Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment

Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment, National Research Council

ISBN: 0-309-11299-0, 300 pages, 6 x 9, (2007)

This free PDF was downloaded from:
http://www.nap.edu/catalog/12037.html

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online, free
- Sign up to be notified when new books are published
- Purchase printed books
- Purchase PDFs
- Explore with our innovative research tools

Thank you for downloading this free PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to comments@nap.edu.

This free book plus thousands more books are available at http://www.nap.edu.

Copyright © National Academy of Sciences. Permission is granted for this material to be shared for noncommercial, educational purposes, provided that this notice appears on the reproduced materials, the Web address of the online, full authoritative version is retained, and copies are not altered. To disseminate otherwise or to republish requires written permission from the National Academies Press.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment

Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology

Board on Environmental Studies and Toxicology

Board on Life Sciences

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

Copyright © National Academy of Sciences. All rights reserved.

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This project was supported by Contract No. N01-OD-4-2139 between the National Academy of Sciences and the U.S. Department of Health and Human Services. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number-13: 978-0-309-11298-7
International Standard Book Number-10: 0-309-1129802

Cover design by Liza R. Hamilton.

Additional copies of this report are available from

The National Academies Press
500 Fifth Street, NW
Box 285
Washington, DC 20055

800-624-6242
202-334-3313 (in the Washington metropolitan area)
http://www.nap.edu

Copyright 2007 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America.

Copyright © National Academy of Sciences. All rights reserved.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

COMMITTEE ON APPLICATIONS OF TOXICOGENOMIC TECHNOLOGIES TO PREDICTIVE TOXICOLOGY

Members

DAVID C. CHRISTIANI (*Chair*), Harvard School of Public Health, Boston, MA
CYNTHIA A. AFSHARI, Amgen, Inc.,Thousand Oaks, CA
JOHN M. BALBUS, Environmental Defense, Washington, DC
JAMES S. BUS, The Dow Chemical Company, Midland, MI
BRUCE F. DEMPLE, Harvard School of Public Health, Boston, MA
LINDA E. GREER, Natural Resources Defense Council, Washington, DC
SHARON L.R. KARDIA, University of Michigan, Ann Arbor
GEORGE D. LEIKAUF, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA
DANIEL C. LIEBLER, Vanderbilt University School of Medicine, Nashville, TN
GARY E. MARCHANT, Arizona State University College of Law, Tempe
JOHN QUACKENBUSH, Harvard School of Public Health, Boston, MA
KENNETH S. RAMOS, University of Louisville, Louisville, KY
MARK A. ROTHSTEIN, University of Louisville School of Medicine, Louisville, KY
RAYMOND E. STOLL, Stoll & Associates, LLC, Storrs-Mansfield, CT
ROGER G. ULRICH, Calistoga Pharmaceuticals Inc., Seattle, WA
HELMUT ZARBL, University of Medicine and Dentistry of New Jersey, Piscataway

Staff

MARILEE SHELTON-DAVENPORT, Project Director
SUZANNE VAN DRUNICK, Senior Program Officer
KARL GUSTAVSON, Senior Program Officer
EILEEN ABT, Senior Program Officer
CAY BUTLER, Editor
MIRSADA KARALIC-LONCAREVIC, Manager, Toxicology Information Center
LIZA R. HAMILTON, Senior Program Assistant
MORGAN R. MOTTO, Senior Program Assistant

Sponsor

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Copyright © National Academy of Sciences. All rights reserved.

BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Members

JONATHAN M. SAMET (*Chair*), Johns Hopkins University, Baltimore, MD
RAMÓN ALVAREZ, Environmental Defense, Austin, TX
JOHN M. BALBUS, Environmental Defense, Washington, DC
DALLAS BURTRAW, Resources for the Future, Washington, DC
JAMES S. BUS, Dow Chemical Company, Midland, MI
RUTH DEFRIES, University of Maryland, College Park
COSTEL D. DENSON, University of Delaware, Newark
E. DONALD ELLIOTT, Willkie Farr & Gallagher LLP, Washington, DC
MARY R. ENGLISH, University of Tennessee, Knoxville
J. PAUL GILMAN, Oak Ridge Center for Advanced Studies, Oak Ridge, TN
SHERRI W. GOODMAN, Center for Naval Analyses, Alexandria, VA
JUDITH A. GRAHAM, American Chemistry Council, Arlington, VA
WILLIAM P. HORN, Birch, Horton, Bittner and Cherot, Washington, DC
WILLIAM M. LEWIS, JR., University of Colorado, Boulder
JUDITH L. MEYER, University of Georgia, Athens
DENNIS D. MURPHY, University of Nevada, Reno
PATRICK Y. O'BRIEN, ChevronTexaco Energy Technology Company, Richmond, CA
DOROTHY E. PATTON (retired), Chicago, IL
DANNY D. REIBLE, University of Texas, Austin
JOSEPH V. RODRICKS, ENVIRON International Corporation, Arlington, VA
ARMISTEAD G. RUSSELL, Georgia Institute of Technology, Atlanta
ROBERT F. SAWYER, University of California, Berkeley
KIMBERLY M. THOMPSON, Massachusetts Institute of Technology, Cambridge
MONICA G. TURNER, University of Wisconsin, Madison
MARK J. UTELL, University of Rochester Medical Center, Rochester, NY
CHRIS G. WHIPPLE, ENVIRON International Corporation, Emeryville, CA
LAUREN ZEISE, California Environmental Protection Agency, Oakland

Senior Staff

JAMES J. REISA, Director
DAVID J. POLICANSKY, Scholar
RAYMOND A. WASSEL, Senior Program Officer for Environmental Sciences and Engineering
KULBIR BAKSHI, Senior Program Officer for Toxicology
EILEEN N. ABT, Senior Program Officer for Risk Analysis
KARL E. GUSTAVSON, Senior Program Officer
ELLEN K. MANTUS, Senior Program Officer
SUSAN N.J. MARTEL, Senior Program Officer
STEVEN K. GIBB, Program Officer for Strategic Communications
RUTH E. CROSSGROVE, Senior Editor

Copyright © National Academy of Sciences. All rights reserved.

BOARD ON LIFE SCIENCES

Members

KEITH YAMAMOTO (*Chair*), University of California, San Francisco
ANN M. ARVIN, Stanford University School of Medicine, Stanford, CA
RUTH BERKELMAN, Emory University, Atlanta, GA
DEBORAH BLUM, University of Wisconsin, Madison
VICKI I. CHANDLER, University of Arizona, Tucson
JEFFERY L. DANGL, University of North Carolina, Chapel Hill
PAUL R. EHRLICH, Stanford University, Stanford, CA
MARK D. FITZSIMMONS, John D. and Catherine T. MacArthur Foundation, Chicago, IL
JO HANDELSMAN, University of Wisconsin, Madison
KENNETH H. KELLER, Johns Hopkins University School of Advanced International Studies, Bologna, Italy
JONATHAN D. MORENO, University of Pennsylvania, Philadelphia
RANDALL MURCH, Virginia Polytechinc Institute and State University, Alexandria
MURIEL E. POSTON, Skidmore College, Saratoga Springs, NY
JAMES REICHMAN, University of California, Santa Barbara
BRUCE W. STILLMAN, Cold Spring Harbor Laboratory, Cold Springs Harbor, NY
MARC T. TESSIER-LAVIGNE, Genentech, Inc., South San Francisco, CA
JAMES TIEDJE, Michigan State University, East Lansing
CYNTHIA WOLBERGER, Johns Hopkins University School of Medicine, Baltimore, MD
TERRY L. YATES, University of New Mexico, Albuquerque

Senior Staff

FRANCES E. SHARPLES, Director
KERRY A. BRENNER, Senior Program Officer
ANN H. REID, Senior Program Officer
MARILEE K. SHELTON-DAVENPORT, Senior Program Officer
EVONNE P.Y. TANG, Senior Program Officer
ROBERT T. YUAN, Senior Program Officer
ADAM P. FAGEN, Program Officer
ANNA FARRAR, Financial Associate
MERCURY FOX, Program Assistant
REBECCA L. WALTER, Program Assistant

Copyright © National Academy of Sciences. All rights reserved.

