mendelian and complex diseases, it has not provided an immediate solution to their genetics, especially for complex diseases. We must still address the daunting prospect of documenting the genetic variation of human genomes at the population level within and across ethnic groups. Only then, correlating this extensive genetic variation with disease, can we uncover the complete genetics of complex diseases and traits, and determine the environmental factors impacting on each disease (5).

2.3. Linkage Disequilibrium

Before discussing linkage and association mapping it is important to distinguish between two concepts: physical linkage and linkage disequilibrium (LD). The description of genetic variation at the population level begins with consideration of allelic variation at a single genetic locus. The next step is to consider genetic variation at two or more loci simultaneously, including non-random associations. The non-random association of alleles at different genetic loci is termed LD, with random association described as linkage equilibrium. The pairwise LD parameter, usually denoted D, is the difference between the observed frequency of a two-locus gametic (chromosome) type—usually referred to as a haplotype, and the frequency expected on the basis of random association of alleles in gametes, i.e., $D = f(AB) - f(A)\,f(B)$, where $f(.)$ denotes the frequency of a gamete (haplotype) or allele.

There is a relationship between D and the recombination fraction θ between two loci: the value of D decreases by a fraction $(1-\theta)$ each generation under random mating and a neutral model. Thus the more loosely two loci are linked, the faster the decay of LD. It is possible, although relatively rare, for loci that are unlinked to be in significant LD, and for loci that are physically very closely linked to be in linkage equilibrium. However, in general, LD is usually only seen between very closely linked loci and is rare otherwise. General population level observations are that there is an overall proportionality between LD and the inverse of the recombination distance (10). This rule however breaks down in very closely linked regions (10–12).

The LD can occur in populations as a consequence of a number of factors (see Ref. 13 for review): mutation—when a new mutant arises, it occurs in one individual and is in LD with all polymorphic loci in the population; selection—either acting directly on the two loci, or transient LD can also be created with neutral loci via a hitchhiking event with a selected locus (14); migration or admixture—generation of LD by these forces requires that the allelic frequencies of both loci in the two populations be different, and this difference must be substantial to generate very much LD; and random genetic drift—while the expected value of pairwise LD due to drift over

many generations is zero, the variance is large for closely linked loci in small populations.

The amount of LD observed in a population is also affected by a number of factors: recombination—significant LD may be maintained for a long time between very closely linked loci; selection—if the selection is sufficiently strong compared to the recombination rate LD can be maintained at an equilibrium state, and will occur transiently with a hitchhiking event; non-random mating—high levels of inbreeding and self-fertilization in plants can retard the rate of approach to linkage equilibrium; and population demographics—a small founder population or a bottleneck in the recent past can cause significant LD due to genetic drift for closely linked loci. While less LD will be generated by genetic drift in a rapidly growing population (15), any LD present before or during the early phase of the expansion will persist.

Selection is expected to produce LD non-randomly across the genome, whereas all the other factors should act randomly over the genome. Some regions, such as HLA, show strong evidence of selection and significant LD which spans 3 cM or more. The LD and haplotype structure of the genome is currently being studied by the National Institutes of Health Haplotype Map project (Sec. 4.5).

Population and evolutionary aspects of a genetic region are highly relevant to the study of Mendelian and complex diseases. These relate to our ability to localize and identify disease predisposing variants, and to our understanding of the mechanisms by which some disease predisposing genes can become relatively common in a population.

3. LINKAGE MAPPING

3.1. Linkage Analyses

Linkage analyses test for co-segregation of a marker and disease phenotype within a pedigree (implicating physical linkage between the marker gene and a gene involved in the disease process, and such loci are then said to be ''linked'') (16–18). Linkage analyses do not require LD between the marker and disease loci and different marker alleles may segregate with disease in each pedigree.

Linkage analyses can be performed using single marker locus variation, and any kind of inherited difference can potentially be a genetic marker. For genome-wide linkage analyses, testing for co-segregation with disease of about 300 to 400 highly polymorphic markers, usually microsatellites, distributed approximately evenly over the genome (average spacing between markers is then on the order of 10 cM, i.e., 10% recombination) is the usual practice. The increasing availability of more markers across the genome, combined with multi-point analyses using a number of

closely linked markers, such as microsatellites and SNPs—hence increasing the informativeness with respect to co-segregation, will increase the power of linkage studies.

When large multi-generation pedigrees are available, linkage analysis is a powerful technique for localizing disease genes and has been successfully applied to a number of monogenic traits, e.g., Huntington disease, and the familial breast cancer genes *BRCA1* and *BRCA2*. With diabetes, it has been used to map maturity onset diabetes of the young (*MODY*) genes. The occurrence of multiple alleles at one genetic locus involved in disease, as in cystic fibrosis, does not affect the power of linkage analyses. However, if different Mendelian genes are involved in disease in different families as in *MODY*, power can be reduced if all the families are analyzed together.

For complex diseases, the involvement of many genes and the strong influence of environmental factors means that large multi-generation pedigrees are rarely, or never, seen. Linkage analysis of nuclear families with both parents and two children affected with the disease, although less powerful, is therefore more commonly used to map complex traits (5,19,20). Consider an affected sib pair family (Fig. 1), and assume for simplicity we can distinguish all four parental chromosomes in the genetic region under study. Deviation from the Mendelian random expectations of 25%, 50%, and 25%, that the affected sibs will on average share two, one, and zero parental chromosomes in common that are identical by descent (IBD) implicates a disease predisposing gene in the region.

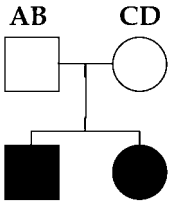
3.2. Linkage Studies in Complex Diseases

Linkage of the HLA region to type-1 diabetes was demonstrated initially with 15 affected sib pairs (21) (Table 1a). This linkage, termed IDDM1, has been confirmed in many studies (Table 1b), with a mean IBD sharing of parental alleles for HLA of 72% (22–25). For some other HLA associated diseases, e.g., multiple sclerosis and rheumatoid arthritis, the initial number of affected sib pairs required to show evidence of linkage has been larger, at least 50 and sometimes around 100 (26,27).

The existence of non-*HLA* genes in many of the HLA associated diseases was established from theoretical considerations involving population prevalence, risks to relatives, and HLA IBD values in affected sib pairs (28–30). Type-1 diabetes shows an increased risk in siblings over population prevalence (λ_s) of 15 in Caucasian populations (30), of which HLA contributes 3.4: approximately 45% of the type-1 diabetes genetic component under a multiplicative model (31).

Given the relative ease with which linkage was demonstrated for many HLA associated diseases, it seemed a logical progression to use genome-wide linkage scans on affected sib pair families to investigate all complex diseases. Such studies of many complex disorders are in progress: to map

Affected Sib pairs



			AFBACs
AC	AC	share 2 (25%)	B, D
	AD	share 1 (25%)	B
	BC	share 1 (25%)	D
	BD	share 0 (25%)	_

Figure 1 Affected sib pair families. A nuclear family pedigree is shown with the father (*square*) and mother (*circle*) in the first row and the two affected children, which can be of either sex in the second row. Assume for simplicity we can distinguish all four parental alleles, denoted A, B, C, and D, in the genetic region under study, with the parental alleles ordered such that A and C are transmitted from the father and mother, respectively, to the first affected child (19,20). Four possible configurations among the two offspring with respect to the alleles inherited from the parents are possible: they can share both parental alleles (A and C); they can share an allele from the father (A) but differ in the alleles received from the mother (C and D); they can share an allele from the mother (C) but differ in the alleles received from the father (A and B); or they can share no parental alleles in common. These four configurations are equally likely (25%) if there is no influence of the genetic region under consideration on the disease. If there is a linked gene influencing disease, there will be deviation from these ratios in the direction of increased sharing. The parental alleles that are never transmitted to the affected sib pair in each family type are used as a control population in association studies using nuclear family data, the so-called AFBAC (affected family based control) sample (19,20, and Sec. 4.2).

Table 1 The HLA and IDDM Affected Sib-Pair Values

(a) Ref. 21 (from Ref. 19)				
IBD sharing	2	1	0	
Observed	10	4	1	Total 15
Observed (%)	67	27	6	
$\chi^2_2 = 14.07$, $p < 0.001$, mean IBD sharing $= 0.805$				
(b) Ref. 22, 538 families (from Ref. 20)				
IBD sharing	2	1	0	
Observed	373	283	55	Total 711
observed (%)	52	40	8	
$\chi^2_2 = 314.03$, $p < 10^{-5}$, mean IBD sharing $= 0.724$				

the non-*HLA* genes in a number of diseases such as type-1 diabetes, multiple sclerosis, rheumatoid arthritis, celiac disease, and Crohn's disease; and for many other complex diseases, for example type-2 diabetes, hypertension, coronary artery disease, alcoholism, and schizophrenia.

Type-1 diabetes was the first complex disease for which genome-wide linkage scans in affected sib pairs was carried out (32–34). Many other studies using genome-wide scans to detect non-*HLA* type-1 diabetes genes have followed since these papers (reviewed in Ref. 23). Apart from *HLA*, no evidence of major gene effects has been found. Using approximately 500 affected sib pairs, linkages for the chromosomal regions IDDM4 (11q13), IDDM5 (6q25), and IDDM8 (6q27) were confirmed (35) using the criteria of Lander and Kruglyak (36). The mean IBD sharing values in these cases were much closer to the 50% expected randomly than seen with IDDM1: 0.58, 0.58, and 0.60, respectively for IDDM4, IDDM5, and IDDM8. The IDDM6 (18q12–21), IDDM10 (10p11-q11), and IDDM12 (2q33) have more recently been considered confirmed, using a combination of affected sib pair and association/linkage tests (23,37–39). Seven additional chromosomal regions have shown evidence of linkage to type-1 diabetes in one or more studies (reviewed in Refs. 23–25).

3.3. Quantitative Trait Loci

Not all human disease traits are binary (affected/unaffected). Mapping the genes affecting continuous traits such as blood pressure, bone mineral density, or serum lipid levels is also of major importance. In order to map such quantitative trait loci (QTL) variance components methods (40–42) are commonly used. In a variance component approach the variability in the value of a quantitative trait from individuals within a pedigree is modeled as a function of a major locus affecting the trait plus any relevant covariates plus residual genetic and non-genetic effects. The effect attributable to a locus linked to a marker will be a function of the additive (V_a) and dominance (V_d) components of the variance of the locus, the recombination fraction, and the proportion of alleles IBD at the marker locus (40). The parameters involved are then estimated using maximum-likelihood methods. Although this approach has the deficiency of a strong assumption of normality, it also has some very attractive features, namely its interpretability and its computational tractability in the incorporation of covariates, multivariate phenotypes, interactions, etc.

3.4. Observations from Genome-Wide Linkage Scans in Complex Diseases

A common feature of all complex disease studies has been difficulty in both detecting and replicating linkages, with considerable heterogeneity seen between data sets both within and between populations and ethnic groups.

The possible involvement in disease of relatively common alleles of a large number of loci, each with a relatively small effect overall, requiring study of many hundreds, and usually thousands, of affected sib pairs to establish linkage, can explain this phenomenon (24,43,44). The additional possible involvement in some cases of complex disease of relatively rare variants further confounds our abilities to detect all genes involved in the disease process.

Two second generation type-1 diabetes genome scans continue to illustrate the heterogeneity seen in linkage studies in complex diseases (24,25). Even more surprising, little or no support was found for most reported type-1 diabetes loci in one of these studies (25) even with a very large sample size. The use of stratification analyses based on known IDDM loci such as in the HLA region (IDDM1) may have contributed to more examples of evidence of IDDM loci in the second study (Ref. 24, Sec. 7.2).

Despite initial anxieties with genome-wide linkage scans that we would be flooded with false positives (36), this has not been the case. An excess of false linkages has never been a problem with complex diseases. Instead we are scrambling to find any evidence of disease-predisposing genes. In an insightful review of the many factors that need to be considered in the design of genome-wide scans and interpretation of linkage results, Rao (45) pointed out that application of stringent criteria drastically reduces power, and leads to many disease genes remaining undetected (46,47). Rao's recommendation was that "we tolerate/accept, on average, one false positive per individual scan." The issue is to have a reasonable balance between false positives and false negatives, while understanding that false negatives are the more serious concern (45,48,49).

A major challenge for human genetics remains to understand the genetic etiology of common diseases such as cardiovascular diseases, osteoporosis, osteoarthritis, type-2 diabetes, and psychiatric disorders. Findings on the genetics of these complex disorders have often been equivocal. Complex diseases and most traits of clinical interest are influenced by many genetic and environmental factors (50,51). In spite of these problems there are examples of genes influencing complex diseases which have been successfully mapped using linkage analysis. Examples include the presenillin genes for Alzheimer's disease (52) and the calpain 10 gene for type-2 diabetes (53).

3.5. Linkage Analysis Without Parental Genotypes

For late onset diseases such as Alzheimer's disease, detection of genes is even more difficult, because parental information may be difficult or impossible to obtain. The use of siblings has the advantage of avoiding the difficulties of recruiting parents for the study of late-onset conditions. The utility of sib pairs for QTL linkage analysis is well established (17,18). If a marker locus is linked to a QTL the difference in the value of a trait between two sibs in

a pair is expected to decrease as they share more markers IBD. In the absence of parental data, the proportion of IBD alleles at the trait locus can be estimated by using a multiple regression that incorporates all the linked markers on a chromosome containing the putative QTL. Variance components procedures which can make use of sib pair data without parental information exist (54).

A major problem of such approaches is that they have extremely low statistical power. For example, for a tightly linked marker ($\theta = 0$) to trait-related locus which in turn accounts for as much as 15% of the variance in the quantitative trait ($V_a = 0.15$) Sham et al. (55) estimated that 5322 sib pairs would be needed. Although a number of twin cohorts with very extensive phenotyping exist (4) which could potentially be used for linkage studies, the low statistical power involved renders them less attractive for mapping QTLs.

A transmission disequilibrium test that does not require parental data has also been developed (56). The sib-TDT method uses the marker data from unaffected sibs instead of from parents, thus allowing application of the principle of the TDT to sibships without parental data.

4. ASSOCIATION (LD) MAPPING

4.1. Association Analyses in Case/Control Data

Association studies compare marker frequencies in unrelated cases and controls. If a particular variant increases susceptibility to a disease of interest, we will find that it is more common among affected individuals than among random controls. Random markers in LD with a disease susceptibility mutation may also be associated with the disease and this will usually imply close physical linkage of the marker and disease predisposing gene (Sec. 2.3) (7). In association studies it is essential that the patient and control groups be ethnically matched (13,57). If they are not, a spurious association of an unlinked marker with disease can result from population stratification. In such cases, if the prevalence of the disease is different between the ethnic groups mixed in the samples under study, association with marker alleles whose frequencies vary between the ethnic groups would be falsely detected as being associated with the disease.

The success of association (LD) studies to fine map Mendelian traits following initial localization of chromosomal regions by linkage analysis is well documented in many different populations (11). In these studies LD extended on average over 500 kb (0.5% recombination). There are exceptions to the use of association data in localizing disease genes; the breast cancer gene *BRCA1* could not be localized using LD mapping as different mutations were implicated in each family.

Association mapping may in many cases be more efficient than linkage analyses in detecting genetic regions involved in disease (8,44). IDDM2—the minisatellite variable number of tandem repeats (VNTR) 5′ to the insulin gene (58), which is easily detected by association, is difficult to detect by linkage analysis (59). The well established and strong HLA association with multiple sclerosis is not picked up in all genome linkage scans (60). As with linkage studies, haplotype level analyses can be more powerful than single locus analyses (61,62).

At least 50,000 evenly spaced markers, giving an average distance between markers of 0.06 cM (~60 kb), are required for an initial disease genomic association scan. Association scans are thus over two order of magnitude higher than the 300 to 400 markers at 10 cM typically used for linkage genome scans. The use of pooled samples of DNA for the study of RFLP (63), microsatellite (64–66), and SNP variation (67), as well as the current development of DNA chip technology for the study of SNPs (9,68), has opened the way for future routine and extensive use of association mapping in the study of complex diseases. Genome-wide association scans in complex diseases are starting to be used, currently with DNA pooling and microsatellites. Methods for correction with microsatellites of stutter artifact and preferential amplification have been demonstrated (65). On the one hand we must note that a pooled DNA genomic association screen can be carried out without using these correction factors; no bias is introduced as patient and control samples are similarly affected under the null hypothesis of no marker association with disease. Further, markers should be chosen so as to reduce these artifacts. On the other hand, pooling strategies come at a cost in reduced information about each sample such as the ability to include covariates in an associaton model (50).

Many researchers feel that the number of false positives that will result from genome-wide association scans make them unmanageable. A preliminary discussion of power and a multi-stage strategy to reduce false positives in genome-wide association studies is outlined in Ref. 65. As with linkage studies (45,49), power should be high in an initial association scan, to reduce missing associations (false negatives), and sample sizes of 500–1000 are recommended (65,69). Long and Langley (69) have demonstrated that greater power is achieved in association genome scans by increasing the sample size than by increasing the number of markers.

The potential power of association studies to detect disease genes depends on several unknown parameters, and cannot be determined accurately, especially given that common genetic variants may often be involved in complex diseases (50,70–73). For example, the nature of mutations involved in the disease process will influence the power of association studies to detect disease genes (70). Also, the power of genome association scans will vary across the genome, due to variation in LD values (10), as well as between populations. The full development of SNPs will eventually permit

routine typing for variation in every human gene and its regulatory region, the ultimate association study (9,74,75).

Association mapping is appropriate for monogenic and complex diseases, and may always be preferable to linkage analyses for late onset diseases where it is difficult to obtain nuclear families, rare diseases for which multiplex pedigrees may not be available, e.g., type-1 diabetes and multiple sclerosis in China, and infectious diseases. Study design can incorporate disease heterogeneity and interaction effects between loci (Sec. 7.2).

4.2. Association Analyses in Nuclear Family Based Data

The use of nuclear family data in association studies was initially developed to avoid possible ethnic mismatching between patients and randomly ascertained controls (76). The parental marker alleles not transmitted to an affected child, or never transmitted to an affected sib pair, form the so-called AFBAC population (19,20) (Fig. 1). In a random mating population, when there is a marker association with disease, the AFBAC population provides an unbiased estimate of the overall population (control) marker alleles when the recombination fraction between the marker and disease genes is sufficiently small that it can be taken as zero ($\theta = 0$), and differences between patient and AFBAC frequencies can be tested for example by a contingency table analysis for heterogeneity.

Of more importance, however, is the fact that concerns about population stratification effects creating associations between unlinked loci are less problematic with family based data. Testing for a 50% transmission ratio from parents heterozygous for a marker allele (the TDT—transmission disequilibrium test) detects significant differences with marker alleles that are in both LD and linked to the disease predisposing gene (77–79). This is because a compound null hypothesis is always being tested which involves a term which is zero if the marker and disease are unlinked, and a term which is zero if there is no LD at the population level. Thus, associations at the population level of unlinked marker loci with disease caused by population stratification, migration, or admixture, are eliminated, and significant effects are only seen for marker loci linked and in LD with the disease loci. Neverthless, it is preferable to use ethnically matched families to ensure that the same allele markers are associated consistently.

The recent emphasis on use of only family based association/linkage tests has ignored the readily available resource of case/control data (57), where sampling and then study of many thousands of samples are feasible with new techniques. Provided the patient and control groups are carefully matched for ethnicity, population stratification effects creating spurious associations are eliminated. The large collections of multiplex families now available for linkage studies in many complex diseases are obviously also a valuable resource for association screens (65,80).

4.3. Association Studies in Complex Diseases

Association mapping has been the main approach in implicating genes of the HLA region in disease etiology (Table 2). Association studies in the early 1970s distinguished type-1 (HLA associated) from type-2 (not HLA associated) diabetes. In some cases the classical HLA class I or II genes have been implicated as directly involved in disease, e.g., ankylosing spondylitis (class I), and type-1 diabetes, rheumatoid arthritis, celiac disease, and narcolepsy (class II). In other cases, the association with classical *HLA* genes was due to LD, as with hemochromatosis. Association mapping has also been involved in mapping non-*HLA* genes. In a small subset of type-2 diabetes cases association studies have identified the role of candidate genes in disease, e.g., the insulin gene, insulin receptor gene, glukokinase gene, and genes in the mitochondrial genome (81).

The potential of association studies as the optimal strategy for follow-up of regions showing preliminary evidence of linkage for complex diseases has been demonstrated (24,82), and awaits full utilization. As with linkage studies, heterogeneity in association results for complex diseases is also seen between studies, especially between populations, and is expected (24,38,69). In their simulation study, Long and Langley (69) concluded that association studies have a low repeatability unless sample sizes are on the order of 500 individuals.

Table 2 The HLA-Associated Diseases

HLA	Patients (%)	Controls (%)
Ankylosing spondylitis		
B27	90	9
Type-1 diabetes		
DR3	52	23
DR4	74	24
DR3 or DR4	93	43
Multiple sclerosis		
DR2	86	33
Rheumatoid arthritis		
DR4	81	33
Narcolepsy		
DR2	>95	33

For each disease, the frequency of the associated *HLA* allele is given in patients and controls. The letter designation denotes the *HLA* gene, while the number is assigned to a specific allele at the gene. For ease of reading, the data shown are older serological level HLA typing, rather than more recent molecular typing. *Source*: Modified from Ref. 5 (6).

4.4. Candidate Gene Studies

Before the early 1980s, genetic risk factors for a disease or trait could be identified only through direct analysis of candidate genes, usually through association studies. Candidate genes are genes hypothesized to be involved in the disease on the basis of knowledge of the pathogenesis. In general, this is carried out by comparing allele or genotype frequencies between a group of unrelated patients with ethnically, and preferrably age and gender, matched controls.

Candidate genes can be selected using a variety of approaches such as literature searches, genes resulting from experiments in animal models, homology searches, and gene expression experiments. Studies testing even dozens of candidates at once exist for many complex diseases: cardiovascular disease (83), osteoarthritis (84), and asthma (85) to name a few.

A number of methodological problems arise, most notably those of correcting for multiple comparisons and of statistical power. Although most studies are sufficiently powered to detect even modest effects with p-values of 0.01 or greater, if several related phenotypes on dozens of SNPs are analyzed, the number of tests quickly increases to >100. Not all these tests are necessarily independent (the phenotypes can be closely related and the SNPs can be in LD with each other). One way of correcting for multiple tests without an excessive loss of statistical power is to carry out permutation-based tests. For example Valdes et al. (84) derived a list of 24 candidate genes based on the different abundance of the transcripts in cDNA libraries derived from affected (with osteoarthritis) and unaffected tissue. Having genotyped one or more SNPs at each of the candidate genes, a standard association analysis was carried out comparing individuals with and without osteoarthritis. To correct for multiple comparisons a permutation test had to be carried out which could address, given the correlations between phenotypic traits on the one hand, and between genotypes on the other, two questions: (a) what is the empirical (corrected) probability of observing by chance the total number of p-values under 0.05 that are observed in the true set, and (b) what is the probability of observing by chance a SNP associated at a certain significance level with a certain number of phenotypes. Although other tests for multiple comparisons (e.g., Bonferroni and Sidak-Dunn) would be perfectly legitimate, with 156 comparisons to correct for, an initial p-value of 3×10^{-4} would be required. The sample sizes needed to achieve 80% power at such a p-value level assuming an odds ratio of 1.8 and an allele or genotype frequency of 10% are 1200 cases and 1200 controls. This problem becomes even more acute when dealing with genome-wide association scans.