OTHER REPORTS OF THE BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Models in Environmental Regulatory Decision Making (2007)
Toxicity Testing in the Twenty-first Century: A Vision and a Strategy (2007)
Sediment Dredging at Superfund Megasites: Assessing the Effectiveness (2007)
Environmental Impacts of Wind-Energy Projects (2007)
Scientific Review of the Proposed Risk Assessment Bulletin from the Office of Management and Budget (2007)
Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues (2006)
New Source Review for Stationary Sources of Air Pollution (2006)
Human Biomonitoring for Environmental Chemicals (2006)
Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment (2006)
Fluoride in Drinking Water: A Scientific Review of EPA's Standards (2006)
State and Federal Standards for Mobile-Source Emissions (2006)
Superfund and Mining Megasites—Lessons from the Coeur d'Alene River Basin (2005)
Health Implications of Perchlorate Ingestion (2005)
Air Quality Management in the United States (2004)
Endangered and Threatened Species of the Platte River (2004)
Atlantic Salmon in Maine (2004)
Endangered and Threatened Fishes in the Klamath River Basin (2004)
Cumulative Environmental Effects of Alaska North Slope Oil and Gas Development (2003)
Estimating the Public Health Benefits of Proposed Air Pollution Regulations (2002)
Biosolids Applied to Land: Advancing Standards and Practices (2002)
The Airliner Cabin Environment and Health of Passengers and Crew (2002)
Arsenic in Drinking Water: 2001 Update (2001)
Evaluating Vehicle Emissions Inspection and Maintenance Programs (2001)
Compensating for Wetland Losses Under the Clean Water Act (2001)
A Risk-Management Strategy for PCB-Contaminated Sediments (2001)
Acute Exposure Guideline Levels for Selected Airborne Chemicals (five volumes, 2000-2007)
Toxicological Effects of Methylmercury (2000)
Strengthening Science at the U.S. Environmental Protection Agency (2000)
Scientific Frontiers in Developmental Toxicology and Risk Assessment (2000)
Ecological Indicators for the Nation (2000)
Waste Incineration and Public Health (2000)
Hormonally Active Agents in the Environment (1999)
Research Priorities for Airborne Particulate Matter (four volumes, 1998-2004)
The National Research Council's Committee on Toxicology: The First 50 Years (1997)
Carcinogens and Anticarcinogens in the Human Diet (1996)
Upstream: Salmon and Society in the Pacific Northwest (1996)
Science and the Endangered Species Act (1995)
Wetlands: Characteristics and Boundaries (1995)
Biologic Markers (five volumes, 1989-1995)
Review of EPA's Environmental Monitoring and Assessment Program (three volumes, 1994-1995)
Science and Judgment in Risk Assessment (1994)
Pesticides in the Diets of Infants and Children (1993)
Dolphins and the Tuna Industry (1992)
Science and the National Parks (1992)
Human Exposure Assessment for Airborne Pollutants (1991)
Rethinking the Ozone Problem in Urban and Regional Air Pollution (1991)
Decline of the Sea Turtles (1990)

Copies of these reports may be ordered from the National Academies Press
(800) 624-6242 or (202) 334-3313
www.nap.edu

Copyright © National Academy of Sciences. All rights reserved.

OTHER REPORTS OF THE BOARD ON LIFE SCIENCES

Recognizing the Role of Theory in Biology (2007)
2007 Amendments to the National Academies' Guidelines for Human Embryonic Stem Cell Research (2007)
Assessing the Medical Risks of Human Oocyte Donation for Stem Cell Research: A Workshop Report (2007)
The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet (2007)
Validation of Toxicogenomic Technologies: A Workshop Summary (2007)
Protecting Building Occupants and Operations from Biological and Chemical Airborne Threats (2007)
Understanding Interventions That Encourage Minorities to Pursue Research Careers: Summary of a Workshop (2007)
Exploring the Role of Antiviral Drugs in the Eradication of Polio: Workshop Report (2006)
Path to Effective Recovering of DNA from Formalin-Fixed Biological Samples in Natural History Collections: Workshop Summary (2006)
Application of Toxicogenomics to Cross-Species Extrapolation (2006)
Review of the Department of Energy's Genomics: GTL Program (2006)
Overcoming Challenges to Develop Countermeasures Against Aerosolized Bioterrorism Agents: Appropriate Use of Animal Models (2006)
Status of Pollinators in North America (2006)
Treating Infectious Diseases in a Microbial World: Report of Two Workshops on Novel Antimicrobial Therapeutics (2005)
Guidelines for Human Embryonic Stem Cell Research (2005)
Toxicogenomic Technologies and Risk Assessment of Environmental Carcinogens: A Workshop Summary (2005)
Reopening Public Facilities After a Biological Attack: A Decision-Making Framework (2005)
Bridges to Independence: Fostering the Independence of New Investigators in Biomedical Research (2005)
Smallpox: Transforming Biological Information into New Therapies (2004)
Dietary Supplements: A Framework for Evaluating Safety (2004)
Communicating Toxicogenomics Information to Nonexperts: A Workshop Summary (2004)
Safety of Genetically Engineered Foods: Approaches to Assessing Unintended Health Effects (2004)
Biological Confinement of Genetically Engineered Organisms (2004)
Seeking Security: Pathogens, Open Access, and Genome Databases (2004)
Indicators for Waterborne Pathogens (2004)
Enhancing the Vitality of the National Institutes of Health: Organizational Change to Meet New Challenges (2003)
Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences (2003)
Bio2010: Transforming Undergraduate Education for Future Research Biologists (2003)
The National Ecological Observatory Network (NEON): Addressing the Nation's Environmental Challenges (2003)
Scientific and Medical Aspects of Human Reproductive Cloning (2002)
The National Plant Genome Initiative: Objectives for 2003-2008 (2002)
Exploring Horizons for Domestic Animal Genomics (2002)
Animal Biotechnology: Science-Based Concerns (2002)
Environmental Effects of Transgenic Plants (2002)
Countering Agricultural Bioterrorism (2002)
Predicting Invasions of Nonindigenous Plants and Plant Pests (2002)
Funding Smithsonian Scientific Research (2002)
Stem Cells and the Future of Regenerative Medicine (2001)
Environmental Contamination, Biotechnology, and the Law: The Impact of Emerging Genomic Information (2001)