4.5. Genome-Wide Association Studies and Haplotype Blocks

One limitation of candidate gene studies is that the genes must be defined in advance having prior information about their role in the trait being

considered. On the other hand, a genome-wide association scan offers the possibility, without the need to collect family data, to evaluate if regions in the genome are involved in a certain trait and to find novel genes. Carrying out a genome-wide association scan would make it possible to discover genes involved in susceptibility to complex diseases that are not immediately evident given our current knowledge of the metabolic pathways involved.

Escamilla et al. (86), in a study of bipolar mood disorder in an isolated population from Costa Rica, using microsatellite markers spaced ~6 cM intervals across chromosome 18, concluded that LD methods will be useful in this case in a larger sample. The Finnish and Costa Rican populations are considered ideal, since they are relatively homogeneous and show LD over a wider recombination distance than other populations. However, LD is routinely seen for closely linked loci and around disease genes in all populations. With sufficiently closely linked markers, including haplotype level analyses (61,62), association mapping should be a powerful and informative approach in many, and possibly most, populations (57).

Since finding an association depends largely on the underlying genomic patterns of LD in the studied population the NIH Haplotype Map ("HapMap") project was devised. The HapMap project is an extension of the genome sequencing project which aims to characterize patterns of haplotype structure and LD across the human genome to facilitate mapping of complex disease genes (87,88).

If the haplotype structure of the genome reveals blocks of limited haplotype diversity in which most of a global human sample can be characterized by a few common haplotypes this would greatly facilitate LD mapping (50,89). Alleles making up blocks of SNPs in close physical proximity are often correlated and result in reduced genetic variability which define a limited number of "SNP haplotypes." Each of those SNP haplotypes reflects descent from a single, ancient ancestral chromosome. In this scenario, in order to encompass a very large proportion (e.g., 80%) of human variation at a given genomic region it would be necessary only to genotype those SNPs that marked or defined the haplotype blocks in that region and not every SNP [these are termed haplotype tagging SNPs (htSNPs Table 3)]. This would greatly reduce the effort needed while reducing the possibility of missing any genetic variants that could affect the trait under study.

Block-like patterns have been observed within the HLA region (90), in the cytokine cluster on 5q31 (91), and on chromosomes 21 (92) and 22 (93). In one study (90), the blocks were flanked by precisely localized recombination hotspots, leading to suggestions that punctate recombination could be a general phenomenon underlying block structure (87). Extrapolating from one large scale study of 51 genomic regions Gabriel et al. (88) proposed that through careful SNP selection all common haplotypes in the genome could be surveyed typing less than 1 million SNPs.

Table 3 Glossary of Some Terms Commoly Used in Gene Mapping

Term	Abbreviation	Definition
Haplotype		A unique combination of two or more genetic markers present in a chromosome
Haplotype block	Hapblock	A discrete genome region of high LD and low haplotype diversity, hypothesized to be a region of low recombination flanked by hotspots of high recombination
Identity by descent	IBD	Refers to alleles that can be traced back to a common ancestor, in sib-pairs refers to inheritance of the same allele from a given parent
Linkage analysis		Genes that lie on the same chromosome tend to be inherited as a group, a tendency which declines with increasing recombination distance between loci. Linkage analysis makes use of marker segregation within families to identify markers that are physically linked to a gene influencing a trait.
Linkage disequilibrium	LD	Non-random association of alleles at tightly linked loci, i.e., alleles at these loci occur together more (or less) often than would be predicted by chance
Maximum likelihood	ML	A general statistical procedure to estimate one or more parameters (e.g., recombination fraction) of a distribution provided that the distribution is specified.
Penetrance (of a genotype)	—	The probability that a random individual carrying that genotype displays the disease or trait. Incomplete penetrance happens when not all the individuals with the susceptible genotype display the disease
Quantitative trait locus	QTL	A locus (gene) which influences a continuous (also called quantitative) trait such as blood pressure, height, weight, etc
Sibling relative risk	λs	The disease risk for a sibling of an individual affected with disease divided by the risk of disease in the general population
Single nucleotide polymorphism	SNP	Variation in the DNA sequence due to a change at a single nucleotide position
Haplotype tagging SNP	HtSNP	SNP that can define the haplotype block structure within a genomic region

However, Wall and Pritchard's (94) analysis of that same study showed that although recombination "hotspots" do exist, ~50% of the genome cannot be categorized as "block-like." Several authors have therefore stressed the importance of a more dense SNP map to better define both the length and integrity of haplotype blocks (hapblocks) in the human genome (94,95).

Even if the genes influencing a complex disease or trait all fall within haplotype blocks, the haplotype tagging approach will only be effective if the alleles associated with the disease or trait are relatively common. On the contrary, if the alleles associated with a complex disease are very rare (frequency <1%) such variants are likely to be missed using the LD methods so far described. In fact, if a disease is caused by one or only a few genes with many distinct and individually very rare alleles, linkage analysis can perform better than association-based tests (50).

5. ANIMAL MODELS OF DISEASE

When a particular genetic region is implicated in an animal model of disease, both linkage and association studies can focus on the corresponding syntenic chromosomal region in humans (23,24,96), or study of the specific gene if one has been identified. Demonstration of the involvement of the HLA region in type-1 diabetes led to study, and demonstration of a role of, the corresponding major histocompatibility regions (MHCs) in the diabetic mouse and rat. Genome-wide QTL linkage analyses in the mouse have identified genetic regions and genes involved in several multifactorial traits and diseases, leading to the identification of human homologs, for example in hypertension (7). The ability with animal and plant studies to control crosses and in most cases to also study large numbers of offspring makes them very powerful.

Animal models of human disease can also be informative for physiological studies of genetic and environmental factors, and can be used to test novel therapeutics (7,8). For both Mendelian and complex diseases, genes that modify disease severity can also be investigated in animal models by studying animal strains that are severely or mildly affected by the same gene defect, combined with knock-out and knock-in models.

6. DETERMINING THE GENETIC COMPONENTS INVOLVED IN DISEASE

Initial study of the HLA component to type-1 diabetes gave results compatible with a simple recessive model (with incomplete penetrance) (22,97). Only later did the extensive heterogeneity of the HLA component to type-1 diabetes emerge. Genetic heterogeneity was definitively established by the demonstration that the HLA class II DR3 and DR4 serological

associations at the DRB1 locus showed increased risk for DR3/DR4 heterozygotes (98). In a joint collaborative study of Caucasian type-1 diabetes patients and controls, Thomson et al. (99) confirmed further heterogeneity by study of relative predispositional effects (100): after removal of the DR3 and DR4 effects (approximately 93% of patients have 1 or 2 copies of DR3 and/or DR4 compared to 43% of controls), DR2 was shown to be protective, followed by predisposing effects of DR1 and DR8. Cross ethnic studies have now confirmed that the three closely linked class II genes, DRB1, DQA1, and DQB1, all contribute directly to type-1 diabetes (101). A hierarchy of very susceptible, through intermediate, to very protective HLA type-1 diabetes allele, haplotype, and genotype effects are seen, which no current molecular model fully explains (101).

Determining the precise HLA components of disease is difficult for the following primary reasons: multiple genetic factors are involved; the HLA genes are in strong LD; HLA region loci have many polymorphic residues; common HLA alleles may be involved in disease; and interaction effects among HLA, and/or non-HLA, disease loci may be important. These characteristics make necessary a multi-strategy approach, in which complementary methods are used. These include analysis of data from various ethnically or geographically defined groups, as differences in HLA polymorphism and disease prevalence across groups can point toward the precise genetic factors that are important in disease. The methods which have been applied to HLA data to identify the actual disease predisposing factors, and to determine all disease factors in the region, and the difficulties encountered, apply equally to non-HLA regions.

While LD is our ally in detecting genetic regions involved in disease, it confounds attempts to identify the actual gene in the region involved in disease and to identify additional genes in the region contributing to disease. The so-called haplotype method to identify disease-predisposing alleles and amino acids in a genetic region is a stratification analysis to take account of the effects of LD (99,101,102). If all alleles, or amino acid sites, directly involved in disease in a genetic region have been identified, then considering haplotypes (chromosome combinations) containing all these sites, the relative frequencies of alleles and amino acid sites on these haplotypes not involved in disease should be the same in patients and controls. The absolute frequencies of these haplotypes will differ between patients and controls, but the ratio of variants at sites not involved in disease will not differ between patients and controls, provided all alleles or amino acids involved in disease have been identified and all these sites are included in the haplotypes considered. Obviously, the haplotype method cannot distinguish between sites which are very highly correlated in a population, however ethnic comparisons can help identify the predisposing factors. The method however can unequivocally determine if all disease predisposing amino acid sites in LD in a genetic region have *not* been identified.

The original application of the haplotype method was to allele frequency data (99). Direct roles of only HLA DR3 and DR4 serological types in type-1 diabetes were excluded, since HLA B locus variation on these haplotypes was different in patients and controls. At the amino acid level, a combination of sites at DRB1, DQA1, and DQB1 was shown to be highly correlated with the DR-DQ contribution to type-1 diabetes and to correlate with disease incidence (101).

The role of HLA region genes additional to HLA DR-DQ in type-1 diabetes was first demonstrated using affected sib pairs with parents homozygous for the DR3 haplotype (103). The HLA class I B locus data was used to distinguish between the two DR3 haplotypes of the homozygous parent. Under the null hypothesis that no HLA region variation additional to that defined by the DR3 haplotype is involved in type-1 diabetes, the affected sib pairs should share the two parental DR3 haplotypes equally frequently. Significant deviation from 50% sharing was observed. Since the DR3 haplotypes examined in this study could be assumed to be homogeneous for their *DR-DQ* alleles at the molecular level (DRB1∗0301, DQA1∗0501, and DQB1∗0201), this test implicated other HLA region loci in type-1 diabetes.

Using the haplotype method, LD patterns, matching of case/control data for specific DR-DQ combinations, and TDT analysis of heterozygote microsatellite data from parents homozygous for the *DR-DQ* genes, the role of additional HLA region genes in type-1 diabetes has been clearly demonstrated in a number of additional studies. These include other HLA peptide-presenting molecules, such as DPB1∗0301 (104,105), as well as microsatellite associations implicating other regions containing genes involved in the disease process (106,107).

7. THE FUTURE

7.1. Overview

With completion of the human genome project, the functional and positional cloning of Mendelian traits and diseases can become fairly routine (2). Yet, considerable work still lies ahead even with Mendelian diseases: document worldwide variation in mutations, identify the effects of modifier genes on heterogeneity with respect to age of onset and severity, obtain accurate measurements of the penetrance values of different allelic mutations, study the effects of environmental factors on disease expression, and develop appropriate therapies.

For genetically complex human diseases and traits the human genome project will not provide an immediate solution to their genetics, although the ability to identify candidate genes de novo and in a region identified by positional cloning will greatly reduce the time required to home in on

the actual genes involved in complex diseases. A concomitant outcome of the human genome project will be an increase in the characterization of human gene expression patterns, which will greatly aid our studies of the disease process (108,109).

A multi-strategy approach to the mapping of complex diseases and traits is still appropriate: no single method is sufficient or optimal. Genome-wide linkage studies are now routine, although still expensive and time consuming. Despite extensive efforts by many groups, progress in the mapping of complex diseases has been exceedingly slow; only a few genes and some genetic regions involved in complex diseases have been identified. The general picture is one of difficulty in locating disease genes and replication of reported linkages.

The practical application of association studies in Mendelian and complex diseases in fine-mapping of genetic regions identified by linkage analyses has been demonstrated for many, but not all, disease predisposing loci. Genome-wide association studies are now feasible and are starting to be used. As with linkage studies, the overall power of association genome scans to detect disease genes is unknown. Power for both linkage and association studies depends on a number of unknown parameters, will vary across the genome, and also between populations and ethnic groups. The increasing availability of more markers across the genome (50), combined with multi-point analyses using closely linked markers (61,110) will increase the power of linkage and association studies, including the follow-up of regions potentially involved in disease.

When animal models accurately reflect the human disease, these will increasingly be informative. All life forms share many molecular features of gene structure, inheritance and expression, and protein synthesis and function. Given our increasing understanding of the unifying biological aspects of all life forms, including gene expression, protein synthesis and protein function, animal models can be used to identify key genes that influence human multifactorial traits.

Identification of the actual genetic variants involved in disease is not a trivial task for complex diseases. Cross ethnic studies can aid in this venture, as patterns of LD may differ. Animal models also provide valuable information from study of knock-out and knock-in mice and from the mapping of QTLs in inbred strains.

Increasing technological advances in molecular biology, such as the development of SNPs combined with DNA chip technology, and biocomputing will allow screening and analysis of large sample sizes, with predicted future rates of data output which are astronomical. The full development of SNPs will eventually permit routine typing for variation in every human gene, and their regulatory regions, the ultimate association study (9,71,74,75). However, attention must also be given to aspects of genetic epidemiology and population genetics in study design. The issue

of disease heterogeneity, including diagnostic criteria, will be of increasing importance.

7.2. Disease Heterogeneity

Rao (45) emphasized that power should be high in an initial linkage genome-wide scan, to reduce missing linkages (false negatives). The same argument applies to genome-wide association scans (69). Further, all studies of complex diseases should now be seen as exploratory data analyses, without correction for multiple comparisons, or as confirmational studies, as appropriate.

In the context of genome-wide significance levels and false positives, the issue is also raised of whether conditional linkage analyses, stratified by linkage results from established loci such as the confirmed IDDM loci, should be carried out on all genome scan data (24), or whether conditional analyses should be performed only after linkage is established for a specific region (25,110). Given the intrinsic heterogeneous nature of complex disease genetics, stratification analyses with respect to all aspects of disease definition and population heterogeneity should increase our chances of finding the genes involved in disease. Even though type-1 error will be increased (48,110), stratification approaches have proven their worth as an aid in identifying linkages (23,24). The reduction in sample size resulting from stratification by genetic and environmental factors means that very large sample sizes are needed.

The use of a number of different disease phenotype definitions in linkage studies is similarly debated. Success stories based on study of a range of disease definitions (62), subsets of the disease phenotype (111), sex effects (112), and age of onset effects (113–115) support these approaches. The existence of interaction effects between genes involved in complex diseases, as well as parental effects reportedly modifying the transmission and expression of some genes further complicates studies (23). To investigate the possibility of interaction effects between disease loci, linkage evidence at one or more regions can be incorporated into analyses of other positions (116). Standard application of multilocus methods of linkage analysis which are under development should also provide more power (110).

7.3. Sample Sizes, Study Designs

Exceedingly large sample sizes of case/control, simplex, and multiplex family based data, as well as multi-generation pedigrees when available, are needed for our continuing studies of complex diseases. The goal is to achieve a balance between the information content and the ease of obtaining the different types of data points to requiring very large sample sizes of each. Sample sizes in the thousands are required to allow detection and confirmation of linkages, association genomic level as well as fine mapping,

stratification analyses based on various disease phenotypes, including complications, age of onset, genetic, and epidemiological parameters, as well as analyses to determine if the true predisposing factors, and all factors in a genetic region, have been identified.

It was previously argued that detailed study of environmental factors on disease would best be studied once the genetic basis of the disease was established. In this way genetically predisposed individuals could be identified and the impact of environmental factors more clearly identified. However, given the difficulty in identifying the genetic components of complex diseases, it now seems appropriate that we concentrate on environmental factors in conjunction with genetic studies, as the effect of environmental factors on many complex diseases is quite strong.

Provided the patient and control groups are carefully matched for ethnicity, population stratification effects creating spurious associations are eliminated. The large collections of multiplex families now available for linkage studies in many complex diseases are obviously also a valuable resource for association screens. Further, they allow for study of combined linkage and association results (117,118). This can yield strong evidence for linkage when the affected sib pairs are analyzed based on the presence or not of the associated allele in the proband, e.g., with type-2 diabetes, where the overall evidence of linkage is very weak (119). Further, the direct role of a putative disease predisposing gene can be tested (118), and interactive effects of candidate genes studied (119).

Our studies must also include many different populations, including extensive ethnic variation, and not be restricted to relatively homogeneous populations. Only with this approach can we identify all the genes involved, in all populations, in predisposition to and protection from disease, including their interaction effects and their influence on response to environmental factors. The genes involved in multi-generation pedigrees may often be different from those in affected sib pair and/or "sporadic" cases of disease, so both types of pedigrees must be studied.

National and international cooperative efforts for sharing data are mandatory to achieve the large sample sizes required and allow meta-analyses of data (120,121), which can be very powerful. Study of genetic effects common to multiple diseases will also increasingly be of considerable interest (121,122). The most efficient scheme for completing a pooled association screen is for a large collaborative study of several diseases with division of microsatellites among laboratories (65).

Further study of population level data is also obligatory (70,89,95,123), including development of methods to understand the evolutionary history of a region (10,70,124). Only then can the power of different linkage and association methods be completely assessed and the processes understood by which disease predisposing variants become established, and often frequent, in populations.

REFERENCES

1. Collins FS, McKusick VA. Implications of the human genome project for medical science. JAMA 2001; 285(5):540–544.
2. Gelehrter TD, Collins FS, Ginsberg D. Principles of Medical Genetics. Baltimore: Williams and Wilkins, 1998.
3. King MC, Lee GM, Spinner NB, Thomson G, Wrensch MR. Genetic epidemiology. Ann Rev Public Health 1984; 5:1–52.
4. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet 2002; 3(11):872–882.
5. Thomson G, Esposito MS. The genetics of complex diseases: an overview. Trends Genet 1999; 15:M17–M20.
6. Thorsby E. Invited anniversary review: HLA associated diseases. Hum Immunol 1997; 53:1–11.
7. Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994; 265:2037–2048.
8. Risch NJ. Searching for genetic determinants in the new millenium. Nature 2000; 405:847–856.
9. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 1998; 8:1229–1231.
10. Huttley GA, Smith MW, Carrington M, O'Brien SJ. A scan for linkage disequilibrium across the human genome. Genetics 1999; 152:1711–1722.
11. Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M. Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. Am J Hum Genet 1994; 54:884–898.
12. Klitz W, Stephens JC, Grote M, Carrington M. Discordant patterns of linkage disequilibrium of the peptide—transporter loci within the HLA class II region. Am J Hum Genet 1995; 57:1436–1444.
13. Hedrick PW. Genetics of Populations. Boston: Science Books International, 1983.
14. Thomson G. The effect of a selected locus on linked neutral loci. Genetics 1977; 85:753–788.
15. Slatkin M. Linkage disequilibrium in growing and stable populations. Genetics 1994; 137:331–336.
16. Ott J. Analysis of human genetic linkage. 3d ed. Baltimore: John Hopkins University Press, 1999.
17. Elston RC. Linkage and association. Genet Epidemiol 1998; 15:565–576.
18. Elston RC. Methods of linkage analysis—and the assumptions underlying them. Am J Hum Genet 1998; 63:931–934.
19. Thomson G. Mapping disease genes: family based association studies. Am J Hum Genet 1995; 57:487–498.
20. Thomson G. Analysis of complex human genetic traits: an ordered-notation method and new tests for mode of inheritance. Am J Hum Genet 1995; 57:474–486.
21. Cudworth AG, Woodrow JC. Evidence for HL-A-linked genes in "juvenile" diabetes mellitus. Br Med J 1975; 3:133–135.

22. Payami H, Thomson G, Motro U, Louis EJ, Hudes E. The affected sib method IV. Sib trios. Ann Hum Genet 1985; 49:303–314.
23. Pugliese A. Unraveling the genetics of insulin-dependent type 1A diabetes: the search must go on. Diabetes Rev 1999; 7:39–54.
24. Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ, Pritchard LE, Cucca F, Barnett AH, Bain SC, Todd JA. A search for type 1 diabetes susceptibility genes in families from the UK. Nat Genet 1998; 19:297–300.
25. Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst B, Morrison VA, Stirling B, Mitra M, Farmer J, Williams SR, Cox NJ, Bell GI, Risch N, Spielman RS. A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. Nat Genet 1998; 19:292–296.
26. Payami H, Louis EJ, Klitz W, Lo SK, Thomson G. Family and population analysis of multiple sclerosis. Genet Epidemiol 1986(suppl 1):381–386.
27. Payami H, Thomson G, Khan MA, Grennan DM, Sanders P, Dyer P, Dostal C. Genetics of rheumatoid arthritis. Tissue Antigens 1986; 27:57–63.
28. Thomson G. A two locus model for juvenile diabetes. Ann Hum Genet 1980; 43:383–398.
29. Rotter JI, Landaw EM. Measuring the genetic contribution of a single locus to a multilocus disease. Clin Genet 1984; 26:529–542.
30. Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 1987; 40:1–14.
31. Noble J, Valdes AM, Cook M, Klitz W, Thomson G, Erlich H. The role of HLA class II genes in insulin-dependent diabetes mellitus (IDDM): molecular analysis of 180 Caucasian, multiplex families. Am J Hum Genet 1996; 59:1134–1148.
32. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Farrall M, Barnett AH, Bain SC, Todd JA. A genome-wide search for human type 1 diabetes susceptibility genes. Nature 1994; 371:130–136.
33. Hashimoto L, Habita C, Beressi JP, Delepine M, Besse C, Cambon-Thomsen A, Deschamps I, Rotter JI, Djoulah S, James MR, Froguel P, Weissenbach J, Lathrop GM, Julier C. Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. Nature 1994; 371:161–164.
34. Field LL, Tobias R, Magnus T. A locus on chromosome 15q26 (IDDM3) produces susceptibility to insulin—dependent diabetes mellitus. Nat Genet 1994; 8:189–194.
35. Luo DF, Buzzetti R, Rotter JI, Maclaren NK, Raffel LJ, Nistico L, Giovannini C, Pozzilli P, Thomson G, She JX. Confirmation of three susceptibility genes to insulin-dependent diabetes mellitus: IDDM4, IDDM5 and IDDM8. Hum Mol Genet 1996; 5:693–698.
36. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995; 11:241–247.
37. Reed P, Cucca F, Jenkins S, Merriman M, Wilson A, McKinney P, Bosi E, Joner G, Ronningen K, Thorsby E, Undlien D, Merriman T, Barnett A Bain S, Todd J. Evidence for a type 1 diabetes susceptibility locus (IDDM10) on human chromosome 10p11–q11. Hum Mol Genet 1997; 6:1011–1016.