Copyright © National Academy of Sciences. All rights reserved.

Ecological Monitoring of Genetically Modified Crops (2001)
Bioinformatics: Converting Data to Knowledge (2000)
Environmental Issues in Pacific Northwest Forest Management (2000)

Copies of these reports may be ordered from the National Academies Press
(800) 624-6242 or (202) 334-3313
www.nap.edu

Copyright © National Academy of Sciences. All rights reserved.

Preface

In recent years, completion of the sequencing of the human genome as well as the genomes of dozens of other organisms and subsequent development of tools for comprehensive analysis of other cellular constituents have revolutionized biology. These new technologies, referred to broadly as "genomics," have integrated biologic sciences with information sciences and engineering.

The application of these new technologies to toxicology has opened a new era in which genetic variation and expression signatures might be used to screen compounds for hazard identification, to assess cellular responses to different doses, to classify toxicants on the basis of mechanisms of action, to monitor exposure of individuals to toxicants, and to predict individual variability in sensitivity to toxicants. In pharmacology, these technologies have been used both to detect desired cellular responses to drugs and to monitor potential toxicity.

Although the combination of genomic technologies with toxicology has the potential to alter risk assessment by improving the predictive capabilities of toxicology for human health, it is equally clear that significant challenges remain to be overcome, and realistic limitations must be acknowledged.

An issue well highlighted by the development of toxicogenomic technologies is the wide impact of these technologies. The applications of these technologies affect two broad areas: assessment of risk from exposure to chemicals in the environment and assessment of pharmaceutical safety. The National Institutes of Health, regulatory agencies (Environmental Protection Agency, Occupational Safety and Health Administration, Food and Drug Administration), the chemical and pharmaceutical industries, health professionals, attorneys, and the entire public are affected by the applications that toxicogenomics brings to environmental health risk assessment as well as drug efficacy and toxicity.

Recognizing the challenges of dealing with these new types of scientific information, the director of the National Institute of Environmental Health Sciences (NIEHS) of the U.S. Department of Health and Human Services, Dr. Kenneth Olden, asked the National Academies in 2002 to convene a committee of experts to facilitate discussion and communication among stakeholders on technical, regulatory, and ethical issues that need to be considered when deciding how data from this rapidly evolving technology can be used most appropriately to inform public policy and promote human health. The National Academies' standing Committee on Emerging Issues and Data on Environmental Contaminants was established to facilitate exchanges of information among the academic community, government, industry, environmental advocates, and public interest groups about these new approaches.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

The Emerging Issues standing committee, chaired by David Eaton of the University of Washington and Kenneth Ramos of the University of Louisville, has held workshops to encourage discussion on a variety of topics, including toxicogenomics and bioinformatics, toxicogenomics and carcinogenic risk assessment, communicating toxicogenomic information, toxicogenomic application to cross-species extrapolation, sharing toxicogenomic data, validating toxicogenomic data, genomic signatures, toxicogenomics and early life exposures, and intellectual property concerns. These workshop discussions are briefly described in the committee's newsletters (currently available at http://dels.nas.edu/emergingissues/newslet.shtml) and summarized in documents titled *Communicating Toxicogenomics Information to Nonexperts: A Workshop Summary* (http://newton.nap.edu/catalog/11179.html), *Toxicogenomic Technologies and Risk Assessment of Environmental Carcinogens: A Workshop Summary* (http://newton.nap.edu/catalog/11335.html), *Application of Toxicogenomics to Cross-Species Extrapolation: A Report of a Workshop* (http://fermat.nap.edu/catalog/11488.html), and *Validation of Toxicogenomic Technologies: A Workshop Summary* (http://books.nap.edu/catalog.php?record_id=11804).