38. Merriman T, Twells R, Merriman M, Eaves I, Cox R, Cucca F, McKinney P, Shield J, Baum D, Bosi E, Pozzilli P, Nistico L, Buzzetti R, Joner G, Ronningen K, Thorsby E, Undlien D, Pociot F, Nerup J, Bain S, Barnett A, Todd J. Evidence by allelic association-dependent methods for a type 1 diabetes polygene (IDDM6) on chromosome 18q21. Hum Mol Genet 1997; 6:1003–1010.
39. Esposito L, Hill NJ, Pritchard LE, Cucca F, Muxworthy C, Merriman ME, Wilson A, Julier C, Delepine M, Tuomilehto J, Tuomilehto-Wolf E, Ionesco-Tirgoviste C, Nistico L, Buzzetti R, Pozzilli P, Ferrari M, Bosi E, Pociot F, Nerup J, Bain SC, Todd JA. Genetic analysis of chromosome 2 in type 1 diabetes: analysis of putative loci IDDM7, IDDM12, and IDDM13 and candidate genes NRAMP1 and IA-2 and the interleukin-1 gene cluster. IMDIAB Group. Diabetes 1998; 47:1797–1799.
40. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 1994; 54(3):535–543.
41. Blangero J, Williams JT, Almasy L. Variance component methods for detecting complex trait loci. Adv Genet 2001; 42:151–181.
42. de Andrade M, Amos CI, Thiel TJ. Methods to estimate genetic components of variance for quantitative traits in family studies. Genet Epidemiol 1999; 17(1):64–76.
43. Suarez BK, Hampe CL, Van Eerdewegh P. Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR, eds. Genetic Approaches to Mental Disorders. Washington, DC: American Psychiatry Press, 1994:23–46.
44. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273:1516–1517.
45. Rao DC. CAT scans, PET scans, and genomic scans. Genet Epidemiol 1998; 15:1–18.
46. Witte JS, Elston RC, Schork NJ. Genetic dissection of complex traits. Nat Genet 1996; 12:355–356.
47. Curtis D. Genetic dissection of complex traits. Nat Genet 1996; 12:356–358.
48. Todorov AA, Rao DC. Trade-off between false positives and false negatives in the linkage analysis of complex traits. Genet Epidemiol 1997; 14:453–464.
49. Rao DC, Gu C. False positives and false negatives in genomic scans. In: Rao DC, Province MA, eds. Genetic Dissection of Complex Traits: Challenges for the Next Millenium. San Diego: Academic Press, 2001:487–498.
50. Clark AG. Finding genes underlying risk of complex disease by linkage disequilibrium mapping. Curr Opin Genet Dev 2003; 13(3):296–302.
51. Broeckel U, Schork NJ. Identifying genes and genetic variation underlying human diseases and complex phenotypes via recombination mapping. J Physiol 2004; 554(pt 1):40–45.
52. Schellenberg GD, D'Souza I, Poorkaj P. The genetics of Alzheimer's disease. Curr Psychiatry Rep 2000; 2(2):158–164.
53. Cox NJ, Hayes MG, Roe CA, Tsuchiya T, Bell GI. Linkage of calpain 10 to type 2 diabetes: the biological rationale. Diabetes 2004; 53(suppl 1):S19–S25.
54. Fulker DW, Cherny SS, Sham PC, Hewitt JK. Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 1999; 64(1):259–267.

55. Sham PC, Cherny SS, Purcell S, Hewitt JK. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 2000; 66(5):1616–1630.
56. Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 1998; 62(2):450–458.
57. Morton NE, Collins A. Tests and estimates of allelic association in complex inheritance. Proc Natl Acad Sci USA 1998; 95:11389–11393.
58. Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Kawaguchi Y, Dronsfield MJ, Pociot F, Nerup J, Bouzekri N, Cambon-Thomsen A, Ronningen KS, Barnett AH, Bain SC, Todd JA. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 1995; 9:284–292.
59. Spielman RS, Baur MP, Clerget-Darpoux F. Genetic analysis of IDDM: summary of GAW 5 IDDM results. Genet Epidemiol 1989; 6:43–58.
60. Sawcer S, Goodfellow PN, Compston A. The genetic analysis of multiple sclerosis. Trends Genet 1997; 13:234–239.
61. Barcellos LF, Thomson G, Carrington M, Schafer J, Begovich AB, Lin P, Xu XH, Min BQ, Marti D, Klitz W. Chromosome 19 single-locus and multilocus haplotype associations with multiple sclerosis: evidence of a new susceptibility locus in Caucasian and Chinese patients. JAMA 1997; 278:1256–1261.
62. Valdes AM, McWeeney SK, Thomson G. Evidence for linkage and association to alcohol dependence on chromosome 19. Genet Epidemiol 1999; 17(suppl 1):S367–S372.
63. Arnheim N, Strange C, Erlich H. Appendix by Thomson G. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proc Natl Acad Sci USA 1985; 82:6970–6974.
64. Pacek P, Sajantila A, Syvanen AC. Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. PCR Methods Appl 1993; 2:313–317.
65. Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, Nelson MP, Nagatomi J, Thomson G. Association mapping of disease loci, by use of a pooled DNA genomic screen. Am J Hum Genet 1997; 61:734–747.
66. Shaw SH, Carrasquillo MM, Kashuk C, Puffenberger EG, Chakravarti A. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. Genome Res 1998; 8:111–123.
67. Germer S, Holland MJ, Higuchi R. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. Genome Res: Genome Res 2000; 10(2):258–266.
68. Mohlke KL, Erdos MR, Scott LJ, Fingerlin TE, Jackson AU, Silander K, Hollstein P, Boehnke M, Collins FS. High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. Proc Natl Acad Sci USA 2002; 99:16928–16933.

69. Long AD, Langley CH. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res 1999; 9:720–731.
70. Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. Haplotype structure and population genetic inferences from nucleotide—sequence variation in human lipoprotein lipase. Am J Hum Genet 1998; 63:595–612.
71. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 1998; 19: 233–240.
72. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999; 22:139–144.
73. Lonjou C, Collins A, Morton NE. Allelic association between marker loci. Proc Natl Acad Sci USA 1999; 96:1621–1626.
74. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalayanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 1999; 22:231–238.
75. Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 1999; 22:239–247.
76. Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 1987; 51:227–233.
77. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993; 52:506–516.
78. Monks SA, Kaplan NL, Weir BS. A comparative study of sibship tests of linkage and/or association. Am J Hum Genet 1998; 63:1507–1516.
79. Martin ER, Kaplan NL, Weir BS. Tests for linkage and association in nuclear families. Am J Hum Genet 1997; 61:439–448.
80. Risch N, Teng J. The relative power of family based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. Genome Res 1998; 8:1273–1288.
81. Thomson G. Strategies involved in mapping diabetes genes: an overview. Diabetes Rev 1997; 5:106–115.
82. Ertekin-Taner N, Ronald J, Asahara H, Younkin L, Hella M, Jain S, Gnida E, Younkin S, Fadale D, Ohyagi Y, Singleton A, Scanlin L, de Andrade M, Petersen R, Graff-Radford N, Hutton M, Younkin S. Fine mapping of the alpha-T catenin gene to a quantitative trait locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. Hum Mol Genet 2003; 12(23):3133–3143.
83. Zee RY, Hoh J, Cheng S, Reynolds R, MA Grow, Silbergleit A, Walker K, Steiner L, Zangenberg G, Fernandez-Ortiz A, Macaya C, Pintor E, Fernandez-Cruz A, Ott J, Lindpainter K. Multi-locus interactions predict risk

for post-PTCA restenosis: an approach to the genetic analysis of common complex disease. Pharmacogenomics J 2002; 2(3):197–201.
84. Valdes AM, Hart DJ, Jones KA, Surdulescu G, Swarbrick P, Doyle DV, Schafer AJ, Spector TD. Candidate gene association study of the prevalence and progression of knee osteoarthritis. Arthritis and Rheumatism 2004; 50: 2497–2507.
85. Immervoll T, Loesgen S, Dutsch G, Gohlke H, Herbon N, Klugbauer S, Dempfle A, Bickeboller H, Becker-Follmann J, Ruschendorf F, Saar K, Reis A, Wichmann HE, Wjst M. Fine mapping and single nucleotide polymorphism association results of candidate genes for asthma and related phenotypes. Hum Mutat 2001; 18(4):327–336.
86. Escamilla MA, McInnes LA, Spesny M, Reus VI, Service SK, Shimayoshi N, Tyler DJ, Silva S, Molina J, Gallegos A, Meza L, Cruz ML, Batki S, Vinogradov S, Neylan T, Nguyen JB, Fournier E, Araya C, Barondes SH, Leon P, Sandkuijl LA, Freimer NB. Assessing the feasibility of linkage disequilibrium methods for mapping complex traits: an initial screen for bipolar disorder loci on chromosome 18. Am J Hum Genet 1999; 64: 1670–1678.
87. Couzin J. New mapping project splits the community. Science 2002; 296: 1391–1393.
88. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science 2002; 296: 2225–2229.
89. Cardon LR, Abecassis GR. Using haplotype blocks to map human complex trait loci. Trends Genet 2003; 19(3):135–140.
90. Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 2001; 29:217–222.
91. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet 2001; 29(2):229–232.
92. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 2001; 294:1719–1723.
93. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Lohmussaar E, Zernant J, Tonisson N, Remm M, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott A, Metspalu M, Bentley G, Cardon LR, I Dunham I. A first-generation linkage disequilibrium map of human chromosome 22. Nature 2002; 418:544–548.
94. Wall JD, Pritchard JK. Assessing the performance of the haplotype block model of linkage disequilibrium. Am J Hum Genet 2003; 73:502–515.

95. Tishkoff SA, Verrelli BC. Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. Curr Opin Genet Dev 2003; 13:569–575.
96. Morahan G, Huang D, Tait BD, Colman PG, Harrison LC. Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. Science 1996; 272:1811–1813.
97. Thomson G, Bodmer WF. The genetics of HLA and disease associations. In: Christiansen FB, Fenchel T, Barndorff-Nielson O, eds. Measuring Selection in Natural Populations. Heidleberg: Springer-Verlag, 1977:545–564.
98. Svejgaard A, Platz P, Ryder LP. Insulin-dependent diabetes mellitus. In: Terasaki PI, ed. Histocompatibility 1980. Los Angeles: University of California Press, 1980:638–656.
99. Thomson G, Robinson WP, Kuhner MK, Joe S, MacDonald MJ, Gottschall JL, Barbosa J, Rich SS, Bertrams J, Baur MP, Partanen J, Tait B, Schober E, Mayr WR, Ludvigsson J, Lindblom B, Farid NR, Thompson C, Deschamps I. Genetic heterogeneity, modes of inheritance and risk estimates for a joint study of Caucasians with insulin dependent diabetes mellitus. Am J Hum Genet 1988; 43:799–816.
100. Payami H, Joe S, Farid NR, Stenszky V, Chan SH, Yeo PPB, Chan JS, Thomson G. Relative predispositional effects (RPEs) of marker alleles with disease: HLA-DR alleles and Graves disease. Am J Hum Genet 1989; 45: 541–546.
101. Valdes AM, McWeeney S, Thomson G. HLA class II DR/DQ amino acids and insulin-dependent diabetes mellitus: application of the haplotype method. Am J Hum Genet 1997; 60:717–728.
102. Valdes AM, Thomson G. Detecting disease-predisposing variants: the haplotype method. Am J Hum Genet 1997; 60:703–716.
103. Robinson WP, Thomson G, Rich SS, Barbosa J. The homozygous parent affected sib pair method for detecting disease predisposing variants: application to insulin dependent diabetes mellitus. Gen Epidemiol 1993; 10:273–288.
104. Erlich HA, Rotter JI, Chang JD, Shaw SJ, Raffel LJ, Klitz W, Bugawan TL, Zeidler A. Association of HLA-DPB1*0301 with IDDM in Mexican-Americans. Diabetes 1996; 45:610–614.
105. Noble JA, Valdes AM, Thomson G, Erlich HA. The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. Diabetes 2000; 49(1):121–125.
106. Hadifi Moghaddam P, deKnijf P, Roep BO, VanderAuwera B, Naipal A, Gorus F, Schuit F, Giphart MJ. Genetic structure of IDDM1: two separate regions in the major histocompatibility complex contribute to susceptibility or protection. Diabetes 1998; 47:263–269.
107. Lie B, Todd J, Pociot F, Nerup J, Akselsen H, Joner G, Dahl-Jorgensen K, Ronningen K, Thorsby E, Undlien D. The predisposition to type 1 diabetes linked to the HLA complex (IDDM1) includes at least one non-class II gene. Am J Hum Genet 1999; 64:793–800.
108. Weeraratna AT, Nagel JE, Mello-Coelho Vd V, Taub DD. Gene expression profiling: from microarrays to medicine. J Clin Immunol 2004; 24(3):213–224.
109. Knight JC. Allele-specific gene expression uncovered. Trends Genet 2004; 20(3):113–116.

110. Lernmark A, Ott J. Sometimes it's hot, sometimes it's not. Nat Genet 1998; 19:213–214.
111. Gibbs M, Stanford JL, McIndoe RA, Jarvik GP, Kolb S, Goode EL, Chakrabarti L, Schuster EF, Buckley VA, Miller EL, Brandzel S, Li S, Hood L, Ostrander EA. Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. Am J Hum Genet 1999; 64:776–787.
112. Paterson AD, Petronis A. Sex of affected sibpairs and genetic linkage to type 1 diabetes. Am J Med Genet 1999; 84:15–19.
113. Day EW, Heath SC, Wijsman EM. Multipoint oligogenic analysis of age-at-onset data with applications to Alzheimer disease pedigrees. Am J Hum Genet 1999; 64:839–851.
114. Valdes AM, Thomson G, Erlich HA, Noble JA. Association between IDDM age of onset and HLA among sib pairs. Diabetes 1999; 48:1658–1661.
115. Li H, Hsu L. Effects of age at onset on the power of the affected sib pair and transmission/disequilibrium tests. Ann Hum Genet 2000; 64:239–254.
116. Cox NJ, Frigge M, Nicolae DL, Concannon P, Hanis CL, Bell GI, Kong A. Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. Nat Genet 1999; 21:213–215.
117. Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J. A new method to test genetic models in HLA associated diseases: the MASC method. Ann Hum Genet 1988; 52:247–258.
118. Clerget-Darpoux F, Babron MC, Deschamps I, Hors J. Complementation and maternal effect in insulin-dependent diabetes. Am J Hum Genet 1991; 49: 42–48.
119. Dizier MH, Babron MC, Clerget-Darpoux F. Interactive effect of two candidate genes in a disease: extension of the marker-associaton-segregation X^2 method. Am J Hum Genet 1994; 55:1042–1049.
120. Gu C, Province M, Todorov A, Rao DC. Meta-analysis methodology for combining non-parametric sibpair linkage results: genetic homogeneity and identical markers. Genet Epidemiol 1998; 15:609–626.
121. Wise LH, Lanchbury JS, Lewis CM. Meta-analysis of genome searches. Ann Hum Genet 1999; 63:263–272.
122. Becker KG, Simon RM, Bailey-Wilson JE, Freidlin B, Biddison WE, McFarland HF, Trent JM. Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. Proc Natl Acad Sci USA 1998; 95:9979–9984.
123. Anderson EC, Slakin M. Population-genetic basis of haplotype blocks in the 5q31 region. Am J Hum Genet 2004; 74:40–49.
124. Grote M, Klitz W, Thomson G. Constrained disequilibrium values and hitchhiking in a three-locus system. Genetics 1998; 150:1295–1307.

24

Positional Cloning and Disease Gene Identification

Richard F. Wintle
Ellipsis Biotherapeutics Corporation, Toronto, Ontario, Canada
Katherine A. Siminovitch
Department of Medicine, University of Toronto, Toronto, Ontario, Canada

1. INTRODUCTION

Advances in recombinant DNA technologies and the sequence data provided by the human genome project have provided an unprecedented opportunity to characterize the molecular pathophysiology of disease. It is now evident that the majority of human diseases arise as a result of the interplay between environmental factors and genetic background. Genetic and environmental interactions play an essential role not only in disease predisposition, but also in the modulation of disease characteristics (such as age at onset and expression of various subphenotypes), therapeutic responsiveness, and outcome. Recently, the elucidation of these complex interactions has been greatly facilitated by the completion of the human genome project and publication of the full genome sequence, currently estimated to encompass about 22,000 protein coding genes (1). The major challenge now facing medical science is to exploit this genetic information so as to identify the sets of genes responsible for specific illnesses and thereby unravel the molecular processes coupling such genetic determinants to expression and persistence of disease (2). Such knowledge will allow for definition of molecular biomarkers for use in diagnosis and prediction of disease risk, prognosis and drug

responsiveness, and will thereby enable ''individualized'' medical care wherein both diagnosis and therapy are predicated on the individual's profile of susceptibility gene variants. Thus, definition of the full complement of susceptibility alleles involved in common diseases will pave the way for earlier, more precise diagnosis and for developing the pharmacogenomic knowledge required to optimize drug therapy (3).

In view of the extraordinary potential for disease gene identification to improve health care delivery, a major thrust of human genetics research has been directed at developing technologies, which accelerate the disease gene discovery process. For many single-gene diseases, knowledge of the pathophysiology has been sufficient to allow for direct screening of newly cloned genes as possible disease gene candidates. This ''candidate gene'' approach, which builds upon understanding of the disease pathophysiology and a given gene/protein function, has, for example, been used to characterize the molecular basis for phenylketonuria, β-thalassemia, and many other single-gene disorders (4–8).

For the majority of human diseases, however, information on etiopathogenesis is insufficient to render candidate gene testing feasible. This situation is true for most single-gene diseases and for virtually all of the common human diseases having a significant genetic or inherited component. These latter conditions are referred to as ''complex genetic'' diseases, conditions which collectively account for the majority of the healthcare burden in developed countries and which include, for example, atherosclerotic cardiovascular disease, rheumatoid arthritis, and diabetes. Discovery of genes involved in complex genetic diseases is further confounded by many other issues such as lack of knowledge of mode of inheritance (dominant, recessive, multifactorial, etc.), incomplete penetrance, genetic and allelic heterogeneity (9), and the fact that many disease-associated genetic variants are likely to occur commonly in the general (healthy) population (10). Additionally, until relatively recently, most of the genome had remained uncharacterized. For these and other reasons, most of the initial work on disease gene identification utilized an alternative gene discovery approach, ''positional cloning,'' wherein disease genes are isolated based on their chromosomal localization (11). Positional cloning, a strategy allowing disease gene discovery despite lack of knowledge of disease pathophysiology or mode of inheritance, has proven extraordinarily successful to gene discovery in single-gene disease and, more recently, also in relation to uncovering genes involved in complex genetic disease. While the steps involved in positional cloning and candidate gene analysis are initially distinct, positional cloning approaches lead ultimately to the identification of sets of candidate genes at specific locations in the genome. Thus, these two methodologies converge and the challenge becomes identifying which one of a (generally small) number of ''good candidate'' genes represents the gene responsible for the disease under study. In this chapter, these technologies and challenges are described and

illustrated using the specific example of the inflammatory bowel disease, Crohn's disease.

2. CANDIDATE GENE APPROACHES

Detailed understanding of disease pathogenesis provides a knowledge framework for predicting the genes likely contributing to disease susceptibility. The relevance of such "candidate" genes to a specific disease can be directly established by demonstrating that the disease state is associated with variants (alleles) of the gene(s) in question. Candidate gene analysis involves investigation of fully characterized genes encoding putative disease-relevant products (e.g., tumor suppressor genes in cancer, cytokines in inflammatory disease, etc.). Alternatively, candidate gene analysis may require the isolation of a gene encoding a protein that seems likely to be related to disease, for example by using oligonucleotide-based probes or antibodies to screen cDNA or expression libraries. This latter strategy was used, for example, to isolate the factor VIII gene involved in Haemophilia A (4), the globin genes involved in thalassemia and sickle cell anemia (8) and one of the genes responsible for Fanconi anemia (12). However, as the specific pathophysiological processes and proteins underlying the majority of human diseases are unknown, this approach has limited scope for application.

As gene cloning has progressed and the human genome has been increasingly well characterized, candidate gene investigation has largely shifted from the direct cloning approach to the screening of already-identified genes for their potential relevance to disease. Although this approach involves some degree of "educated" guessing, with candidates being selected on the premise of relevance to disease etiology, the candidate gene strategy has proven successful in many instances. Identification of the genetic defect underlying Marfan's syndrome, for example, was based on the realization that a newly cloned gene encoded a protein, fibrillin, with functional properties potentially relevant to the cellular defects found in Marfan's patients (13,14). Genes encoding the human leukocyte antigens (HLA) have also been investigated as candidate susceptibility genes for a broad spectrum of diseases, including autoimmune and other immunological disorders such as asthma, inflammatory bowel disease, multiple sclerosis, type 1 diabetes mellitus (T1D), and psoriasis (15). Indeed, certain variants of the HLA genes have been shown to contribute to a significant proportion of the genetic risk for T1D, rheumatoid arthritis (16), and inflammatory bowel disease (17). As the functions of the full complement of human genes become increasingly well characterized, opportunities for disease gene identification by candidate gene analysis will continue to grow and likely permit discovery of many important disease susceptibility genes.

3. POSITIONAL CLONING

In contrast to the candidate gene approach, positional cloning involves disease gene identification in the absence of assumptions as to the functional properties of the relevant genes (11). Thus, at least in its initial stages, this approach does not require prior knowledge of the disease pathophysiology. The first and most well known example of disease gene discovery by positional cloning was the identification of the gene (CFTR) responsible for cystic fibrosis (18–20), a discovery involving the use of physical and genetic mapping methods to "home in" on the chromosome 7 region harboring the gene mutation underlying this disease.

The basic strategy used in positional cloning involves definition of chromosomal region(s) containing the genes of interest and the subsequent analysis of the genes within such regions for disease-associated mutations. This approach involves screening the >3 billion bases of the haploid human genome for as minor a change as a single-base alteration, a magnitude of task that can be likened to the challenge associated with finding a specific individual person on this planet with no prior knowledge as to his/her address. However, despite the enormity of the challenge, many genes associated with single-gene disorders have been identified by positional cloning. These successes have validated positional cloning as a valuable strategy for disease gene isolation and have provided incentive for the field to progress toward the search for genes underlying the chronic multifactorial disorders constituting the bulk of human disease (21).

4. CLINICAL MATERIALS

Regardless of the approach taken, disease gene identification generally involves a series of discrete steps, the earliest and often rate limiting of which is the ascertainment and clinical characterization of affected individuals. The importance of patient collection cannot be overstated, as the quality of the phenotypic data collected in the population under study has immense impact on the results and interpretation of the data obtained from genetic analysis. In contrast to candidate gene association analysis, positional cloning involves the analysis of families as well as affected individuals. Multigenerational pedigrees including multiple affected individuals such as affected relative pairs and affected sib pairs, as well as parents and ideally, grandparents, are most useful for such studies, but are often difficult to obtain, particularly for diseases where age of onset is late in life. Other unaffected relatives may also be included in order to "link" affected individuals in the pedigree. Genetic material, usually genomic DNA extracted from a peripheral blood sample, is collected from the affected and unaffected study participants and clinical records obtained and reviewed so as to ensure the affectation status of all subjects. In more sophisticated studies, the affectation

status may be denoted as "unknown," if a subject is young enough to be pre-symptomatic for a late-onset disease. The importance of establishing stringent diagnostic criteria for the phenotype of interest is crucial as the failure to do so may obfuscate subsequently obtained linkage data (22). These clinical data should, in turn, be formatted and stored in databases in a fashion that facilitates easy retrieval and flexible queries, so as to enable multivariate analyses, stratification of the population based on various demographic/clinical characteristics and definition of subpopulations treated with and/or responding to specific medications (3).

5. FAMILY-BASED LINKAGE ANALYSIS

Following completion of family collection, genotyping studies can be initiated so as to enable linkage-based disease gene localization. Linkage refers to the cosegregation of loci in close proximity on the same chromosome. By contrast, loci which map to different chromosomes or far apart on the same chromosome segregate independently due to meiotic recombination and therefore show no linkage. The greater the (physical) distance between two loci, the higher the likelihood that these loci will be separated by recombination events, and thus show a greater interlocus recombination fraction. By contrast, loci mapping close to each other are unlikely to be separated by recombination and are referred to as "linked." Importantly, "linkage" implies a physical relationship between genetic loci and needs to be distinguished from the concept of "association," the latter of which refers to the coincident presence of a specific genetic variant and a given disease, but which does not necessarily imply physical proximity between the disease gene and the marker locus. A number of different approaches to determining linkage between disease and DNA markers are commonly used, including both parametric and non-parametric approaches (23–25).

6. GENETIC MARKERS FOR LINKAGE OR ASSOCIATION ANALYSIS

In recognition of the potential for linkage to localize and ultimately reveal novel disease susceptibility genes, enormous effort has been directed at identifying genetic markers that are sufficiently polymorphic and/or sufficiently frequent in the genome to be of value in linkage studies. Older marker typing technologies such as restriction fragment length polymorphism (RFLPs) (26) and variable number of tandem repeats (VNTR) marker (27) testing have since given way to the two most frequently used genotyping tools: short tandem repeats (STRs), also commonly called "microsatellites," and single-nucleotide polymorphisms (SNPs). The SNPs also form the basis of newer mapping approaches such as haplotype-tag mapping, a strategy underpinning the international haplotype mapping effort (www.hapmap.org) directed at

delineating all haplotype "blocks" across the genome and sets of SNPs representative of each block (28).

The STRs represent regions of the genome containing stretches of tandem repeats of about two to seven nucleotides in length and occur frequently throughout the genome. The STRs are generally highly polymorphic and can be detected by PCR amplification using primer pairs corresponding to conserved STR flanking sequences. Detection is typically via size-separation on fluorescent capillary or slab gel DNA sequencers and automated software is used to obtain genotype "calls," i.e. to identify the different sized alleles from either chromosome of an individual DNA sample. The vast majority of genome-wide scans for disease susceptibility loci use STR-based approaches, although SNP-based approaches are now becoming feasible and are gaining in popularity (29). Most STR-based scans involve the use of approximately 300 di-, tri-, or tetranucleotide repeat markers spaced at about 10-cM intervals, so as to provide coverage of the approximately 3300 cM genome. Standardized primer sets for such genome-wide scans are readily available from commercial vendors.

While STR markers have proven highly useful for disease gene mapping, the level of resolution provided by these markers is sufficient only for regional localizations of genes involved in complex disease (i.e. refinement to, at best, perhaps 1 cM, or roughly 1 million base pairs of DNA). Typically, such regions are defined through family-based linkage analysis, although linkage disequilibrium (LD) mapping using STRs can also be very successful in many regions of the genome (30). Indeed, our own search for inflammatory bowel disease susceptibility genes involved an initial standard STR-based genome-wide linkage scan approach (31), followed by LD mapping with microsatellite STRs to refine the Crohn's disease-associated *IBD5* locus to a small region of chromosome 5q31 (32). Regions of this magnitude in size are expected to contain multiple genes, perhaps even dozens in some particularly gene-rich areas of the genome. In this context, SNPs, the most common form of genetic variation, are useful for refining regions and identifying specific disease-associated genes (33). The SNPs represent sites in which single-base pair variation occurs from person-to-person and in which the least frequent allele has a population frequency of 1% or greater. In contrast to STRs, SNPs occur at an average frequency of about 1 per 1000 bp and are generally mutationally stable, facilitating LD mapping strategies which rely upon co-segregation of ancestral alleles that are in close physical proximity to each other. Many SNPs occur within genes, and accordingly, some will represent the genetic lesions of interest (e.g. a missense or nonsense mutation in a coding region). By contrast, the utility of SNPs is reduced by their biallelic nature, meaning that heterozygosity cannot be greater than 50% and thus more SNPs than STRs are required to provide adequate information for disease gene discovery. Estimates of the numbers of SNPs required for genome-wide genetic association studies have been as

high as 0.5–1.0 million (34), and even a recent estimate suggests that no fewer than 200,000 SNPs may be required to achieve this goal (35). This problem has been addressed in part with the development of microarray-based SNP genotyping approaches incorporating greater than 100,000 SNPs per array (Affymetrix, Santa Clara, CA), or through other very high-throughput genotyping platforms such as the Illumina BeadArray system (Illumina Inc., San Diego, CA). New approaches such as single-molecule sequencing using nanoscale fluidics may also further optimize such studies (36). These and other technologies, coupled to robotic systems for sample and liquid handling, should render high-throughput SNP genotyping increasingly feasible and, in conjunction with worldwide efforts to identify and validate all common SNP variants in the human genome (37), enable the scale of whole-genome association studies to be ramped up to the level required for disease gene mapping in complex genetic diseases.

7. USE OF FAMILIES FOR POSITIONAL CLONING

In conjunction with polymorphic markers spanning the genome, family collections represent a cornerstone of gene mapping studies. The choice of family structures to be used in positional cloning is very much influenced by the disease under study. For example, large pedigrees including multiple affected individuals are particularly relevant to the mapping of single-gene disorders in which the pattern of inheritance is known and parametric linkage methods (those including parameters specifying mode of inheritance, penetrance, etc.) thus easily applied. By contrast, in instances in which the mode of inheritance is unknown, as is the case for most complex genetic diseases, a more widely used family structure is the affected sibling pair (ASP), which can be studied by non-parametric methods that assess sharing of chromosomal regions by the affected individuals. The ASP analysis is predicated upon the fact that siblings are expected to share 0, 1, or 2 parental alleles at frequencies of 0.25, 0.5, and 0.25, respectively, if the alleles segregate at random. In contrast, siblings affected by a given disease would predictably share the chromosomal region containing the disease relevant loci (and thus polymorphic alleles located within such a region) at frequencies greater than would be seen under the null hypothesis of random chance segregation. The ASP analysis has proven very valuable in the search for complex disease gene loci and remains the state of the art in this regard. However, there are drawbacks to this approach, including, for example, the fact that sib pairs share many chromosomal segments by chance alone. Thus, in order to avoid false-positive results, a large number of ASPs must be examined.

Once genotypes have been derived from genome-wide scans, the data can be analyzed so as to determine whether any markers cosegregate with the disease/phenotype. Segregation of alleles is evaluated using linkage

analysis and is generally quantitated using a lod (logarithm of odds) score (Z), which depicts the likelihood of linkage vs. the null hypothesis of no linkage (i.e., random segregation between marker and disease phenotype). For single-gene disorders and two-point linkage analysis, a lod score of 3.0 is considered definitive evidence of linkage; under these conditions, linkage is rejected if $Z < -2.0$ while values between -2.0 and $+3.0$ are considered inconclusive. However, these standards do not hold in the context of multiple testing in genome-wide scans, wherein the simultaneous analysis of >300 rather than one or two individual loci markedly increases the chance for false-positive results. Accordingly, for linkage analysis in which genome-wide scans are used to identify single-gene disease loci, the accepted threshold for significance is a lod score of 3.3 or greater (9).

For complex genetic diseases, defining the level of significance in a genome-wide study is crucial to establishing which disease linked marker loci represent real disease susceptibility loci. The difficulties inherent to this issue arise in part from the incorporation of multiple test markers. While it is generally agreed that genome-wide scans require a threshold p value more stringent than required for two-point analysis, controversy has existed as to the level of this value. A set of criteria for interpreting linkage data have been proposed by Lander and Kruglyak and these have been widely accepted (38), but the correct thresholds of significance for use in such studies remains controversial (39). Based on the 1995 criteria set out by Lander and Kruglyak, a lod score of 3.6 ($p < 2\times10^{-5}$) can be considered as the threshold for significant linkage, a lod score of >2.2 ($p < 7\times10^{-4}$) can be considered suggestive of linkage in ASP analyses, and replication of previously reported significant linkage data requires a $p < 0.01$. These guidelines are likely to be revisited in the near future in view of the increasing complexity and density of genotype data derived from genome-wide scans, particularly those based on tens to hundreds of thousands of SNPs.

As evidenced by the extensive variation in the linkage data derived from independent groups working on the same complex genetic disorder, linkage analysis in multifactorial disease is not straightforward, independent genome-wide screens typically revealing overlapping, but non-identical, sets of loci contributing to a specific disease. One solution to resolving such discrepancies involves the meta-analysis of data from multiple genome screens, a strategy which may identify the most consistently linked loci and thereby redirect research efforts towards discovery of disease genes most likely to be relevant in many different populations (10,40). While some of the disparities between linkage data may reflect easily correctable errors such as misread gels, switched samples, incorrect scoring of family relationships or misdiagnosis/misclassification of phenotype, other key contributing factors such as incomplete penetrance and genetic heterogeneity are difficult to address. An additional problem is the relatively low resolution of STR-based marker scans, and the consequent necessity for analysis of very large

numbers of families so as to delineate a target region amenable to gene identification (i.e., <2 cM). Most commonly, linkage data emanating from STR marker scans has revealed target regions that are far too large (e.g., 10–40 Mb) to attempt disease gene isolation. Accordingly, genome-wide scanning needs to be complemented by additional mapping strategies, which enable refinement of regions of interest to sizes amenable to physical mapping and candidate gene analysis.

One other confounding factor in the use of high-throughput genotype data for disease gene mapping is the possibility of undetected genotype errors in large datasets. While Mendelian (inheritance) errors can be identified readily, genotype errors that yield data consistent with Mendelian inheritance can significantly interfere with downstream genetic analyses, for example, through the introduction of spurious "double recombination" events or through the elimination of *bona fide* "real" recombinations. Even in the highest-quality datasets, problems such as missing genotypes, sample mix-up, or miscalled genotypes can introduce significant biases that can be difficult to detect and which can reduce ability to localize disease genes with the accuracy required for such studies (41).

8. REFINEMENT OF CHROMOSOMAL LOCALIZATIONS

The disease gene localizations provided by STR marker-based genome-wide scans are usually broad, and thus regions of interest identified in this manner need to be refined using additional genetic strategies. One commonly used approach involves screening for associations between the disease locus and markers in the candidate region. To this end, additional microsatellite markers that increase the density of coverage (e.g., from 10 to 1 cM resolution) are placed across the region so as to search for disease-associated LD. The term LD is applied in instances in which association occurs in the presence of linkage, as opposed to situations in which association is found without linkage (e.g., in candidate gene analysis, or, spuriously, as a result of other factors such as population stratification). In the presence of LD, the cosegregation of two or more alleles across many generations and the consequent transmission of a "haplotype" of alleles arranged along a single chromosome can be demonstrated (42). The use of LD mapping in disease gene localization is predicated upon the notion that disease gene alleles segregate in association with a set of flanking alleles that comprise an ancestral haplotype derived from a founder individual who originally introduced the disease gene allele into the population under study. Thus, affected individuals may share a set of contiguous alleles within a chromosomal region that comprise a disease-associated haplotype encompassing the disease-related allele. Regions of LD are generally much smaller than "linked" regions and thus LD mapping can facilitate refinement of candidate regions elucidated from linkage data. Additionally, the increasing definition of comprehensive

maps of "blocks" of LD in the human genome should allow for LD mapping with SNPs "tagging" specific haplotypes and should therefore markedly reduce the numbers of SNPs required for LD mapping studies. Thus, for example, one analysis of such haplotype-tagging SNPs revealed the information provided by 122 genotyped SNPs to be completely reproduced by analysis of 34 SNPs, a 3.6-fold reduction in genotyping effort (28). Similarly, we have shown that any one of 11 SNPs within a 250 region on chromosome 5q31 serves to tag a Crohn's disease-associated haplotype at this locus (32). Thus, while complex haplotype structures are likely to be the rule rather than the exception in most regions of the genome (43–45), the identification of haplotype blocks, a major goal of the international HapMap consortium, should greatly reduce the number of SNPs required for both genome wide and refinement mapping efforts (28).

The transmission disequilibrium test (TDT) is a non-parametric test that is commonly used to identify association in the presence of linkage (i.e., LD) in datasets derived from "trio" families, (two parents and their affected child) (46). This family-based association test involves comparison of the number of times a specific allele is transmitted to an affected offspring versus the number of times that allele is not transmitted, the expectation being that a marker allele in LD with the disease gene variant will be disproportionately transmitted to affected progeny. As TDT is typically carried out by analysis of trio families including a single affected individual, it is usually possible to ascertain many more families for TDT than for affected sib pair analysis. To accommodate parent inavailability, variant approaches to TDT, incorporating, for example, affected/unaffected sib pairs in the absence of parental samples (47,48) and extended family structures (49), have also been developed and can be combined with traditional TDT to increase power in datasets lacking some of the parental information. Although the TDT circumvents some of the major problems inherent to case-control association approaches (most notably, population stratification), under some circumstances, the latter strategy may have greater statistical power to detect disease associations with similar sized sample sets (33,50). Thus, the choice of study design and methodologies for association analyses is most often influenced by the presence or absence of suitable family structures in the sample set (51).

9. CASE–CONTROL GENETIC ASSOCIATION STUDIES

The increasing feasibility of candidate gene-based disease gene identification has engendered enormous interest in case-control genetic association studies, wherein relevance of a specific gene to a disease/phenotype is examined by comparing affected individuals and matched controls with respect to frequencies of particular allelic variants of the gene of interest (52). Such case-control studies involve the analysis of putative susceptibility (positively associated)

or protective (negatively associated) alleles. This methodology has many advantages over traditional linkage analysis for disease gene discovery. Such advantages include, for example, lack of requirement for parental DNA samples, which may be very difficult or impossible to obtain in the context of late-onset diseases. Both family based and case-control association methods also involve the analysis of gene variants (polymorphisms) which may not only exhibit LD with the gene defect of interest, but which may actually represent the disease-relevant defect per se (52). To render possible LD mapping in the context of a case-control design, statistical methods such as the expectation-maximization (EM) algorithm have been developed to enable definition of haplotypes (arrangements of alleles along a single chromosome) associated with disease susceptibility (53). One major consideration, however, which needs to be factored into any genetic case-control study, is the ethnic constitution of the cases versus the controls. Because of the potential for allele frequencies to vary significantly between different ethnic populations independent of differences related to disease susceptibility, divergence in the ethnic representation in cases compared to controls creates a significant potential for spurious false positive results. This problem, referred to as population stratification, together with inadequate sample size and incomplete or inaccurately derived assignment of phenotype, may account for the high rate of failure in replicating case-control genetic association data in independent studies. To address these problems, standard guidelines identifying criteria required for rigorous case-control genetic association studies have been derived (54,55) and include: large sample sizes (on the order of 1000 unrelated cases and at least as many controls); a requirement for low p values, appropriately corrected for multiple testing, and correspondingly high odds ratios for susceptibility; inclusion of a replication study performed on an independently ascertained population; and "biological sense" for the demonstrated association—that is, the disease-associated allele(s) should affect the gene product in a physiologically meaningful way. The last of these criteria, however, is often difficult to fulfill and even under circumstances in which all these criteria are met, some genetic association data still prove non-reproducible. Nevertheless, with the increasing availability of high-density SNP maps, better technologies for automated high-throughput SNP genotyping, international haplotype mapping (HapMap) data (www.hapmap.org), and higher quality DNA sequence data for most of the human genome (56), the quality of genetic association data is continually improving and this study design is almost certain to become the most practical and rapid approach to disease gene discovery, particularly for identifying genes relevant to late-onset diseases.

The potential for SNP-based genome-wide association studies to identify susceptibility genes in complex diseases have been investigated by assessing the extent of LD between known disease susceptibility genes and specific SNP markers flanking these genes. Data from one such analysis, for example, involving the *APOE* gene implicated in late-onset Alzheimer

disease, revealed that of 10 SNPs lying within the 800 kb flanking the gene, only three SNPs located 10, 12–30, and 440 kb, respectively, from *APOE* showed significant LD with *APOE* and strong association with Alzheimer disease (57). By contrast, other SNPs located as little as 60–80 kb from the gene showed no evidence of LD or disease association. LD has also been reported between a number of other genes and STR markers mapping up to 400 kb from the genes of interest (58), although more typically, LD between disease genes and associated SNPs is in the range of 60 kb, at least for European Caucasian populations (59). The size of LD blocks across the genome is, however, highly variable, some regions of LD on chromosome 22, for example, spanning >700 kb (33). Moreover, extent of LD across specific chromosomal segments is likely to vary between different ethnic groups, a point which needs to be considered in selecting haplotype-tagging SNPs for association studies. These considerations illustrate the complexity inherent to SNP-based candidate gene analysis dependent on marker-disease LD and further highlight the need to analyze many SNPs in or near a gene of interest and to assess the impact of suspected disease-causing variants on the function and/or expression of the relevant gene product. The SNPs for such studies can be selected and prioritized based on their physical location near or within genes of interest, presence within a specific haplotype block, and/or by computational approaches enabling functional impact of SNP variants to be predicted and analysis thereby focused on putative deleterious SNPs (see below).

10. UTILITY OF ISOLATED POPULATIONS

Isolated populations have been used successfully to identify a number of complex disease loci, including for example, putative susceptibility loci for types 1 and 2 diabetes mellitus (60,61), multiple sclerosis (62), hyperlipidemia (63), and Hirschsprung's disease (64). In such populations, inbreeding and genetic isolation alters disease allele frequencies and often reduces genetic heterogeneity substantively compared with that found in outbred populations. Because affected individuals are more likely to share functional alleles identical by descent in this setting than in more admixed populations, the number of alleles contributing to a complex trait is likely to be lower and recessive genetic effects on phenotypes are more easily identified. Data from association studies have in fact directly shown the detection of small genetic effects on a quantitative trait to be enhanced in consanguineous populations (65). Similarly, inbreeding coefficients of approximately 0.01 have been shown to be associated with increased power in detecting the effect of a candidate gene (66). In addition to consanguinity, isolated societies often contain large, multigenerational families with multiple affected individuals, a pedigree structure that greatly facilitates linkage analysis. In founder populations of recent origin, the extent of LD across a genetic interval

may be larger than in older populations and therefore associations between polymorphic markers and a disease trait can be detected over greater genetic distances (67). This phenomenon may, however, represent a double-edged sword, potentially diminishing the extent to which genetics can be used to refine intervals of interest to sizes amenable to candidate gene analysis. In this latter situation, examination of a more outbred population may allow tighter refinement of the relevant interval, provided that the disease association seen in the inbred cohort is also present in the outbred population. Finally, population isolates may also have common traditions, lifestyles, and/or environmental exposures, which also contribute to overall lessening of the confounding effects of background environmental heterogeneity that may mask genetic effects in outbred populations.

Despite these advantages, there are a number of caveats to the interpretation of genetic data garnered from inbred populations. For example, distinguishing a disease-causing allele (which might be comparatively common in an inbred population) from a rare polymorphism unique to the population may be difficult in a population isolate (68). Consanguinity may also increase the type 1 error rate in sibling pair linkage analysis, leading to false positive linkage results (66). Finally, alleles identified as disease causative variants in such populations may be rare alleles that are not relevant to disease etiology in more admixed populations and the practical clinical utility of such data may therefore be limited to the isolated population studied.

11. GENOME SEQUENCE

The strategy of positional cloning, in which a disease gene is identified by virtue of its position within the genome rather than by functional characterization, has been used in the mapping and identification of many disease loci. However, with the completion of the human genome project, there has been a radical change in positional cloning technologies, the need for laborious physical mapping techniques such as radiation hybrid mapping, ordered library contigs, and pulsed-field gel electrophoresis mapping having been essentially abrogated by availability of the full genome sequence. Nevertheless, such resources are still of value in filling remaining sequence gaps and in understanding the biology of duplicated regions of the genome, which may not be correctly detected or assembled in genome sequence databases (69). The success of candidate gene identification by inspection of the genome sequence within a small region of linkage or LD is very much dependent on the accuracy and resolution of both genome sequence data and the precise positioning of both genetic markers and candidate genes. While the quality of human DNA sequence data in both public and private databases continues to improve, at the current time, not all genes are accurately represented or correctly predicted, and examination of multiple

DNA sequence databases and gene prediction algorithms is still warranted (70).

Genetic maps provide critical tools for positioning disease gene loci ascertained from genome-wide or regional refinement linkage data within the framework of the sequenced genome. Integrating physical and genetic maps was, for many years, a laborious process and ambiguity as to the physical positions of some STRs often hampered efforts to identify disease genes. However, the physical positions of the markers used in contemporary genetic maps are now well established in relation to the genome sequence and the possibility to positionally clone disease genes has thus been greatly enhanced. A wide variety of genetic maps, including those derived by CEPH/Généthon (Paris, France) and the Marshfield Clinic (Marshfield, WI), are currently available and serve as the start-point for gene mapping studies. Nevertheless, it seems likely that in the near future, comprehensive SNP maps based on standardized panels of haplotype tagging SNPs (28) will supplant these STR-based maps as the template for disease gene mapping projects.

Once the full complement of genes within a region of interest has been elucidated, the potential relevance of each gene to disease can then be explored. Identification of the sought after gene may be facilitated by prior knowledge of the gene product functions or expression patterns in tissues relevant to disease pathology. Our group's identification of *SLC22A4* and *SLC22A5* as Crohn's disease susceptibility genes, for example, was greatly facilitated by our data revealing these genes to be expressed in tissues relevant to intestinal inflammation, i.e., in cells of the colonic epithelium and in T cells and macrophages infiltrating colonic tissue (71). Based on such biological data, genes of interest can be prioritized for further investigation and studied by direct sequence analysis or other mutation detection methods (e.g., denaturing high-performance liquid chromatography or DHPLC) (72) to identify disease-associated variants. As the genetic lesions responsible for complex genetic disease are likely to be more subtle (i.e., SNPs) than those found in single-gene disease, and will not be entirely correlated with the presence or absence of disease, verification of a candidate gene as etiologically relevant to this class of genetic disease will generally require some functional data linking the gene defect to the clinical phenotype.

12. COMPUTATIONAL ALGORITHMS AND DATABASE MINING

At the time of this writing, genome sequence for 9 of the 24 human chromosomes has been fully completed and the remainder will soon be available (reviewed in Ref. 56). These sequence data have dramatically changed the positional cloning process, substantively improving capacity to integrate genetic mapping with DNA sequence data and allowing for the application

of computational methods to identify, within regions of interest, candidate genes, SNPs with potential pathological consequences, and other sequence properties of possible relevance to disease pathophysiology, such as transcriptional regulatory regions. Analysis of selected chromosomal regions for candidate genes has traditionally involved such strategies as exon trapping, CpG island identification, and isolation of cDNA clones representing genes within the region of interest. While such approaches may still be required to identify all of the genes within a given region, the need to involve these labor intensive strategies is progressively diminishing as new computational approaches and databases are used to prioritize and inform "wet lab" experimentation.

A number of programs incorporating algorithms for gene or exon prediction have been developed and output from many such programs is already available for the public builds of the genome sequence. The algorithms incorporated in such programs have varying sensitivities and specificities for exon/gene identification and a robust analysis would therefore require use of several such programs. Regions of interest can also be scanned for transcription start sites, potential splice sites, polyadenylation signals, CpG islands, transcription factor binding sites, and other indicators of coding sequence (73). In addition, sequences of interest can be compared by automated software methods to a database of expressed sequence tags (ESTs) in order to identify previously known transcribed sequences that correspond to genes in the region. The ESTs that occur in clusters, align to genomic sequence, are predicted by exon/gene prediction algorithms, and exhibit intron–exon structure are highly likely to represent fragments of *bona fide* genes, which will thus warrant further investigation. Much of this work has already been performed and the results are available online in the public portals to the human genome sequence, including, for example, the Human Genome Browser Gateway hosted by the University of California, Santa Cruz, at www.genome.ucsc.edu, and the well-curated set of human genome tools and resources hosted by the National Center for Biotechnology Information at www.ncbi.nlm.nih.gov/genome/guide/human/.

Informatics is also key to many other aspects of positional cloning, such as, the mining of cumulative SNP databases. Novel computational approaches have been developed, for example, to assist in the identification of SNP variants with likely functional impact, such strategies incorporating epidemiological and evolutionary approaches and utilizing multiple sequence alignment comparisons of known or suspected functional domains (74–76). While some SNPs introduce easily predictable changes in protein function (e.g., a change of a conserved residue in an enzyme's catalytic site), others have more subtle effects, altering, for example, important non-coding regulatory regions or mRNA stability or engendering amino acid substitutions that appear conservative but have functional impact. In silico approaches to the identification of such functional variants are also emerging and include, for

example, comparisons of comparable sites in genomes from different species and prediction of likely functional relevance based on the average rate of conservation between these species (75).