The Emerging Issues standing committee was also asked to identify areas where an in-depth study from the National Research Council would be beneficial. As a result, the Emerging Issues standing committee described the need for a consensus report on the impacts of toxicogenomic technologies on predictive toxicology. Recognizing this need, in April 2004, NIEHS asked the National Academies to direct its investigative arm, the National Research Council (NRC), to examine the impacts of toxicogenomic technologies on predictive toxicology. In response, the NRC formed the Committee on Applications of Toxicogenomic Technologies for Predictive Toxicology, which authored this report. The committee has a panel of 16 members, including experts in toxicology, molecular and cellular biology, epidemiology, genetics, law and ethics, bioinformatics, statistics, medicine, and public health. The committee approached its charge by focusing on the current and potential uses of toxicogenomics in the study and risk assessment of human responses to environmental and pharmaceutical chemicals. The committee did not consider ecologic assessment within its focus.

The committee owes a great debt of gratitude to the NRC staff members who guided it through production of the final report. Marilee Shelton-Davenport, of the Board on Life Sciences (BLS); and Suzanne van Drunick, former member, and Karl Gustavson, of the Board on Environmental Studies and Toxicology (BEST), provided valuable guidance and contributions. Liza Hamilton, Morgan Motto, and Mirsada Karalic-Loncarevic, of BEST, provided excellent staff support throughout the study.

I, the NRC staff, and the committee are indebted to a number of scientists who presented background information, both orally and in writing, that made the committee's understanding of these issues more complete. Several persons involved at the beginning of the study were Wylie Burke, Thomas Caskey, Barbara Culliton, and Rick Kittles. We also thank the external reviewers for their insightful criticisms and suggestions.

David Christiani, *Chair*
Committee on Applications of Toxicogenomic
Technologies to Predictive Toxicology

Copyright © National Academy of Sciences. All rights reserved.

Acknowledgments

We appreciate the generous support and outstanding assistance provided by Kenneth Olden, David A. Schwartz, and Samuel H. Wilson, all of the National Institute of Environmental Health Sciences (NIEHS); and David Brown, formerly of NIEHS and currently of the Constella Group.

Many people assisted the committee and the National Research Council by providing data and reports. We are grateful for the information and support provided by the following:

Richard A. Becker, American Chemistry Council
Yvonne P. Dragan, National Center for Toxicological Research
Jennifer M. Fostel, National Center for Toxicogenomics, National Institute of Environmental Health Sciences
Kathryn Gallagher, U.S. Environmental Protection Agency
David Jacobson-Kram, Center for Drug Evaluation and Research
Kurt Jarnagin, Iconix Pharmaceuticals
Supriya Jayadev, Boehringer Ingelheim Pharmaceuticals, Inc.
Jeremy K. Nicholson, University of London
Donald G. Robertson, Pfizer Global Research and Development
Bennett Van Houten, National Institute of Environmental Health Sciences
Brenda K. Weis, National Institute of Environmental Health Sciences
Douglas C. Wolf, U.S. Environmental Protection Agency

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Acknowledgment of Review Participants

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We thank the following individuals for their review of this report:

Christopher Amos, University of Texas
Wylie Burke, University of Washington
C. Thomas Caskey, University of Texas
George Daston, Procter & Gamble Company
Lutz Edler, German Cancer Research Center
David Goodlett, University of Washington
William Mattes, Critical Path Institute
Roger McClellan, Chemical Industry Institute of Toxicology
Stephen Roberts, University of Florida
Martha Sandy, California Environmental Protection Agency
Gerald Wogan, Massachusetts Institute of Technology
Timothy Zacharewski, Michigan State University
Lauren Zeise, California Environmental Protection Agency

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Robert A. Goyer, University of Western Ontario, and Gilbert S. Omenn, University of Michigan Medical School. Appointed by the National Research Council, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Abbreviations