In addition to improvements in database mining strategies, increased sophistication in statistical methodologies for disease gene mapping is also a major factor in accelerating disease gene discovery research. Among the many difficult issues emerging from the trend to case-control study designs, a particularly important problem has been the analysis of non-phased genotype data (i.e., data in which the arrangement of alleles in haplotypes on individual chromosomes is unknown) in which haplotypes that may contribute to disease susceptibility cannot be explicitly determined. This problem has, however, been resolved to some extent through the development of algorithms such as the EM algorithm, a method now widely used to reconstruct haplotype frequency distributions from unphased genotype data (53,77). Association of specific haplotypes with disease susceptibility can then be determined using a likelihood ratio test based on the haplotype frequencies in cases as compared with controls. We have previously utilized EM analysis to show that a 250 kb haplotype block delineated by 11 SNPs in strong LD contains a two SNP haplotype involving variants in the *SLC22A4* and *SLC22A5* genes, respectively, with genetic, functional and histopathological links to Crohn's disease (71). Identification of the *SLC22A4*/*SLC22A5* haplotype, which demonstrated the role of these genes in susceptibility to Crohn's disease, was made possible by use of these recently developed statistical methods.

13. CONCLUSION

Discovery of the genetic variants underlying the most prevalent and devastating human diseases is one of the most significant goals of the human genome project. As exemplified by the identification of *SLC22A4* and *SLC22A5* gene variants as susceptibility factors for Crohn's disease, highly integrated approaches to gene discovery can yield new knowledge of disease mechanisms, potential drug targets, and diagnostic tools. Ultimately, many such discoveries will lead to new insights into the mechanism of drug actions and toxicity and to the personalization of therapies based on the underlying genetic lesions specific to an individual's disease. While significant issues such as genetic heterogeneity, late onset of disease, and incomplete penetrance complicate disease gene isolation in complex, multifactorial disease, novel statistical, computational, and bioinformatic approaches, coupled to tremendous increases in throughput of genotyping technologies, have markedly accelerated the pace at which the genome can be screened and mined. In particular, the generation of genome-wide SNP microarrays and other very high throughput SNP genotyping technologies has brought genome-scale

SNP association studies into the realm of possibility, from the standpoints of both cost and practicality.

To fully capitalize on the wealth of current and future disease gene discoveries, an acceleration in the pace at which such discoveries are translated to the clinic will also be required. Although DNA-based diagnostic products are now routinely used in many settings (e.g., mutation panels for diagnosis/risk prediction in Cystic Fibrosis, breast cancer, or cardiovascular disease), many attendant problems associated with the application of genetic data to the clinic remain unresolved. These include, for example, complex ethical, sociological, and legal issues that require continuous and proactive attention and management and the many problems inherent to communicating and even interpreting complex genetic data. Despite these caveats, disease gene discovery has many potential immediate benefits, including, for example, the definition of disease-specific targets for pharmacological intervention and the possibility to personalize medical treatments so as to reduce risk and enhance efficacy. Disease gene discovery also opens the door to large-scale population studies directed at definition of the environmental–gene interactions which underlie disease, and for meta-analytical studies integrating data from multiple groups and thereby identifying the gene–gene interactions contributing to disease. These data, in turn, will inform and direct biological experimentation to elucidate underlying disease mechanisms. Such knowledge is anticipated to provide further novel drug targets, new molecular markers that can be used to stratify patient populations so as to predict severity, prognosis, and drug responsiveness, as well as basic knowledge of disease etiology of potential use in risk reduction.

Ultimately, knowledge of the complete complement of genes which cause and modulate the common diseases afflicting humankind will shift the paradigm of medical care to "individualized" care, wherein disease risk can be identified and mitigated in healthy individuals and treatment for those affected by disease can be tailored so as to minimize toxicity and optimize efficacy of selected therapies. The era of genomic medicine is well underway, and the challenge in the early part of the 21st century will be to delineate the genetic basis of all human disease and to translate this extraordinary achievement to radical improvements in the quality of health care delivery.

ACKNOWLEDGMENTS

Compilation of this chapter was made possible by funding from the Crohn's and Colitis Foundation of Canada, Canadian Institutes of Health Research, Ellipsis Biotherapeutics Corporation and Genome Canada. K.A.S. is a Canadian Institutes of Health Research Senior Scientist and a McLaughlin Centre for Molecular Medicine Scientist.

REFERENCES

1. Bentley DR. Genomes for medicine. Nature 2004; 429:440–445.
2. Hudson TJ. The human genome project: tools for the identification of disease genes. Clin Invest Med 1998; 21:267–276.
3. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. Nature 2004; 429:464–468.
4. Gitschier J, Wood WI, Gopralka TM, Wion KL, Chen EY, Eaton DH, Vehar GA, Capon DJ, Lawn RM. Characterization of the human factor VIII gene. Nature 1984; 312:326–330.
5. Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LP. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet 1999; 21:379–384.
6. Robson KJH, Chandra T, MacGillivray TRA, Woo SLC. Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. Proc Natl Acad Sci USA 1982; 79:4701–4705.
7. Schalling M, Hudson TJ, Buetow KH, Houseman DE. Direct detection of novel expanded trinucleotide repeats in the human genome. Nat Genet 1993; 4:135–139.
8. Weatheral DJ. The Thalassemias. In: Stamatoyannopoulos G, Nienhuis AW, Majerus PH, Varmus H, eds. The Molecular Basis of Blood Diseases. Philadelphia: Saunders WB, 1994:157–205.
9. Lander ES, Schork N. Genetic dissection of complex traits. Science 1994; 265:2037–2048.
10. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003; 33:177–182.
11. Collins FS. Of needles and haystacks: finding human disease genes by positional cloning. Clin Res 1991; 39:615–634.
12. Strathdee CA, Gavish H, Shannon WR, Buchwald M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. Nature 1992; 356:763–767.
13. Lee B, Godfrey M, Vitale E, Hori H, Mattei MG, Sarfarazi M, Tsipouras P, Ramirez F, Hollister DW. Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes. Nature 1991; 352:330–334.
14. Dietz HC, Cutting CR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, Stetten G, Meyers DA, Francomano CA. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature 1991; 352:337–339.
15. Hill AV. Immunogenetics and genomics. Lancet 2001; 357:2037–2041.
16. Silman AJ, Pearson JE. Epidemiology and genetics of rheumatoid arthritis. Arthritis Res 2002; 4(suppl 3):S265–S272.
17. Duerr RH. The genetics of inflammatory bowel disease. Gastroenterol Clin North Am 2002; 31:63–76.
18. Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M, Buchwald M, Riordan JR, Tsui LC, Collins F. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 1989; 245:1059–1065.

19. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. Science 1989; 245:1073–1080.
20. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, Drumm ML, Iannuzzi MC, Collins F, Tsui LC. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989; 245:1066–1073.
21. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273:1516–1517.
22. Silverberg MS, Daly MJ, Moskovitz DN, Rioux JD, McLeod RS, Cohen Z, Greenberg GR, Hudson TJ, Siminovitch KA, Steinhart AH. Diagnostic misclassification reduces the ability to detect linkage in inflammatory bowel disease genetic studies. Gut 2001; 49:773–776.
23. Weeks DE, Lange K. A multilocus extension of the affected pedigree member method of linkage analysis. Am J Hum Genet 1992; 50:859–868.
24. Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 1995; 57:439–454.
25. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996; 58:1347–1363.
26. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 1989; 44:388–396.
27. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 1987; 235:1616–1622.
28. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet 2001; 29:233–237.
29. John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC. Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. Am J Hum Genet 2004; 75:54–64.
30. Schulze TG, Chen Y-S, Akula N, Hennessy K, Badner JA, McInnis MG, DePaulo JR, Schumacher J, Cichon S, Propping P, Maier W, Rietschel M, Nöthen MM, McMahon FJ. Can long-range microsatelliate data be used to predict short-range linkage disequilibrium? Hum Mol Genet 2002; 11: 1363–1372.
31. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM, Green T, Brettin TS, Stone V, Bull SB, Bitton A, Williams CN, Greenberg GR, Cohen Z, Lander ES, Hudson TJ, Siminovitch KA. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet 2000; 66:1863–1870.

32. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, Kocher K, Miller K, Guschwan S, Kulbokas EJ, O'Leary S, Winchester E, Dewar K, Green T, Stone V, Chow C, Cohen A, Langelier D, Lapointe G, Gaudet D, Faith J, Branco N, Bull SB, McLeod RS, Griffiths AM, Bitton A, Greenberg GR, Lander ES, Siminovitch KA, Hudson TJ. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 2001; 29:223–228.
33. Bell JI. Single nucleotide polymorphisms and disease gene mapping. Arthritis Res 2002; 4(suppl 3):S273–S278.
34. Kruglyak L. The use of a genetic map of biallelic markers in linkage studies. Nat Genet 1997; 17:1–4.
35. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. Nature 429:446–452.
36. Smith T. Whole genome variation analysis using single molecule sequencing. Drug Discov Today: Targets 2004; 3:112–116.
37. Kwok P-Y, Gu Z. Single nucleotide polymorphism libraries: why and how are we building them? Mol Med Today 1999; 5:538–543.
38. Lander ES, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995; 11:241–247.
39. Morton NE. Significance levels in complex inheritance. Am J Hum Genet 1998; 62:690–697.
40. van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM. Genome scan meta-analysis group of the IBD international genetics consortium. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum Mol Genet 2004; 13:763–770.
41. Broman KW. Cleaning genotype data. Genet Epidemiol 1999; 17 (suppl 1): S79–S83.
42. Xiong M, Guo S-W. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. Am J Hum Genet 1997; 60:1513–1531.
43. Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T, Suzuki M, Nagasaki M, Ohtsuki M, Ono M, Furukawa H, Nagashima M, Yoshino S, Mabuchi A, Sekine A, Saito S, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet 2003; 35:341–348.
44. Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, Ankener WM, Alfisi SV, Kuo F-S, Camisa AL, Pazorov V, Scott KE, Carey BJ, Faith J, Katari G, Bhatti HA, Cyr JM, Derohannessian V, Elosua C, Forman AM, Grecco NM, Hock CR, Kuebler JM, Lathrop JA, Mockler MA, Nachtman EP, Restine SL, Varde SA, Hozza MJ, Gelfand CA, Broxholme J, Abecasis GR, Boyce-Jacino MT, Cardon LR. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. Nat Genet 2003; 33:382–387.
45. Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. Trends Genet 2003; 19:135–140.
46. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993; 52:506–516.

47. Sun F, Flanders WD, Yang Q, Khoury MJ. Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. Am J Epidemiol 1999; 150:97–104.
48. Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 1998; 62:450–458.
49. Martin ER, Monks SA, Warren LL, Kaplan NL. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 2000; 67:146–154.
50. McGinnis R, Shifman S, Darvasi A. Power and efficiency of the TDT and case-control design for association scans. Behav Genet 2002; 32:135–144.
51. Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 2000; 72:63–132.
52. Long AD, Langley CH. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res 1999; 9:720–731.
53. Polanska J. The EM algorithm and its implementation for the estimation of frequencies of SNP-haplotypes. Int J Appl Math Comput Sci 2003; 13: 419–429.
54. Todd JA. Interpretation of results from genetic studies of multifactorial diseases. Lancet 1999; 354 (suppl 1):15–16.
55. Page GP, George V, Go RC, Page PZ, Allison DB. "Are we there yet? Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. Am J Hum Genet 2003; 73:711–719.
56. Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, Bajorek E, Black S, Chan YM, Denys M, Escobar J, Flowers D, Fotopulos D, Garcia C, Gomez M, Gonzales E, Haydu L, Lopez F, Ramirez L, Retterer J, Rodriguez A, Rogers S, Salazar A, Tsai M, Myers RM. Quality assessment of the human genome sequence. Nature 2004; 429:365–368.
57. Roses AD, Pericak-Vance MA. Alzheimer's disease and other dementias. In: Rimoin DL, ed. Emery and Rimoin's Principles and Practice of Medical Genetics. New York: Churchill-Livingston, 1997:1807–1825.
58. Koch HG, McClay J, Loh E-W, Higuchi S, Zhao J-H, Sham P, Ball D, Craig IW. Allele association studies with SSR and SNP markers at known physical distances within a 1 Mb region embracing the *ALDH2* locus in the Japanese, demonstrates linkage disequilibrium extending up to 400 kb. Hum Mol Genet 2000; 9:2993–2999.
59. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. Nature 2001; 411:199–204.
60. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM. A genome-wide search for human type 1 diabetes susceptibility genes. Nature 1994; 371:130–136.
61. Hanis CL, Boerwinkle E, Chakraborty R, Ellsworth DL, Concannon P, Stirling B, Morrison VA, Wapelhorst B, Spielman RS, Gogolin-Ewens KJ, Shepard JM, Williams SR, Risch N, Hinds D, Iwasaki N, Ogata M, Omori Y, Petzold C, Rietzch H, Schroder HE, Schule J, Cox NJ, Menzel S, Boriraj VV, Chen X. A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. Nat Genet 1996; 13:161–166.

62. Kuokkanen S, Sundvall M, Terwilliger JD, Tienari PJ, Wikstrom J, Holmdahl R, Pettersson U, Peltonen L. A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntenic to the murine locus *Eae2*. Nat Genet 1996; 13: 477–480.
63. Pajukanta P, Nuotio I, Terwilliger JD, Porkka KV, Ylitalo K, Pihlajamaki J, Suomalainen AJ, Syvanen AC, Lehtimaki T, Viikari JS, Laakso M, Taskinen MR, Ehnholm C, Peltonen L. Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. Nat Genet 1998; 18:369–373.
64. Puffenberger EG, Hosoda K, Washington SS, Nakao K, deWit D, Yanagtisawa M, Chakravarti A. A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. Cell 1994; 79:1257–1266.
65. Genin E, Clerget-Darpoux F. Association studies in consanguineous populations. Am J Hum Genet 1996; 58:861–866.
66. Genin E, Clerget-Darpoux F. Consanguinity and the sib-pair method: an approach using identity by descent between and within individuals. Am J Hum Genet 1996; 59:1149–1162.
67. Peltonen L, Uusitalo A. Rare disease genes—lessons and challenges. Genome Res 1997; 7:765–767.
68. Sheffield VC, Stone EM, Carmi R. Use of isolated inbred human populations for identification of disease genes. Trends Genet 1998; 14:391–396.
69. Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. Genome Biol 2003; 4:R25.
70. Scherer SW, Cheung J. Discovery of the human genome sequence in the public and private databases. Curr Biol 2001; 11:R808–R811.
71. Peltekova VD, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X, Newman B, Van Oene M, Cescon D, Greenberg G, Griffiths AM, St George-Hyslop PH, Siminovitch KA. Functional variants of OCTN cation transporter genes are associated with Crohn disease. Nat Genet 2004; 36:471–475.
72. Giordano M, Oefner PJ, Underhill PA, Cavalli-Sforza LL, Tosi R, Richiardi PM. Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. Genomics 1999; 56:247–253.
73. Milanesi L, Rogozin I. Prediction of human gene structure. In: Bishop M, ed. Guide to Human Genome Computing. Cambridge: Academic Press, 1998; 215–259.
74. Zhu Y, Spitz MR, Amos CI, Lin J, Schabath MB, Wu X. An evolutionary perspective on single-nucleotide polymorphism screening in molecular cancer epidemiology. Cancer Res 2004; 64:2251–2257.
75. Bazykin GA, Kondrashov FA, Ogurtsov AY, Sunyaev S, Kondrashov AS. Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. Nature 2004; 429:558–562.
76. Sunyaev S, Ramensky V, Koch I, Lathe W III, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet 2001; 10:591–597.
77. Cheng R, Ma JZ, Wright FA, Lin S, Gao X, Wang D, Elston RC, Li MD. Nonparametric disequilibrium mapping of functional sites using haplotypes of multiple tightly linked single-nucleotide polymorphism markers. Genetics 2003; 164:1175–1187.

25

Genome Variation Influencing Gene Copy Number and Disease

Lars Feuk and Stephen W. Scherer

Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada

Paul Buckland

Department of Psychological Medicine, University of Wales College of Medicine, Cardiff, U.K.

Xavier Estivill

Program of Genes and Diseases, Center for Genomic Regulation, Barcelona, Spain

1. INTRODUCTION

In the past few years there has been significant progress towards completing the sequences and beginning to characterize the content of the genomes of several mammalian species. The most notable advances have been made in the study of the human genome. The availability of a near-finished reference DNA sequence has been most crucial in clinical genetic- and genomic-investigations because for the first time it provides a common template for comprehensive comparative studies aimed at cataloguing genotypes and their influence on phenotypic outcomes. Although much progress has been made in this endeavor (~2000 disease genes or associated variants identified) there are still some 4000 genetic diseases for which the molecular etiology is unknown. There are also numerous phenotypic traits in the apparently "healthy" population that can have a strong genetic component; one important example being genetic factors affecting drug metabolism. Genetic

variation in the human genome has, until recently, mainly been studied either at the single nucleotide- or the karyotypic-level. The most common class of variation is single nucleotide (nt) substitutions. These mostly benign changes are now well studied with an estimated 11 million single-nucleotide polymorphisms (SNPs) currently described in the human population. Most of the Human Genome Project-coordinated endeavors to catalogue common variants in the genome sequence have been focused on SNPs. Small insertions and deletions are also usually grouped into this category. Genomic variation detected by karyotyping would include larger tracts of usually contiguous DNA that can vary in copy number (deletion and duplication), distribution (translocations and insertions), or orientation (inversions) along the chromosomes. In most cases these large genomic rearrangements are associated with clinical outcomes. To date, there has not been an exhaustive assessment of the frequency, extent, or distribution of variants in the kilobase (kb) to megabase (Mb) size range, mainly due to lack of robust genome scanning technologies available for this resolution of analysis. However, to partially address this problem there have been recent technical advances that capitalize on the genome sequence as a reference substrate allowing rapid assessment of gains or losses of sequences along chromosomes. As the data begin to accumulate it is becoming increasingly apparent that these "so-called" large-scale copy number variants (or LCVs), often averaging hundreds of kb in size, are present in the genomes of apparently healthy individuals at a much higher frequency than originally thought. In many cases these genomic variants partially or entirely encompass genes, which can affect their copy number. Moreover, in some cases they overlap with nearly identical segmentally duplicated DNA (called low-copy repeats or duplicons). Given that segmental duplications (and possibly LCVs) are implicated in a growing list of over 30 human diseases, which arise due to a gain, loss, or disruption of dosage sensitive genes or regulatory regions these new observations may also be relevant to other unresolved genetic diseases. In this chapter we will describe three different categories of genomic variation and discuss how each type of variation may influence human disease. The different types of variation are outlined in Figure 1. We will describe what are known about genomic disorders and the mechanisms that cause them. We will then discuss how similar molecular events may underlie certain phenotypic variation and susceptibility to common complex diseases as well as influence the dynamic structuring of the human genome.

2. MECHANISMS OF GENOME REARRANGEMENTS

There are a large number of diseases that are caused by genomic rearrangements involving several genes and they are usually referred to as "genomic

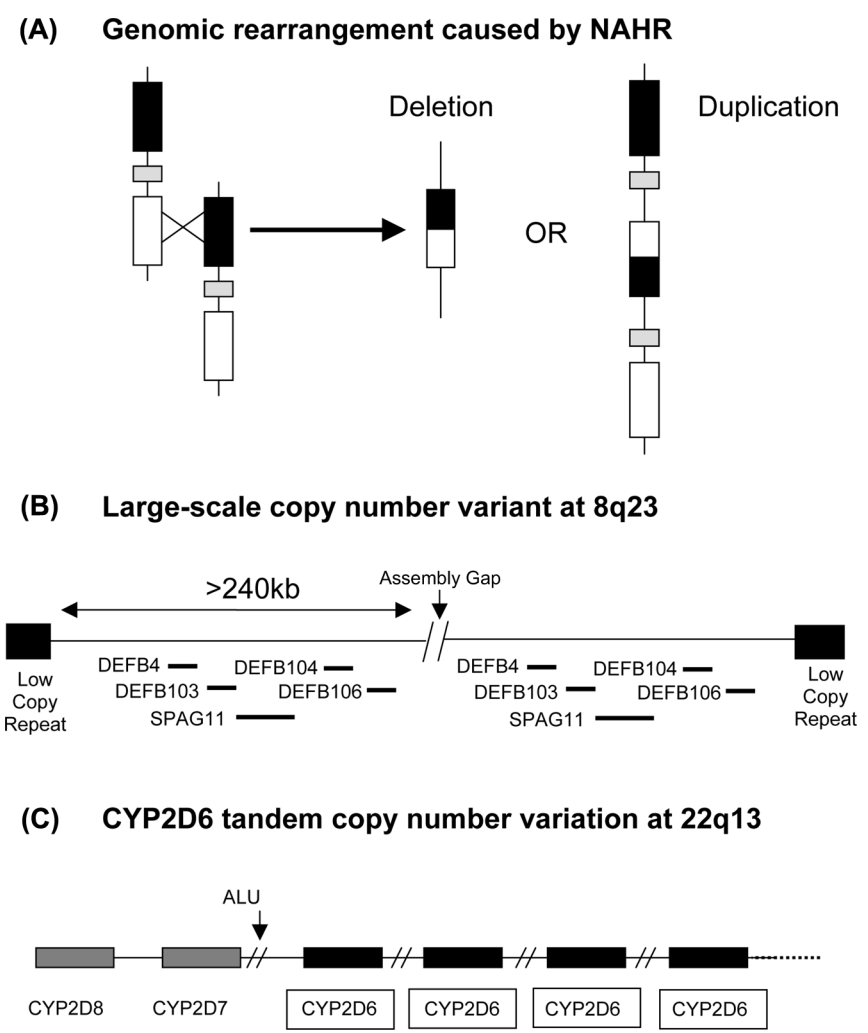


Figure 1 This figure outlines the three types of genomic variation discussed in the text. (**A**) Genomic rearrangements can be induced by NAHR. This mechanism is involved in the causation of a number of genomic disorders, e.g., DiGeorge Syndrome. (**B**) A number of recent publications indicate that there is a high incidence of large-scale variation in the human genome. The defensin gene cluster at 8q23 is an example of a large-scale genomic variant (LCV). A region of >240 kb has been found to exist as 2 to 12 copies in diploid genomes. (**C**) One type of variation in the human genome is a difference in gene copy number. An increase in copy number is normally correlated with an increase in expression of the gene. A classic example is the *CYP2D6* gene, where up to 12 copies of the gene have been found in individuals with extremely high CYP2D6 activity.

disorders.'' In most cases genomic disorders arise due to errors occurring in the chromosomal recombination process. Molecular genetic studies have now led to the precise positioning of chromosomal breakpoints for many of these genomic disorders. While the regions affected by the rearrangements are mostly made up of unique sequence, the intervals immediately surrounding the breakpoints often show a high degree of DNA sequence similarity. These findings indicate that the majority of recurrent rearrangements are caused by misalignment of nearly identical sequences in the genome. This type of mechanism of aberrant recombination is referred to as non-allelic homologous recombination (NAHR) (1). The rearrangements caused by NAHR include deletions, duplications, inversions, or more complex combinations of these rearrangements.

Homologous sequences have also been found near the breakpoints for the few known examples of recurrent translocations, suggesting that similar mechanisms exist for interchromosomal recombination. Large deletions and duplications will give rise to genomic disorders when the gene(s) located within the affected region are dosage sensitive. Inversions and translocations do not create differences in copy number of the genes in the genomic segment affected, but may cause disease by directly disrupting genes or acting on them indirectly by dysregulating their control elements in the breakpoint regions.

2.1. Recurrent Genomic Rearrangements

Large stretches of highly homologous DNA sequences usually referred to as segmental duplications or low copy repeats or duplicons have been shown to be the primary cause for NAHR. The type of rearrangement caused by segmental duplications depends on the architecture of the specific sequences involved. If the direction of the duplicated sequence is the same, it will result in either deletion or duplication, while inverted segmental duplications give rise to inversion. One of the best-studied examples of how low-copy repeat regions having a simple organization gives rise to genomic disorders is based on extensive studies of the p12 band of chromosome 17. Rearrangements in this region can cause one of two diseases, namely, Charcot-Marie-Tooth disease type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP). CMT1A is caused by a 1.4 Mb genomic duplication, while HNPP is caused by a deletion of the equivalent region (2). Molecular studies indicate that these diseases are the reciprocal products of NAHR events between two flanking 24 kb segmental duplications sharing 98.7% sequence identity (3). Similar mechanisms cause two genomic disorders on 17p11.2, Smith–Magenis Syndrome (SMS) and dup(17)(p11.2p11.2) syndrome. In these cases a large DNA segment about 3.7 Mb in size, which is deleted in SMS and duplicated in dup(17)(p11.2p11.2) patients, is flanked by long duplicated sequences of $\sim$200 kb arranged in the same orientation along

the chromosome (4). However, the organization of segmental duplications is often complex with several sub-regions organized in different directions compared to the ancestral segments. Series of duplicated regions may therefore give rise to a number of different size deletions, duplications, or more complex rearrangements causing different disorders. The most studied such region is located at 22q11 and it contains a number of segmental duplications of complex organization, which are now known to be involved in deletions, duplications, inversions, translocations, and marker chromosomes (5,6).

2.2. Deletions and Duplications

The most common rearrangement caused by illegitimate recombination triggered by the presence of segmental duplications is chromosomal deletion of the unique sequence. If reciprocal exchanges occurred equally as would be predicted from modeling an equal proportion of deletions and duplications should be observed. The data indicate, however, that deletions are vastly over-represented. Disorders arising due to recurrent deletions are referred to as microdeletion syndromes. True microdeletion syndromes are caused by haploinsufficiency of several genes. One of the most common microdeletion syndromes is DiGeorge Syndrome/Velocardiofacial Syndrome (DGS/VCFS), with a prevalence of 1.3–1.5/10.000 births (7,8). The deletion is induced by low copy repeats located at 22q11.2. The frequency of de novo deletions in this region is at least one magnitude higher than de novo point mutation rates for human autosomal dominant disorders.

When large deleted regions including several genes give rise to a complex clinical phenotype, it can often be difficult to distinguish which of the genes in the segment contribute to the phenotype. Williams–Beuren syndrome (WBS) offers one such example. WBS is caused by a 1.6 Mb deletion at 7q11.2 region, which contains more than 20 genes (9). The phenotype of WBS displays unique facial features, growth retardation, infant hypercalcemia, and supravalvular aortic stenosis. Patients with WBS have a distinctive personality and are highly sociable, musical, and talk remarkably well, but have severely impaired visuospatial constructive abilities. It has been shown that disruption of only the elastin gene, harbored within the region commonly deleted in WBS patients, gives rise to dominant supravalvular aortic stenosis, thus explaining part of the phenotype (10–12). Using a similar "genotype–phenotype" correlation additional disease contributing genes in the region are being discerned for other sub-phenotypes.

One of the most recent additions to the family of genomic disorders is Kabuki Syndrome, a form of congenital mental retardation with a phenotype that appears to be genetic in etiology. However, little evidence of a genetic abnormality, either chromosomal or Mendelian, had previously been established. Recently a duplication of a 3.5 Mb region was found at

8p23.1-p22 in six unrelated patients, suggesting that this may be the common cause for this disorder (13).

Parkinson disease is a late-onset disease and the second most common neurodegenerative disorder after Alzheimer disease, affecting approximately 1% of the population over age 50. In 1997 kindreds with a familial form of Parkinson disease were shown to carry mutations in the alpha-synuclein gene (*SNCA*) (14). A study last year reported on a family showing linkage to the SNCA locus but where no disease causing mutations could be found. Further investigations revealed that the affected family members all carried a duplication of a ~2 Mb stretch of DNA spanning the *SNCA* gene (15). This finding not only suggests a mechanism of the role of SNCA in Parkinson disease, but also that large duplications and deletions may more often than originally thought be involved in disease etiology in families where mutation screening of causative genes appears negative.

2.3. Inversions and Translocations

Genomic disorders caused by inversions and translocations are far fewer in occurrence than those arising due to deletions and duplications. In most cases, inversions and translocations normally do not cause disease by leading to a direct gain or loss of dosage sensitive genes, but the genomic lesions may disrupt genes residing at the respective breakpoints. Approximately 40% of all hemophilia A patients carry an inversion at chromosomal region Xq28 (16,17). This recurrent inversion spans ~400 kb and is mediated by two inverted segmental duplications, one of which is located in intron 22 of the factor VIII gene, with two other copies being located approximately 400 kb telomeric of the same gene. When the inversion occurs, it disrupts the factor VIII gene, giving rise to hemophilia A.

Recurrent de novo translocations are rarely observed in the human genome with only three so far being described in any detail. Two of these events include the region on chromosome 22q, which is also involved in causing the deletions and duplication leading to DiGeorge and velocardiofacial syndromes. The most frequent translocation is called t(11q;22q). The breakpoints on both chromosome 11q23 and chromosome 22q11 are clustered in multiple unrelated families (18,19). Both breakpoints have been shown to contain palindromic AT-rich sequences that can mediate hairpin structure formation (20,21) suggesting they mediate t(11q;22q). Moreover, the same region on chromosome 22 and a similar mechanism of rearrangement has been suggested to underlie the t(17q;22q) translocation (22) that involves the neurofibromatosis (*NF1*) gene on chromosome 17. This translocation has been found in a number of patients with neurofibromatosis type I. The third translocation that has been reported numerous times in the literature is the t(4;8)(p16;p23) translocation (23). Two pairs of olfactory receptor gene clusters are located in close vicinity to each other on

two separate chromosomes (4p16 and 8p23). The olfactory clusters have evolved by gene duplication events and thus share high sequence similarity. As would be predicted the breakpoints of the translocations were found to be located within these olfactory genes in all translocation carriers identified.

2.4. Gene Conversion

Many human genes have highly related family members existing as functional genes or pseudogenes that can be clustered or dispersed elsewhere in the genome. Unequal crossing over between a functional gene and a related pseudogene can result in the deletion of the functional gene, or the formation of fusion genes with one part derived from the pseudogene. However, similar DNA sequences may also give rise to "gene conversion" events. Gene conversion is the nonreciprocal transfer of sequence information between a pair of sequences. For example, one (donor) sequence remains unchanged while the other (acceptor) sequence is replaced by the nucleotide content of the donor. It has been suggested that this mechanism involves mismatch repair of a heteroduplex involving highly similar sequences. The mismatch repair system identifies the nucleotide bases that are not perfectly matched in the heteroduplex and changes them such that both copies are identical. This unusual recombinational outcome has been shown to be a recurrent mutational mechanism in some diseases. A gene conversion event may include a functional gene and its pseudogene copy, which carries a premature stop codon. In this case gene conversion would transfer the stop codon mutation to the functional gene, thereby disrupting the functional copy. The classical example of pathogenesis caused by gene conversion between a gene and its pseudogene is steroid 21-hydroxylase deficiency. More than 95% of mutations causing this disease are caused by sequence exchange between *CYP21B*, and the closely related, tandemly duplicated pseudogene *CYP21A* (24). The same mechanism has been shown to be involved in causing Schwachman–Diamond syndrome, where stop codons are created in the *SBDS* gene due to gene conversion with its 97% similar pseudogene copy (25). In the *SBDS* example the gene and pseudogene are part of non-contiguous larger segmental duplications located on the long arm of chromosome 7.

2.5. Balanced Genomic Variants Can Induce Rearrangements

One interesting explanation for a few of the genomic disorders caused by deletions is that one of the parents of the proband carries a balanced genomic variant. These balanced genomic variants then appear to predispose to unbalanced chromosomal rearrangements associated with disease in the offspring. As such, in several disorders caused by large deletions, inversion of the exact same fragment that is deleted in the patient has been found in one of the parents. For example, an inversion of the 15q11-q13 region, which is

deleted in patients with Angelman syndrome, has been identified in 2/3 of mothers of patients with deletions. Interestingly, this inversion is found in 9% of the general population (26). Similarly, an approximately 1.9 Mb inversion has been found in 1/3 of all transmitted chromosomes of the deletion that causes Williams–Beuren syndrome. This inversion is not so common (estimated at less than 5%) in the general population, but seems to predispose to deletions in the offspring (27). Inversions may also lead to duplications in the offspring. The recently identified duplication at 8p23.1-p22 in patients with Kabuki syndrome may also be mediated by a submicroscopic inversion. In two out of six cases, heterozygous 8p23.1 inversions were found in transmitting mothers of patients with duplications (13). One of the few recurrent translocations, the t(4;8)(p16;p23), may also be mediated by inversions in the transmitting parent. Heterozygous inversions at both chromosomal loci were found in all mothers of five de novo translocation patients (23).

2.6. Non-Recurrent Genomic Rearrangements

Long stretches of nearly identical sequences (>90% DNA sequence identity) seem to be the primary sequence motifs inducing recurrent rearrangements in the genome. However, in many cases of genome rearrangement, the region involved is unique. The mechanisms causing these rare non-recurrent events have also been studied. In a study by Stankiewicz et al.(28), the authors investigated whether the highly homologous regions (i.e., sequences that are 90–100% related at the nucleotide level) could be the explanation also for non-recurrent deletions and translocations. The results did indeed suggest that highly homologous sequences are often involved in large non-recurrent deletions, but are not a common cause for translocations. There are many regions in the genome that show high sequence similarities that do not fulfill the criteria normally used to describe segmental duplications. These include members of gene families, pseudogenes, and common small repetitive elements. There are a number of examples where pseudogenes or homologous functional genes are involved in recombination events creating subsequent rearrangements. One example of this is Hunter syndrome, where more than 10% of patients carry an inversion created by non-homologous recombination between the *IDS* gene and its pseudogene copy (29,30).

Several genetic diseases are caused by non-recurrent rearrangements. It has been speculated that these rearrangements can be caused by clusters of common repetitive elements. The most common repeat sequences in the human genome are ALU repeats, with approximately 500,000 copies in the human genome (31). ALU repeats are ~280 bp in length and are usually flanked by short direct repeats of 6 to 18 bp. The 45-kb low-density lipoprotein receptor gene has an ALU repeat present, on average, every 1.6 kb. The very high frequency of pathogenic deletions in the LDL receptor

gene has been shown to coincide with ALU repeat sequences at both breakpoints (32). Another disease caused by non-recurrent deletions and duplications is Duchenne muscular dystrophy (DMD). Approximately two thirds of DMD cases are caused by deletions of one or more exons of the large (>2 Mb) *DMD* gene. Sequencing of the regions flanking the deletions in patients with DMD showed that 30% of deletion occurred at common repeat sequences (ALU and LTRs) (33), while the remaining deletions were flanked by unique sequence. Some repeat sequences in the human genome are known to be transposable elements. This means that these elements can be copied and inserted in new genomic regions, usually via an RNA intermediate. One mechanism by which these transposable elements can cause disease is if they are inserted into a functional gene, causing a disruption of the transcript. This phenomenon is referred to as insertional inactivation. Obviously, this is more likely to occur if the gene is large, as is the case for the neurofibromatosis I and hemophilia A loci (34,35).

3. SEGMENTAL DUPLICATIONS

It has been known for some time that a large proportion of genomic disorders arise from rearrangements in the genome that is due to aberrant recombination. As discussed above, the major cause for these recombinational errors is the misalignment of long stretches of nearly identical DNA sequences. Analysis of the human genome sequence shows that a large proportion of the genome consists of segmental duplications, also known as low copy repeats (36,37). One of the most important mechanisms for genome evolution and creation of new genes is by duplication of large segments of genetic material. Duplicated copies of genes can then mutate and obtain new or specialized functions. It is therefore not surprising that duplicated segments exist in the human genome. However, the abundance of such sequences in the human genome has only recently been appreciated. Since the completion of the human genome sequence, several analyses show that approximately 5%, or 150 Mb of the genome exists as segmental duplications having at least 90% DNA sequence identity (36,37). This statistic does not include common repetitive elements or short (<1 kb) repeated sequences, which would make the proportion of repeated sequences up to >50% of the genome (38–40). The high sequence identity indicates that these sequences have been created by recent duplications of large stretches of genomic DNA. The size of duplication segments ranges from a few kilobases to a few megabases. Importantly, it is not rare for duplication regions to contain several genes. For example, the human specific 500 kb inverted duplication on 5q13 (41) contains six different genes, including the *SMN1* gene involved in spinal muscular atrophy (42).

Segmental duplications are normally divided into two classes depending on their chromosomal relationship to each other (referred to as being

intrachromosomal or interchromosomal). The intrachromosomal duplications are chromosome-specific and are often clustered in the same region of the chromosome and in some cases even as tandem copies. Interchromosomal duplications are found on non-homologous chromosomes and are less common than the intrachromosomal class. Interchromosomal duplications appear to cluster in pericentromeric and subtelomeric heterochromatic regions of chromosomes (43,44). Due to their high identity segmental duplications have caused problems for the successive assemblies of the human DNA sequence. Tandemly duplicated segments of very high sequence identity are especially difficult to identify. There is a strong correlation between the location of segmental duplications and the remaining gaps in the human sequence assembly, and it is expected that the proportion of segmental duplications in the genome will increase as the assembly of the "complete" genome is finalized.

4. COMMON GENOMIC POLYMORPHISMS

In previous sections we have described the gain or loss of gene copies, which are rare and lead to major phenotypic effects. There are, however, a class of genomic alterations leading to gene copy number differences that lead to less significant phenotypes and which occur more frequently and in some cases in $>1\%$ of the population (indicating they are polymorphisms). A list of some common genomic polymorphisms is shown in Table 1. Many of these variants have already been shown to be directly related to gene expression. With new techniques available to scan the genome for copy number changes, reports of large genomic variants are increasing in frequency. It seems as if the findings of such variants are no longer chance findings of very rare occurrences in the genome, but may be quite common in the normal population. Detailed studies of genes where a variable number of gene copies have been found in the normal population has shown a concomitant increase in mRNA from that gene.

The first indication that a previously unappreciated number of large-scale polymorphism exist in the human genome, as well as the genomes of other species, was a study comparing the copy number of $\sim$29.000 genes in several primate species (69). The data showed a large number of gains or losses of genes that were lineage specific. A substantial number of gene copy number changes were shown to have occurred during the last 5 million years after human divergence from our closest ancestor, the chimpanzee. The data indicate that gene duplication and deletion are strong forces in shaping the genome, and that it is a process that is even more pronounced in the human genome than some of our closest living relatives. Although only five humans were included in the study, close examination of the data revealed several genes that displayed copy number differences within the human population. Using a different approach, Fredman et al.(70) investigated SNPs

Table 1 Large-Scale Polymorphisms in the Human Genome that Have Been Studied in Detail

Locus	Gene/s/	Gene symbol	Size of region	Frequency of polymorphism	Associated phenotype	References
1p21.1	Amylase genes	AMY1A, AMY1B	(100kb)*n*, $n = 1$–3	40%	None known	45
1p13.3	Glutathione-S-transferase M1	GSTM1	18 kb	>3% dup, >70% del	Ultra rapid enzyme activity	46
4p16	Multiple genes	Multiple genes	6 Mb	12.5% inversion	None known	23
5q13.2	Survival of motor neuron 1, 2	SMN1	Entire genes	4% dup	Spinal muscular atrophy	47 and 48
6p21.3	Cytochrome P450 CYP21A2	CYP21A2	32.7 or 26.4 kb	1.6% dup	Congenital adrenal hyperplasia	49
6p21.3	Complement component C4	C4A, C4B	32.7 or 26.4 kb	97% >2 copies	Infection and immunity	50
8p23	Multiple genes	Multiple genes	2.5–5.3 Mb	26% inversion	Panic disorder?	51
8p23.1	Defensin gene cluster	Multiple genes	(240kb)*n*, $n = 2$–12	78% = 3 or 4 copies	None known	52
8q24.23	No known genes	No known genes	200 kb	10% del	None known	53
12p13.2	Killer cell lectin-like receptor 2C	NKG2C	16 kb	20% del	None known	54
14q32.33	Ig heavy variable cluster	VH26	Entire gene	4–15% dup	Ig mediated Immunity?	55
c15q	NF1, IgH D Pseudogenes	No known genes	1 Mb	80% dup	None known	56
15q11-q13	Six genes	Multiple genes	4–5 Mb	9% inversion	Angelman syndrome	26
15q24–26	60 genes	Multiple genes	17 Mb	7%	Panic disorder	57

(*Continued*)

Table 1 Large-Scale Polymorphisms in the Human Genome that Have Been Studied in Detail (*Continued*)

Locus	Gene/s/	Gene symbol	Size of region	Frequency of polymorphism	Associated phenotype	References
17q11-q22	Chemokine ligand 3 and 4	CCL3, CCL4	Entire gene >2 kb	32% dup, 27% dup	Response to infection?	58
19p13.3–13.2	Cytosine 5-methyltransferase	DNMT1	Entire gene >60 kb	93% dup	Neoplastic transformation?	59
19q13.2	Cytochrome P450 CYP2A6	CYP2A6	>7 kb	1.7% dup	Nicotine metabolism	60
22q11.2	Multiple genes	Multiple genes	110 kb	20–44% dup	IgA mediated immunity?	61 and 62
22q11.2	Immunoglobulin l	Igl (IGL)	5.4 kb	28–85% dup	Altered Igk:Igl in B lymphocytes	63
22q11.23	Glutathione-S-transferase T1	GSTT1	54 kb	66% del	Cancer	64
22q13.1	Cytochrome P450 CYP2D6	CYP2D6	Entire gene	1–29% dup, 3–5% del	Resistance to tricyclic antidepressants	65–67
Xq28	Emerin, Filamin 1	EMD, FLN1	48 kb	30% inversion	Emery–Dreifuss muscular dystrophy	68

For several of these regions, there is a high variability in the number of copies between individuals. Associated phenotype refers to the association with genes in the variable region and is not necessarily caused by the polymorphism described in the table. For a comprehensive summary of all large-scale variants described in literature, see http://projects.tcag.ca/variation/.

in segmental duplications. By measuring the ratio between the two alleles for the SNP, they found differences between individuals that could only be explained by either copy number differences or gene conversion. The regions were then investigated using quantitative methods and it was shown that a number of the regions exhibited variable number of copies in the normal populations. Their data indicated that segmental duplications are highly dynamic regions in the genome, often involving copy number differences and recurrent gene conversion events. Although only a small number of copy number variants were verified in the study, it pointed to the fact that this type of variation is substantial.

The extent of large-scale variation in the human genome was investigated in two independent studies published in the summer of 2004. Using a method called ROMA (representational oligonucleotide microarray analysis) Sebat et al. (71) found 80 regions that show copy number variation in healthy, unrelated individuals. The average size of these regions was reported to be ~400 kb and many contain entire coding transcripts. They also showed that these variants were unlikely to be somatic variation by testing DNA from a number of different tissues from the same individuals. The only somatic changes were shown to occur in the T-cell receptor clusters and immunoglobulin related genes. Although the coverage of the human genome is very high using the ROMA method, all regions that are not unique in the genome are excluded from analysis. Since the study by Fredman et al. (70) indicate that segmental duplication regions are likely to be the most dynamic regions in the genome, the findings by Sebat et al. (71) is likely an under-representation of the true variation content of the human genome. In a separate study, Iafrate et al. (72) used CGH (comparative genomic hybridization) to investigate the extent of large-scale variation in the human genome and found 250 regions that displayed copy number variation between 55 unrelated individuals. The method is based on hybridization to genomic clones, indicating that the size of these variants are >50 kb. Many of the regions found to be polymorphic in the population were shown to occur in regions with segmental duplication or gaps in the human genome assembly, reinforcing the role these regions play in genome rearrangement and evolution. A database of large-scale genomic variants has now been created by the authors (http://projects.tcag.ca/variation/) (72). This is a valuable resource for scientists screening for human genome rearrangements in correlation to disease or normal variation. Overall, the results of these studies suggest that the human genome is much more dynamic in nature than was previously thought. In the near future, more detailed information about the extent and frequency of large-scale genomic variants should become available. It will then be important to link these variants to variation in transcription both in the normal population and in disease. These findings may therefore represent an important step towards the understanding of complex genetic disease.

5. GENE COPY NUMBER VARIATION

There are several types of genes that have been shown to display copy number differences in the genome, many being highly relevant to the pharmacologic responses. One class that is conspicuously over-represented are genes involved in the defense against pathogenic or toxic agents, where a number of genes display a high degree of polymorphic variation which includes variation in gene copy number: glutathione-S-transferase (*GST*) genes, cytochrome P450 (*CYP*) genes, the complement component C4, and the defensin genes. In each case, changes to gene copy number have been shown to give rise to concomitant changes in the level of enzyme activity, with phenotypic consequences.

5.1. Glutathione-S-Transferase Genes (*GST*)

The GSTs are phase II enzymes involved in the metabolism of a large range of endogenous and exogenous toxic compounds. Individuals can have large variations in GST activity, which can have a major impact in the sensitivity of cells to these toxins (73,74). This is known, or postulated to result in differences in individual susceptibility to illnesses related to such substances including various carcinomas. In addition, the efficacy of therapeutic drugs is affected by individual variation in the ability of these enzymes to metabolize them (73) with individuals being described as low or high responders. There are two families of *GST* genes: cytosolic and microsomal. There are at least 16 cytosolic GSTs, which are encoded by several classes of genes based on their biochemical, immunological, and structural properties, giving rise to at least 16 different proteins. A further six genes are known to encode microsomal GST. Of these, two genes display polymorphic variations in copy number (*GSTM1* and *GSTT1*) whilst one (*GSTT2*) has two copies in all individuals.

5.2. Microsomal Glutathione-S-Transferase 1 (GSTM1)

The predominant *GSTM1* allele is a null allele where an 18-kb fragment encompassing the entire gene is deleted following unequal crossing over between break points 5 kb downstream from the *GSTM2* gene and 5 kb downstream from the *GSTM1* gene. In fact, *GSTM1* is found in only half of the genomes of Caucasians and Saudi Arabians, the other half of the populations exist with no copy of the gene in their genome. The lack of this gene may predispose to some cancers (46). Most other individuals have either one or two copies, however, some people who have ultra rapid *GSTM1* activity have been found to have three copies of the gene (46). It is unclear if these people are more resistant to cancer, but it is likely that are more resistant to the effects of some toxic compounds in the environment. It is also likely that some individuals who have two copies of the gene

may have both of those copies on one parental chromosome, having one null allele and one duplicated allele.

5.3. Glutathione-S-Transferase Theta 1 (GSTT1)

The gene for *GSTT1* has a null allele, which results from deletion of a 54-kb fragment from 22q11.23, which includes the entire gene (64). The deletion occurs as a result of homologous crossover between two 18 kb flanking sequences, HA3 and HA5, which have greater than 90% sequence homology. This deletion allele is common in the population with 20% of individuals having no copies of *GSTT1*, while 46% are heterozygous and have a single copy of the gene, which fits the Hardy–Weinberg equilibrium. The null allele has been associated with a number of carcinomas (75) however, other research suggests that the effects are mainly seen in sub-populations (74).

5.4. Cytochrome p450

Cytochrome p450 (CYP) is a superfamily of heme-containing phase 1 enzymes (http://www.imm.ki.se/CYPalleles/), which metabolize both endogenous compounds such as steroids, fatty acids, and prostaglandins, and exogenous compounds such as environmental carcinogens, pollutants, drugs, and other xenobiotics (76). At least three genes (*CYP2D6*, *CYP2A6*, and *CYP21A2*) are present in variable copy number.

The *CYP2D6* (22q13.1) gene product (cytochrome P450 CYP2D6) is a highly polymorphic enzyme involved in the oxidative metabolism of many different classes of commonly used drugs, including antidepressants, antipsychotics, beta-blockers, opiates etc. About 7–10% of Northern Europeans have low CYP2D6 activity and are classed as poor metabolizers. These patients may develop adverse drug reactions when treated with standard doses of drug. In contrast, some individuals are ultrarapid metabolizers, and require higher drug dosages in order to achieve the required response. The enzyme is highly polymorphic and the alleles give widely varying enzyme activity, however, another source of variation in activity is changes to gene copy number (77).

A number of studies have shown that many of the alleles are duplicated in a polymorphic manner although the frequency of duplication varies considerably between populations: 1% and 7% in Caucasian and Japanese populations but 21% and 29% in Saudi Arabian and Ethiopian populations, respectively (65,66). The number of gene copies varies up to 13 and increased gene copy number results in increased gene expression and subsequent increased catabolism of drugs by the gene product roughly in proportion to the gene dosage (67). The mechanism for the duplication to give up to five copies of the gene has been proposed to be unequal crossover between Alu repeat sequences (78). Larger numbers of gene copies may have been formed following unequal segregation and/or rolling circle replication followed

by homologous recombination, however, there is little evidence for this theory.

The CYP2A6 is a major hepatic member of the CYP family in humans, which metabolizes pharmaceutical agents including nicotine to the inactive cotinine. A number of reports suggest an association between smoking and null alleles including a deletion allele. The activity of this enzyme is related to the number of cigarettes people smoke; individuals with less active enzyme have more nicotine in their blood for the same number of cigarettes smoked and thus smoke less. Compared with individuals with the same *CYP2A6* alleles, individuals with duplication of *CYP2A6* on a single chromosome showed increased activity of the enzyme of approximately 1.5-fold and had increased blood carbon monoxide levels, the latter being a measure of the amount of tobacco smoked (60).

A deletion allele is very common in Japanese (20% of alleles) Korean (11%), and Chinese (15%) populations (79,80). This deletion is the result of a homologous unequal crossover between the 3′ flanking regions of the *CYP2A7* and *CYP2A6* genes, which result in the deletion of the entire *A6* gene (81). This should lead to the concomitant formation of a tandem duplication of the *A6* gene (80). Although this has been found, the frequency is not equivalent. Duplication of *CYP2A6* (19q13.2) has been found in 1.7% of a smoking population and 1.3% of a normal population both of Caucasian origin (60). A single individual with the duplication was also found in a Chinese population (1 out 114; 0.9%) (82). This may suggest that duplication does not occur frequently, and the deletion occurs through a different mechanism, such as the formation of a loop, or alternatively the duplication may not be detected as southern blots may not show the presence of the extra copy. A novel deletion allele, present in less than 1% of a Japanese population has been described, in which a larger section of DNA has been deleted (83).

The *CYP21A2* gene and the C4 complement component genes (*C4A* or *C4B*) are located close together on 6p21.3 near a highly variable region. The *C4* gene is highly polymorphic with at least 34 allotypes known to exist. These are grouped into two isotypes, *C4A* and *C4B*, which vary in activity; the C4A has a slow reaction rate but a long half-life while the C4B has a fast reaction rate and short half-life. Deficiency of C4 increases the susceptibility or severity of viral and bacterial infections and is also an important risk factor for autoimmune diseases, such as systemic lupus erythematosus. Conversely, excessive C4 or over-activation of C4 could aggravate an inflammatory response and render an individual more vulnerable to tissue injuries. The variable region is bounded on the telomeric side by the genes *RP1* and *C4* (*A* or *B*) and on the centromeric side by the genes *CYP21A2* and *TNXB*. Within these boundaries a unit (termed RCCX) consisting of four genes, may be found in variable copy number typically 0, 1, or 2. The four genes are *CYP21A1P* (a pseudogene of *CYP21A2*), *TNXA*, *RP2*,

and *C4* (*A* or *B*). Thus an individual may have between 2 and 6 copies of the *C4* and the *CYP21A1P* genes. The mechanism leading to the variations is probably unequal crossing over between *C4* genes (50).

The *CYP21A2* is essential for the biosynthesis of glucocorticoid and mineralocorticoid hormones. Gene conversion whereby either the entire active gene or part of it is replaced with corresponding sequences from the pseudogene gives rise to a defective *CYP21A2* gene. Low activity CYP21 activity is the primary cause of congenital adrenal hyperplasia, the severity of symptoms being associated with the enzyme activity level ranging from androgen excess through simple iridizing down to salt wasting in the absence of the gene.

The RCCX unit almost invariably carries the *CYP21A1P* pseudogene and extra copies of the active gene are rare; the frequency has been assessed in a Northern European population and found to be present at a frequency of 1.6% (49) and duplications have also been found in other populations. Although these additional copies of *CYP21A2* are invariably defective, mutational analysis of this gene is often carried out as part of the diagnostic process for congenital adrenal hyperplasia and duplication complicates this analysis, as one *CYP21A2* gene copy may be defective due to partial replacement with pseudogene sequence, whilst the other copy remains intact. This may lead to the misassignment of carrier status.

5.5. β-Defensin Genes

Defensin genes encode small antimicrobial peptides and are an important part of the immune system. The α-defensins are highly expressed in neutrophils, while β-defensins are expressed in a variety of epithelia, especially the airways, and have been shown to have broad antimicrobial properties. All the defensin genes are located in a cluster at 8p23.1, a region known to be frequently involved in chromosomal rearrangements. In the α-defensin family, the highly identical *DEFA1* and *DEFA3* genes have been shown to have variable number of copies in the normal population (84). The *DEFA3* gene currently maps to four places near a gap in the human genome assembly. It was recently shown that three genes in the β-defensin gene cluster, *DEFB4*, *DEFB103*, and *DEFB104*, are located in a 240 kb repeat unit that is polymorphic in the population (52). Individuals carry 2 to 12 copies of this repeat unit in their genomes. The highest number identified on a single chromosome was eight repeats, indicating that individuals with more than 12 copies in total are likely to be found. Quantitative analysis of mRNA from the *DEFB4* gene shows that there is a direct correlation between the copy number and the level of expression of this gene. These defensin polymorphisms have been shown to be the basis for a previously described euchromatic variant in the same region.

6. METHODS FOR STUDYING GENOMIC VARIATION

Despite its fundamental importance for molecular genetics, it is only very recently that effective methods have been developed for studies of submicroscopic gains or losses of specific genomic loci. There is still no method available for genome-wide investigation of inversion variants. Many studies require a higher resolution than what can be attained by using classical fluorescence in situ hybridization (FISH) or comparative genomic hybridization (CGH). Several human diseases have been shown to be caused by small deletions or duplications, involving single exons or parts of genes. The detection of such small copy number variants has previously been limited to quantitative Southern blot, and more recently, real-time PCR. However, neither of these methods is ideal for screening multiple loci in large cohorts of patients or control groups. Several methods have now been developed to meet the demand for techniques to perform analyses of duplications and deletions in a high throughput setting.

6.1. Array Based Methods to Identify Copy Number Differences

The Comparative Genomic Hybridization (CGH) method was originally used for identification of copy number imbalances in metaphase spreads (85). Two differentially labeled DNA samples are hybridized to normal chromosomes and the ratio of fluorescence intensity is measured. Differences in fluorescence are indicative of relative copy numbers in the two DNA samples. However, the resolution of the method was initially limited to several megabases only. Development of microarray technology and the subsequent adaptation of CGH to array format dramatically increased the resolution of the method (86). The basic principle of array-CGH is the same as the original CGH, but the DNA is now hybridized to glass slides containing DNA clones instead of metaphase chromosomes. Array-CGH has been used successfully with both cDNA and genomic clone arrays (87). Although the resolution of the method has improved significantly it is still limited to a lower limit in the kilobase(s) range due to sensitivity of hybridization and detection. Commercially available genome-wide arrays have $\sim$1 Mb resolution (Spectral Genomics, Houston, Texas). Chromosome-specific arrays with contiguous coverage and 75 kb resolution have been successfully implemented (88). A novel approach was recently described, where oligonucleotide probes were used. This method, named ROMA (representational oligonucleotide microarray analysis), achieved an average resolution of 30 kb throughout the genome using 70-mer probes on the array (89). It is likely that the resolution can be improved even further using oligonucleotide arrays. Until recently, CGH has predominantly been used for identification of chromosomal aberrations in tumors. With higher resolution arrays available the use of CGH technology is becoming more widespread. The major advantage of the method is that the whole genome of an individual can

be analyzed for chromosomal imbalances at a comparatively high resolution in a single experiment.

Another type of array-based approach that can be used to identify deletions and duplications is to use whole genome sampling analysis of SNP arrays (e.g., the Affymetrix 10 or 100 k SNP chip) (90). Not only can deletions be identified as contiguous stretches of homozygous markers, but also the SNP discrimination ratios and the perfect match intensities can be normalized across the chip and compared to a reference set of individuals. This way, loss and gain of genetic material can be distinguished in addition to the collection of SNP genotype data.

6.2. Quantitative Multiplex PCR

Real-time quantitative PCR is most commonly used for analysis of differences in gene expression. However, the same principles for quantification can be applied to genomic DNA. Both the 5′-nuclease assay with dual-labeled Taqman probe, and the SYBR-Green I assay formats, have been successfully used to screen for genomic DNA changes in copy number (91,92). However, neither of these approaches is well suited for multiplexing and cost becomes an issue for high-throughput screening. Several other multiplex PCR based methods have therefore been developed for the purpose of screening for copy number differences. These can be summarized by the name quantitative multiplex PCR of short fluorescent fragments (QMPSF) (93). PCRs are designed to span the regions of interest, e.g., exons of a specific gene. In its simplest form, the forward primers are labeled with a fluorescent 6-FAM moiety, and a PCR is run for ~20 cycles until it is in the quantitative range. The product is then purified and run on an automated sequencer and the electropherograms from the individuals to be compared are superimposed. A control sequence that is not expected to be duplicated or deleted is included as a DNA quality and quantity control. More recent versions of this methods includes the addition of chimeric primers, i.e., primers carrying an extension of nucleotides in their 5′-end, rendering the effective annealing temperatures more homogeneous in a multiplex PCR after the first few cycles (94). The use of chimeric primers thereby reduces the amount of optimization needed to find adequate primer concentrations to yield homogeneous fluorescent peak levels. Chimeric primers can also be used to simplify the assay and reduce cost. Normally, each forward primer has to be labeled with a 6-FAM fluorescein. However, if identical 16-mer extensions are added as part of the chimeric primer, this sequence can be used as a template for a universal 6-FAM labeled primer that is added before the last two cycles of the PCR reaction, thereby labeling all PCR products using a single FAM-labeled primer (95). Adding this step substantially reduces the cost of the assay.

6.3. Multiplex Probe Amplification Assays

One approach that has proved successful for identifying copy number differences in the genome is based on using amplifiable probes. The initial version of this methodology was called multiplex amplifiable probe hybridization (MAPH) (96). The test DNA is denatured and bound to a nylon filter and hybridized with amplifiable probes, each recognizing a unique DNA sequence. All probes are of unique length and carry identical end sequences and can thus be amplified in a subsequent step using universal primers that are fluorescently or radioactively labeled. Products are then separated and quantified based on their radioactive or fluorescent intensities compared to control samples. Multiplexing has been performed with up to 40 probes in a single reaction using MAPH (96). A technique based on similar principles is multiplex ligation-dependent probe amplification (MLPA) (97). This assay is performed in solution. Two probes are annealed adjacently to target regions and then ligated. Each probe pair gives rise to a product of unique length, but has identical end sequences, permitting simultaneous PCR amplification with universal primers. One primer is fluorescently labeled allowing gel based separation and quantification. The fact that the method is dependent on ligation also makes it possible to target one specific allele when SNPs are present. The drawback of this method is the requirement for substantial optimization. However, when a set of probes has been optimized they can be used to quickly and effectively score large numbers of samples.

7. SUMMARY

It has been known for a long time that genetic diseases are often caused by rearrangements of large segments of genomic sequence. Many of these rearrangements occur due to NAHR between segmental duplications. The mechanisms causing recurrent rearrangements are now better understood, and this could give clues to identifying hotspots for genomic rearrangements. A wide array of new methodologies to study gain and loss of genomic sequences has led the way to genome-wide studies of large-scale variants in disease as well as in the genomes of healthy individuals. The results of these studies indicate that the genome is much more variable between individuals than was previously suspected, with each individual carrying a number of large genomic polymorphisms. Many of the large variants encompass entire genes, indicating that this could be a primary source for variation in gene expression between individuals. This previously unappreciated source of variation has opened the door to new ways of thinking about disease mechanisms and genome evolution. In the near future we are confident that we will not only get a high resolution map a large-scale variation in the human genome, but we will also see a number of these variants being linked to specific phenotypic traits.

REFERENCES

1. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 1998; 14(10): 417–422.
2. Chance PF, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, Lupski JR. Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. Hum Mol Genet 1994; 3(2):223–228.
3. Potocki L, Chen KS, Park SS, Osterholm DE, Withers MA, Kimonis V, Summers AM, Meschino WS, Anyane-Yeboa K, Kashork CD, Shaffer LG, Lupski JR. Molecular mechanism for duplication 17p11.2—the homologous recombination reciprocal of the Smith-Magenis microdeletion. Nat Genet 2000; 24(1):84–87.
4. Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. Nat Genet 1997; 17(2):154–163.
5. Edelmann L, Pandita RK, Spiteri E, Funke B, Goldberg R, Palanisamy N, Chaganti RS, Magenis E, Shprintzen RJ, Morrow BE. A common molecular basis for rearrangement disorders on chromosome 22q11. Hum Mol Genet 1999; 8(7):1157–1167.
6. Shaikh TH, Kurahashi H, Emanuel BS. Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: an update and literature review. Genet Med 2001; 3(1):6–13.
7. Tezenas Du Montcel S, Mendizabai H, Ayme S, Levy A, Philip N. Prevalence of 22q11 microdeletion. J Med Genet 1996; 33(8):719.
8. Liling J, Cross I, Burn J, Daniel CP, Tawn EJ, Parker L. Frequency and predictive value of 22q11 deletion. J Med Genet 1999; 36(10):794–795.
9. Tassabehji M. Williams-Beuren syndrome: a challenge for genotype–phenotype correlations. Hum Mol Genet 2003; 12(2):R229–R237.
10. Curran ME, Atkinson DL, Ewart AK, Morris CA, Leppert MF, Keating MT. The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. Cell 1993; 73(1):159–168.
11. Li DY, Toland AE, Boak BB, Atkinson DL, Ensing GJ, Morris CA, Keating MT. Elastin point mutations cause an obstructive vascular disease, supravalvular aortic stenosis. Hum Mol Genet 1997; 6(7):1021–1028.
12. Tassabehji M, Metcalfe K, Donnai D, Hurst J, Reardon W, Burch M, Read AP. Elastin: genomic structure and point mutations in patients with supravalvular aortic stenosis. Hum Mol Genet 1997; 6(7):1029–1036.
13. Milunsky JM, Huang XL. Unmasking Kabuki syndrome: chromosome 8p22-8p23.1 duplication revealed by comparative genomic hybridization and BAC-FISH. Clin Genet 2003; 64(6):509–516.
14. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997; 276(5321): 2045–2047.

15. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K. Alpha-synuclein locus triplication causes Parkinson's disease. Science 2003; 302(5646):841.
16. Antonarakis SE, Kazazian HH, Tuddenham EG. Molecular etiology of factor VIII deficiency in hemophilia A. Hum Mutat 1995; 5(1):1–22.
17. Becker J, Schwaab R, Moller-Taube A, Schwaab U, Schmidt W, Brackmann HH, Grimm T, Olek K, Oldenburg J. Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies. Am J Hum Genet 1996; 58(4):657–670.
18. Shaikh TH, Budarf ML, Celle L, Zackai EH, Emanuel BS. Clustered 11q23 and 22q11 breakpoints and 3:1 meiotic malsegregation in multiple unrelated t(11;22) families. Am J Hum Genet 1999; 65(6):1595–1607.
19. Edelmann L, Spiteri E, McCain N, Goldberg R, Pandita RK, Duong S, Fox J, Blumenthal D, Lalani SR, Shaffer LG, Morrrow BE. A common breakpoint on 11q23 in carriers of the constitutional t(11;22) translocation. Am J Hum Genet 1999; 65(6):1608–1616.
20. Kurahashi H, Shaikh TH, Hu P, Roe BA, Emanuel BS, Budarf ML. Regions of genomic instability on 22q11 and 11q23 as the etiology for the recurrent constitutional t(11;22). Hum Mol Genet 2000; 9(11):1665–1670.
21. Kurahashi H, Shaikh TH, Emanuel BS. Alu-mediated PCR artifacts and the constitutional t(11;22) breakpoint. Hum Mol Genet 2000; 9(18):2727–2732.
22. Kurahashi H, Shaikh T, Takata M, Toda T, Emanuel BS. The constitutional t(17;22): another translocation mediated by palindromic AT-rich repeats. Am J Hum Genet 2003; 72(3):733–738.
23. Giglio S, Calvari V, Gregato G, Gimelli G, Camanini S, Giorda R, Ragusa A, Guerneri S, Selicorni A, Stumm M, Tonnies H, Ventura M, Zollino M, Neri G, Barber J, Wieczorek D, Rocchi M, Zuffardi O. Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. Am J Hum Genet 2002; 71(2):276–285.
24. Collier S, Tassabehji M, Sinnott P, Strachan T. A de novo pathological point mutation at the 21-hydroxylase locus: implications for gene conversion in the human genome. Nat Genet 1993; 3(3):260–265.
25. Boocock GR, Morrison JA, Popovic M, Richards N, Ellis L, Durie PR, Rommens JM. Mutations in SBDS are associated with Shwachman–Diamond syndrome. Nat Genet 2003; 33(1):97–101.
26. Gimelli G, Pujana MA, Patricelli MG, Russo S, Giardino D, Larizza L, Cheung J, Armengol L, Schinzel A, Estivill X, Zuffardi O. Genomic inversions of human chromosome 15q11-q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions. Hum Mol Genet 2003; 12(8):849–858.
27. Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, Scherer SW. A 1.5 million-base pair inversion polymorphism in families with Williams–Beuren syndrome. Nat Genet 2001; 29(3):321–325.
28. Stankiewicz P, Shaw CJ, Dapper JD, Wakui K, Shaffer LG, Withers M, Elizondo L, Park SS, Lupski JR. Genome architecture catalyzes nonrecurrent chromosomal rearrangements. Am J Hum Genet 2003; 72(5):1101–1116.

29. Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tonnesen T, Carlberg BM, Pettersson U. Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. Hum Mol Genet 1995; 4(4):615–621.
30. Timms KM, Lu F, Shen Y, Pierson CA, Muzny DM, Gu Y, Nelson DL, Gibbs RA. 130 kb of DNA sequence reveals two new genes and a regional duplication distal to the human iduronate-2-sulfate sulfatase locus. Genome Res 1995; 5(1):71–78.
31. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet 2002; 3(5):370–379.
32. Hobbs HH, Russell DW, Brown MS, Goldstein JL. The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. Annu Rev Genet 1990; 24:133–170.
33. Nobile C, Toffolatti L, Rizzi F, Simionati B, Nigro V, Cardazzo B, Patarnello T, Valle G, Danieli GA. Analysis of 22 deletion breakpoints in dystrophin intron 49. Hum Genet 2002; 110(5):418–421.
34. Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE.Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 1988; 332(6160)164–166.
35. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. A de novo Alu insertion results in neurofibromatosis type 1. Nature 1991; 353(6347):864–866.
36. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. Segmental duplications: organization and impact within the current human genome project assembly. Genome Res 2001; 11(6):1005–1017.
37. Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. Genome Biol 2003; 4(4):R25.
38. Lander ES, Linton LM, Birren B, et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409(6822):860–921.
39. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome.Science 2001; 291(5507):1304–1351.
40. Li WH, Gu Z, Wang H, Nekrutenko A. Evolutionary analyses of the human genome. Nature 2001; 409(6822):847–849.
41. Rochette CF, Gilbert N, Simard LR. SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to *Homo sapiens*. Hum Genet 2001; 108(3):255–266.
42. Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, et al. Identification and characterization of a spinal muscular atrophy-determining gene. Cell 1995; 80(1):155–165.
43. Regnier V, Meddeb M, Lecointre G, Richard F, Duverger A, Nguyen VC, Dutrillaux B, Bernheim A, Danglot G. Emergence and scattering of multiple neurofibromatosis (NF1)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. Hum Mol Genet 1997; 6(1):9–16.
44. Eichler EE, Budarf ML, Rocchi M, Deaven LL, Doggett NA, Baldini A, Nelson DL, Mohrenweiser HW. Interchromosomal duplications of the adreno-

leukodystrophy locus: a phenomenon of pericentromeric plasticity. Hum Mol Genet 1997; 6(7):991–1002.
45. Groot PC, Mager WH, Frants RR. Interpretation of polymorphic DNA patterns in the human alpha-amylase multigene family. Genomics 1991; 10(3):779–785.
46. McLellan RA, Oscarson M, Alexandrie AK, Seidegard J, Evans DA, Rannug A, Ingelman-Sundberg M. Characterization of a human glutathione S-transferase mu cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity. Mol Pharmacol 1997; 52(6):958–965.
47. Schwartz M, Sorensen N, Hansen FJ, Hertz JM, Norby S, Tranebjaerg L, Skovby F. Quantification, by the solid-phase minisequencing, of the telomeric and centromeric copies of the survival motor neuron gene in families with spinal muscular atrophy. Hum Mol Genet 1997; 6(1):99–104.
48. Feldkotter M, Schwarzer V, Wirth R, Wienker TF, Wirth B. Quantification, by solid-phase minisequencing, of the telomeric and centromeric copies of the survival motor neuron gene in families with spinal muscular atrophy. Hum Mol Genet 2002; 70(2):358–368.
49. Koppens PF, Hoogenboezem T, Degenhart HJ. Duplication of the CYP21A2 gene complicates mutation analysis of steroid 21-hydroxylase deficiency: characteristics of three unusual haplotypes. Hum Genet 2002; 111(4–5):405–410.
50. Chung EK, Yang Y, Rennebohm RM, Lokki ML, Higgins GC, Jones KN, Zhou B, Blanchong CA, Yu CY. Genetic sophistication of human complement components C4A and C4B and RP-C4-CYP21-TNX (RCCX) modules in the major histocompatibility complex. Am J Hum Genet 2002; 71(4):823–837.
51. Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O. Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet 2001; 68(4):874–883.
52. Hollox EJ, Armour JA, Barber JC. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. Am J Hum Genet 2003; 73(3):591–600.
53. Yu CE, Dawson G, Munson J, D'Souza I, Osterling J, Estes A, Leutenegger AL, Flodman P, Smith M, Raskind WH, Spence MA, McMahon W, Wijsman EM, Schellenberg GD. Presence of large deletions in kindreds with autism. Am J Hum Genet 2002; 71(1):100–115.
54. Hikami K, Tsuchiya N, Yabe T, Tokunaga K. Variations of human killer cell lectin-like receptors: common occurrence of NKG2-C deletion in the general population. Genes Immun 2003; 4(2):160–167.
55. Sasso EH, Buckner JH, Suzuki LA. Ethnic differences of polymorphism of an immunoglobulin VH3 gene. J Clin Invest 1995; 96(3):1591–600.
56. Fantes JA, Mewborn SK, Lese CM, Hedrick J, Brown RL, Dyomin V, Chaganti RS, Christian SL, Ledbetter DH. Organisation of the pericentromeric region of chromosome 15: at least four partial gene copies are amplified in patients with a proximal duplication of 15q. J Med Genet 2002; 39(3):170–177.
57. Gratacos M, Nadal M, Martin-Santos R, Pujana MA, Gago J, Peral B, Armengol L, Ponsa I, Miro R, Bulbena A, Estivill X. A polymorphic genomic duplication on human chromosome 15 is a susceptibility factor for panic and phobic disorders. Cell 2001; 106(3):367–379.

58. Townson JR, Barcellos LF, Nibbs RJ. Gene copy number regulates the production of the human chemokine CCL3-L1. Eur J Immunol 2002; 32(10):3016–3026.
59. Franchina M, Kay PH. Allele-specific variation in the gene copy number of human cytosine 5-methyltransferase. Hum Hered 2000; 50(2):112–117.
60. Rao Y, Hoffmann E, Zia M, Bodin L, Zeman M, Sellers EM, Tyndale RF. Duplications and defects in the CYP2A6 gene: identification, genotyping, and in vivo effects on smoking. Mol Pharmacol 2000; 58(4):747–755.
61. Rabbani H, Pan Q, Kondo N, Smith CI, Hammarstrom L. Duplications and deletions of the human IGHC locus: evolutionary implications. Immunogenetics 1996; 45(2):136–141.
62. Bottaro A, Gallina R, Brusco A, Cariota U, Boccazzi C, Barilaro MR, Plebani A, Ugazio AG, van Leeuwen AM, DeLange GG, Galanello R, Carbonara AO. Familial clustering of IGHC delections and duplications: functional and molecular analysis. Immunogenetics 1993; 37(5):356–363.
63. van der Burg M, Barendregt BH, van Gastel-Mol EJ, Tumkaya T, Langerak AW, van Dongen JJ. Unraveling of the polymorphic C lambda 2-C lambda 3 amplification and the Ke+Oz- polymorphism in the human Ig lambda locus. J Immunol 2002; 169(1):271–276.
64. Sprenger R, Schlagenhaufer R, Kerb R, Bruhn C, Brockmoller J, Roots I, Brinkmann U. Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype–phenotype correlation. Pharmacogenetics 2000; 10(6):557–565.
65. McLellan RA, Oscarson M, Seidegard J, Evans DA, Ingelman-Sundberg M. Frequent occurrence of CYP2D6 gene duplication in Saudi Arabians. Pharmacogenetics 1997; 7(3):187–191.
66. Mitsunaga Y, Kubota T, Ishiguro A, Yamada Y, Sasaki H, Chiba K, Iga T. Frequent occurrence of CYP2D6*10 duplication allele in a Japanese population. Mutat Res 2002; 505(1–2):83–85.
67. Dalen P, Dahl ML, Ruiz ML, Nordin J, Bertilsson L. 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes. Clin Pharmacol Ther 1998; 63(4):444–452.
68. Small K, Iber J, Warren ST. Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. Nat Genet 1997; 16(1):96–99.
69. Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, Brenton M, Hink R, Burgers S, Hernandez-Boussard T, Karimpour-Fard A, Glueck D, McGavran L, Berry R, Pollack J, Sikela JM. Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol 2004; 2(7):E207.
70. Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ. Complex SNP-related sequence variation in segmental genome duplications. Nat Genet 2004; 36(8):861–866.
71. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science 2004; 305(5683):525–528.
72. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet 2004; 36(9):949–951.

73. Hayes JD, Strange RC. Glutathione S-transferase polymorphisms and their biological consequences. Pharmacology 2000; 61(3):154–166.
74. Strange RC, Spiteri MA, Ramachandran S, Fryer AA. Glutathione-S-transferase family of enzymes. Mutat Res 2001; 482(1–2):21–26.
75. Landi S. Mammalian class theta GST and differential susceptibility to carcinogens: a review. Mutat Res 2000; 463(3):247–283.
76. Ingelman-Sundberg M. Polymorphism of cytochrome P450 and xenobiotic toxicity. Toxicology 2002; 181–182:447–452.
77. Zanger UM, Fischer J, Raimundo S, Stuven T, Evert BO, Schwab M, Eichelbaum M. Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. Pharmacogenetics 2001; 11(7):573–585.
78. Lundqvist E, Johansson I, Ingelman-Sundberg M. Genetic mechanisms for duplication and multiduplication of the human CYP2D6 gene and methods for detection of duplicated CYP2D6 genes. Gene 1999; 226(2):327–338.
79. Yoshida R, Nakajima M, Watanabe Y, Kwon JT, Yokoi T. Genetic polymorphisms in human CYP2A6 gene causing impaired nicotine metabolism. Br J Clin Pharmacol 2002; 54(5):511–517.
80. Oscarson M, McLellan RA, Gullsten H, Yue QY, Lang MA, Bernal ML, Sinues B, Hirvonen A, Raunio H, Pelkonen O, Ingelman-Sundberg M. Characterisation and PCR-based detection of a CYP2A6 gene deletion found at a high frequency in a Chinese population. FEBS Lett 1999; 448(1):105–110.
81. Nunoya K, Yokoi T, Takahashi Y, Kimura K, Kinoshita M, Kamataki T. Homologous unequal crossover within the human CYP2A gene cluster as a mechanism for the deletion of the entire CYP2A6 gene associated with the poor metabolizer phenotype. J Biochem (Tokyo) 1999; 126(2):402–407.
82. Xu C, Rao YS, Xu B, Hoffmann E, Jones J, Sellers EM, Tyndale RF. An in vivo pilot study characterizing the new CYP2A6*7, *8, and *10 alleles. Biochem Biophys Res Commun 2002; 290(1):318–324.
83. Ariyoshi N, Sekine H, Saito K, Kamataki T. Characterization of a genotype previously designated as CYP2A6 D-type: CYP2A6∗4B, another entire gene deletion allele of the CYP2A6 gene in Japanese. Pharmacogenetics 2002; 12(6):501–504.
84. Mars WM, Patmasiriwat P, Maity T, Huff V, Weill MM, Saunders GF. Inheritance of unequal numbers of the genes encoding the human neutrophil defensins HP-1 and HP-3. J Biol Chem 1995; 270(51):30371–30376.
85. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 1992; 258(5083):818–821.
86. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 1997; 20(4):399–407.
87. Mantripragada KK, Buckley PG, de Stahl TD, Dumanski JP. Genomic microarrays in the spotlight. Trends Genet 2004; 20(2):87–94.
88. Buckley PG, Mantripragada KK, Benetkiewicz M, Tapia-Paez I, Diaz De Stahl T, Rosenquist M, Ali H, Jarbo C, De Bustos C, Hirvela C, Sinder Wilen B, Fransson I, Thyr C, Johnsson BI, Bruder CE, Menzel U, Hergersberg M,

Mandahl N, Blennow E, Wedell A, Beare DM, Collins JE, Dunham I, Albertson D, Pinkel D, Bastian BC, Faruqi AF, Lasken RS, Ichimura K, Collins VP, Dumanski JP. A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. Hum Mol Genet 2002; 11(25):3221–3229.

89. Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 2003; 13(10):2291–2305.
90. Bignell GR, Huang J, Greshock J, Watt S, Butler A, West S, Grigorova M, Jones KW, Wei W, Stratton MR, Futreal PA, Weber B, Shapero MH, Wooster R. High-resolution analysis of DNA copy number using oligonucleotide microarrays. Genome Res 2004; 14(2):287–295.
91. Bieche I, Olivi M, Champeme MH, Vidaud D, Lidereau R, Vidaud M. Novel approach to quantitative polymerase chain reaction using real-time detection: application to the detection of gene amplification in breast cancer. Int J Cancer 1998; 78(5):661–666.
92. Ponchel F, Toomes C, Bransfield K, Leong FT, Douglas SH, Field SL, Bell SM, Combaret V, Puisieux A, Mighell AJ, Robinson PA, Inglehearn CF, Isaacs JD, Markham AF. Real-time PCR based on SYBR-Green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. BMC Biotechnol 2003; 3(1):18.
93. Charbonnier F, Raux G, Wang Q, Drouot N, Cordier F, Limacher JM, Saurin JC, Puisieux A, Olschwang S, Frebourg T. Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. Cancer Res 2000; 60(11):2760–2763.
94. Duponchel C, Di Rocco C, Cicardi M, Tosi M. Rapid detection by fluorescent multiplex PCR of exon deletions and duplications in the C1 inhibitor gene of hereditary angioedema patients. Hum Mutat 2001; 17(1):61–70.
95. Casilli F, Di Rocco ZC, Gad S, Tournier I, Stoppa-Lyonnet D, Frebourg T, Tosi M. Rapid detection of novel BRCA1 rearrangements in high-risk breast-ovarian cancer families using multiplex PCR of short fluorescent fragments. Hum Mutat 2002; 20(3):218–226.
96. Armour JA, Sismani C, Patsalis PC, Cross G. Measurement of locus copy number by hybridisation with amplifiable probes. Nucleic Acids Res 2000; 28(2):605–609.
97. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002; 30(12):e57.

26

General Conclusions and Future Directions

Werner Kalow
Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada
Arno G. Motulsky
University of Washington, Seattle, Washington, U.S.A.

1. OVERVIEW

We hope that the different chapters of this book will convey to anyone interested in pharmacogenomics some idea of the various themes and areas of specialized knowledge that will have to come together to convert expectations to reality. The expectations, which have been aroused by pharmacogenomics are substantial and involve the gradual development of personalized medicine or therapeutics, leaving behind the present statistically based medicine. There are variable estimates of the time, which these processes may require; perhaps it will take 15 to 20 years. Eventually, however, pharmacogenomics is hoped to be a major payoff for society from the scientific and technical revolution called ''genomics.''

There is no doubt that some of the specialized techniques described in individual chapters are useful at the present time but may be outdated relatively fast as time goes on. This is always a danger for technical descriptions. In addition, the book covers many principles and problems, which will remain with us for a long time to come. The main hope of the editors is that all who see and decide to read the book (or to read in the book) will feel to have gained new insights or new knowledge.

Pharmacogenetics and pharmacogenomics will have their impact on medicine, the biomedical sciences, and drug development in several ways. Variable responses to drugs and adverse events have always been observed. Modern genetic approaches have the promise of uncovering the genetic and molecular basis of such variability thereby identifying individuals who are non-responders to a drug or at risk for adverse drug reactions. Appropriate testing before a drug is used will often be possible and the potentially ineffective or offending drug can be avoided. For some drugs, pretesting may help in defining an appropriate dosage regimen that is effective and avoids toxicity. Drug development will be aided by finding effective medications whose design is based on disease-specific genomic and phenomic alterations. These scientific developments may lead to the elaboration of ''pharmacogenetic efficiency profiles'' for a given drug and more individualized therapy for the patient.

Clinical trials of new drugs will more frequently include only those individuals who on pharmacogenetic and pharmacogenomic grounds have a greater chance of favorably responding to the new drug. Adverse drug reactions are usually difficult to predict but better understanding of drug metabolism, drug response, and more data on genetic polymorphisms will provide occasional opportunities to screen out those at risk for currently unsuspected adverse events.

The best-understood pharmacogenetic examples relate to processes or reactions mediated by a single gene where a mutation grossly interferes with a crucial step in drug metabolism or drug response (1). However, while more such examples will be discovered, the more frequent role of genetics and genomics will involve multifactorial phenomena where multiple different genes affecting drug metabolism and drug action interact with the environment and with each other. The resultant bell-shaped curve (i.e., number of individuals plotted against a biologic endpoint such as drug half-life) is unlike the bimodal curve classically seen with monogenic drug reactions where ''normals'' and ''abnormals'' represent different populations. As knowledge augments, it will be increasingly possible to define the specific role of each of the various components of polygenic variation by genetic, molecular, and biochemical approaches with careful attention to fixed biologic characteristics (such as age and gender) as well as to various environmental factors.

Pharmacogenetics is not a new science. The role of genes in drug response had been known for about 50 years (2,3) but recognition of a broader applicability of these concepts for drug therapy and drug discovery and development only occurred more recently, particularly as molecular techniques for in vitro studies became available and the need to administer a drug to healthy individuals and their family members no longer existed in pharmacogenetic studies. An additional reason for delay in the development of pharmacogenetics was the existence of only a few examples of

pharmacogenetic variation that were not considered relevant for drug action in general. The demonstration that metabolism for many drugs as assessed by drug half-life differed among unrelated individuals but was very similar in identical twins as compared to non-identical twins (4) pointed to an important role of genetic factors to explain variation of drug metabolism. Identical twins share all their genes while non-identical twins have only half of their genes in common.

Major emphasis in pharmacogenetics has remained on monogenic variation, which can be studied readily with a variety of modern techniques. In this regard, pharmacogenetics mirrors progress in medical genetics in general. While superb advances continue in unraveling the role of single-gene mutations in genetic diseases, success in isolating and elucidating the role of specific genes in the common multifactorial conditions has been much slower, even though considerable effort has been devoted to various approaches to find these genes (5). Chromosomal localizations of possibly involved genes have often been suggested. However, repeat studies have tended to be unsuccessful in confirming initial results. Few genes except those represented by the less common monogenic subtypes of a given common disease (such as the autosomal dominant breast and colon cancer genes) have been identified. Identification of the genes that contribute to "polygenic pharmacogenetics" may be simpler to achieve since the number of genes involved in drug disposal and drug response are likely to be fewer than in multifactorial diseases. The identification of genes in pharmacogenetics may therefore be less difficult.

A consortium of the pharmaceutical industry together with the Wellcome Trust and several academic institutions are now carrying out a search for common DNA variants known as single-nucleotide polymorphisms (SNPs) across the whole human genome in the hope that such DNA variants can be used as markers to signal the presence of very closely linked genes involved in pharmacogenetic processes (6). The likelihood of ultimate success in this endeavor is difficult to predict (5,7). If cSNPs and other DNA changes are detected and related to drug responses, one still has to (a) apply the classical techniques of cloning and identifying the entire gene, (b) express its protein products, and (c) study its function under physiological conditions. The population history, the evolutionary age, and possible selective factors of different SNPs as well as recombination rates between a SNP and the linked gene of interest will vary among populations. The total number of SNPs needed and the number of individuals requiring testing for identifying a relevant gene therefore will often vary between different genes and different populations. Use of genetically more homogeneous populations such as Icelanders in such studies may not necessarily be more helpful (but see Ref. 8). However, data of successful narrowing the genetic region carrying the apo E4 allele (of interest in Alzheimer's disease) have been published by using the SNP approach (9). Linked sets of SNPs that

are inherited together as a unique haplotype simplify the detection of linked genes affecting drug response. Efforts to construct the so-called hap-maps for different populations by this approach are under way, but the ultimate success of this methodology remains to be demonstrated.

2. SOCIETAL PROBLEMS

Fundamental research that led to potential practical applications of pharmacogenetics has usually been carried out in academic institutions. Together with a shift of more biomedical research to pharmaceutical and biotechnology companies, pharmacogenetic and pharmacogenomic approaches now occupy an important place in the research portfolio of such companies. With this shift and an interest of academic institutions to derive income from research discoveries, more patents are being applied for. Efforts have already been made to patent newly discovered human genes without any knowledge of their function. It appears now that the U.S. Patent Office is unlikely to issue such patents unless they describe a novel function or application that has utility such as the commercialization of a discovery. Since introduction of a new drug has become very expensive, granting of patents will make novel drugs more expensive still, increasing the cost of medical care, a major problem in much of the world.

The total impact of pharmacogenetics and pharmacogenomics on the pharmaceutical industry in the long run is hazardous to predict. However, even though some analysts on the business side of the pharmaceutical industry are concerned about reducing market size by targeting smaller segments of the population, most companies have placed large investments in genetic and genomic applications in the belief that these approaches will lead to ultimate success and profitability. The impact of pharmacogenetics and pharmacogenomics on the practice of medicine as compared with the development of new drugs should be considered separately. New drug development is of key interest to industry but of less importance to practicing physicians until a new drug becomes available for use. Genomic concepts and techniques together with the application of proteomics are likely to point out new targets for drug therapy based on better understanding of the biologic mechanism of disease. A more rational therapy is therefore likely to evolve. Effectiveness of a given drug in only some people is currently common. There may be environmental causes like food or smoking. In many cases, the reasons are pharmacogenetic differences between people, which may affect, e.g., the metabolism of a drug. Another reason may be the fact that a common disease to be treated is usually caused, or contributed to, by many genes; however, the set of genes contributing to an apparently identical disease often differs between subjects, thereby causing differences of therapeutic response. This is one factor calling for personalized medicine.

Pharmaceutical companies often want orphan drug status given to a new drug. Under current FDA rules, drugs for rare conditions affecting fewer than 200,000 people in the United States receive tax breaks on clinical trials and 7 years of marketing exclusivity. Drugs that currently have a larger market but that following pharmacogenetic testing would be predictably effective in $<$200,000 people might qualify under current rules. Will more drugs therefore be given orphan status? Developments in this area will be followed with much interest.

3. ETHICAL PROBLEMS

Not covered in this book are ethical issues, which may be harder to solve than some technical problems. Considering ethical issues, it is useful to start by contemplating the Hippocratic Code: the patient, after having agreed to be treated by a physician, trusts the physician to act in the patient's best interest; no other ethical problem arises. This is an entirely different situation from one in which a physician-scientist's interest is the promotion of knowledge.

Protection of a patient's interest who participates in a research project has led to many statements and laws. A major start was the Nuremberg Code (10). In the United States, the Office of Protection from Research Risks (OPRR) has issued regulations known as the "Common Rule." Almost everywhere is institutional review boards, safeguarding the rights of the individuals.

The point to be considered in the present context is that with the aim to create personalized medicine, most investigations will be designed to benefit the patient as well as to promote general knowledge. The logical consequence for the formulation of protective laws should be to combine the new rules with the spirit of the Hippocratic Code. It remains to be seen to which extent this will be possible.

If the medical choice of drug or drug dosage depends on a patient's genes, the prescriber has to know the genes of the patient, which may affect these choices; this requires a look into the patient's genetic privacy. However, it also requires the prescriber to know from a study of pharmacogenetics or pharmacogenomics which genotypes are compatible with which drug, necessitating a restructuring of medical education within the framework of the many other problems posed by developments in genetic and genomic medicine.

Many ethical and societal issues of pharmacogenetics and pharmacogenomics are not unique and raise issues similar to those brought up by the recent advances. While different human populations regardless of geographic origin of their ancestors share the vast majority of their genes, considerable genetic variation remains which renders everyone genetically unique (11). There is considerably more genetic variation within a given population (or

"race") than among populations (12). Despite the universality of this finding in all studies, using non-coding DNA variants (such as microsatellites) and a novel statistical cluster technique, it has become possible to assign anonymous specimens from different populations to their correct geographic origin, such as Africa, East Asia, Europe, West Asia, etc. These designations correlate well with conventional racial classifications. The new method also allows assignment of the extent of admixture in hybrid populations (12). The frequencies of pharmacogenetic traits usually differ in populations from various parts of the world (Chapter 6) due to different selective factors but often for unknown reasons. Since such differences may cause adverse drug reactions or variable drug responses, knowledge of a patient's ethnic origin will often be useful to select appropriate pharmacogenetic tests for clinical trials or in medical practice. If a given allele (or set of alleles) that leads to different drug metabolism or variable drug response differs between ethnic groups, selection of the appropriate PCR reactions (or similar test) that is common in a given population can be made. However, it may often be difficult logistically to assign ethnic origin and ethnic mixture is increasingly common as well. All alleles of a given pharmacogenetic system rather than a population specific set may therefore need to be included in the test system. Utilization of biochemical tests such as measurement of enzyme activity may sometimes be more appropriate since it could detect low enzyme activity regardless of which one of multiple DNA alleles that reduce enzyme activity is involved. Molecular diagnosis, however, is currently simpler and more accurate, and can be more readily automated. The technical and population problems encountered will vary for different biochemical and molecular measurements of a given pharmacogenetic trait.

The DNA testing has raised public worries regarding privacy, and legislative initiatives have been proposed to restrict DNA testing in medical settings. Such trends often relate to failure of differentiating forensic DNA tests from DNA tests designed to diagnose various diseases or disease susceptibilities. Genetic testing in medicine raises problems of informed consent, privacy, confidentiality, stigmatization, and discrimination (health insurance and occupation), and requires appropriate guidelines and oversight (13). Roses (14) has pointed out, however, that pharmacogenetic testing which aims to detect genes and SNP profiles involved in drug metabolism and drug action is different by only searching for a "pharmacogenetic efficiency" or "medicine response" profile. He states that such a goal carries no special ethical or social problems and therefore should be considered differently than other kinds of genetic testing. Attempting to find the most appropriate drug for a patient is therefore considered similar to other diagnostic tests in medicine. However, an abnormal result is also relevant for drug therapy of relatives. As long as DNA specimens are only used for pharmacogenetic tests related to adverse events and effective drug

responses, a good case can be made for treating such tests somewhat differently from more sensitive genetic tests with more serious implications.

Clinical trials on patients with a disease pose additional problems. Here, patients with variable disease mechanisms and often with a different natural history of their disease may require different treatments. Detailed descriptions of dietary habits, smoking, and other environmental exposures (depending on the condition under study) in addition to age, sex, and ethnic origin will be necessary in clinical trials to elucidate the interaction of environmental and genetic factors. Anonymity that strips a specimen of a name but retains demographic, genetic, and environmental information is one solution and allows additional investigation when new laboratory methodology and additional genetic and environmental markers for studies of the same disease become available. Retention of ethnic identity may be important for the reasons already mentioned. To prevent future abuses, it has been suggested to destroy DNA specimens after testing has been completed. Much valuable information in research studies would be lost under these circumstances, particularly if novel tests become available for the study of the same disease, thus causing problems in locating study subjects. Anonymity with retention of specimens therefore appears most appropriate but has the disadvantage that even a trial participant will not have access to his or her own specimen in the future.

Other problems arise. The intellectual property rights related to specimens from clinical trials are not always clear. Do the study participants whose SNPs lead to new drugs and more appropriate therapy own their DNA, or do all potential benefits go to the investigator or company who collect these specimens? Under what conditions could organizations that collect DNA for either clinical trials or other purposes sell this information to third parties?

REFERENCES

1. Weber W. Pharmacogenetics. New York: Oxford University Press, 1997.
2. Motulsky A. Drug reactions, enzymes and biochemical genetics. JAMA 1957; 165:835–837.
3. Kalow W. Pharamacogenetics—Heredity and the Responses to Drugs. Philadelphia: W.B. Saunders, 1962.
4. Vesell ES. Pharmacogenetic perspectives gained from twin and family studies. Pharmacol Ther 1989; 41:535–552.
5. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000; 405:847–856.
6. McCarthy JJ, Hilfiker R. The use of single-nucleotide polymorphism maps in pharmacogenomics. Nat Biotechnol 2000; 18:505–508.
7. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999; 22:139–144.

8. Collins A, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 1999; 96:15173–15177.
9. Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM. SNPing away at complex diseases: analysis if single-nucleotide polymorphism around APOE in Alzheimer disease. Am J Hum Genet 2000; 67:383–394.
10. Sobel ME. Ethical issues in molecular pathology: paradigms in flux. Arch Pathol Lab Med 1999; 123:1076–1078.
11. Cavalli-Sforza LL. Genes, Peoples, and Languages. New York: North Point Press, 2000.
12. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotov LA, Feldman MW. Genetic structure of human populations. Science 2002; 298:2381–2385.
13. Holtzman NA. Promoting safe and effective genetic tests in the United States: work of the task force on genetic testing. Clin Chem 1999; 45:732–738.
14. Roses AD. Pharmacogenetics and the practice of medicine. Nature 2000; 405:857–865.