ALL	acute lympboblastic leukemia
AML	acute myeloblastic leukemia
ANNs	artificial neural networks
ANOVA	analysis of variance
CART	classification and regression trees
CERCLA	Comprehensive Environmental Remediation, Compensation and Liability Act
CNS	central nervous system
CPSC	Consumer Product Safety Commission
EPA	U.S. Environmental Protection Agency
FDA	Food and Drug Administration
FDR	false discovery rate
FQPA	Food Quality Protection Act
HHS	U.S. Department of Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act of 1996
HPV	high production volume
5-HT_6	5-hydroxytryptamine
ICCA	International Council of Chemical Associations
kNN	*k*-nearest neighbors
LOAEL	lowest-observed-adverse-effect level
MOA	monoamine oxidase
NCI	National Cancer Institute
NOAEL	no-observed-adverse-effect level
OHRP	Office for Human Research Protections
OPPT	Office of Pollution Prevention and Toxics
OSHA	Occupational Safety and Health Administration
PKU	phenylketonuria
PMNs	Pre-Manufacturing Notices
QSAR	quantitative structure-activity relationship
RfC	reference concentration
SAM	significance analysis of microarrays
SDWA	Safe Drinking Water Act
SOMs	self-organizing maps
SVM	support vector machines
TRI	Toxic Release Inventory
TSCA	Toxic Substances Control Act

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Contents

SUMMARY **1**

1 **INTRODUCTION** **11**
Background, 11
Committee Charge and Response, 17

2 **TOXICOGENOMIC TECHNOLOGIES** **22**
Genomic Technologies, 24
Transcriptomic Technologies, 27
Proteomic Technologies, 34
Metabolomic Technologies, 41
Technology Advancement and Economy of Scale, 43

3 **EXPERIMENTAL DESIGN AND DATA ANALYSIS** **45**
Experimental Design, 45
Types of Experiments, 47
Toxicogenomics and Biomarker Discovery, 54
Conclusions, 56
Recommendations, 57

4 **APPLICATION TO EXPOSURE ASSESSMENT** **59**
Conventional Biomarkers of Exposure and Response, 60
State of the Art: Toxicogenomic Approaches to the Development of Exposure Biomarkers, 61
Evaluation of Exposure in Risk Assessment, 67
Scientific Challenges, 68
Conclusions, 71
Recommendations, 72

5 **APPLICATION TO HAZARD SCREENING** **73**
Description and Properties of Current Hazard Screening Methods, 73
In Vitro Versus In Vivo Testing, 76
Development of Predicitve Algorithms for Screening Purposes, 77
State of the Art, 77
Potential Applications, 83

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Challenges and Limitations, 86
Conclusions, 89
Recommendations, 90

6 APPLICATION TO ANALYZING VARIATION IN HUMAN SUSCEPTIBILITY 92
State of the Art in Assessment of Individual Variation, 93
Risk Assessment, 100
Challenges, 102
Conclusions, 103
Recommendations, 105

7 APPLICATION TO THE STUDY OF MECHANISMS OF ACTION 107
State of the Art in Transcriptomics Analyses, 108
Conclusions, 118
Recommendations, 119

8 OTHER POTENTIAL APPLICATIONS OF TOXICOGENOMIC TECHNOLOGIES TO RISK ASSESSMENT 121
Potential Applications to Risk Assessment, 121
Implications for Better Use and Potential Reduction in the Use of Experimental Animals, 131
Implications of Toxicogenomics for Risk Assessment Infrastructure Needs, 132
Conclusions, 133
Recommendations, 133

9 VALIDATION 135
Levels of Validation, 135
Conclusions, 149
Recommendations, 151

10 SAMPLE AND DATA COLLECTION AND ANALYSIS 152
Leveraging Samples and Information in Population-Based Studies and Clinical Trials for Toxicogenomic Studies, 152
Existing Toxicogenomic Data Repositories: Standardization, Availability and Transparency, 161
Conclusions, 166
Recommendations, 169

11 ETHICAL, LEGAL, AND SOCIAL ISSUES 172
Research Issues, 172
Ethical and Social Issues, 178
Regulatory Implications, 186
Litigation, 191
Communication and Education, 195
Conclusions and Recommendations, 197

12 CONCLUSIONS AND RECOMMENDATIONS 201
A New Human Toxicogenomics Initiative, 202
Data Analysis Tools, 203

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Exposure Assessment, 204
Hazard Screening, 205
Variability in Susceptibility, 206
Mechanistic Information, 207
Dose-response Relationships, 208
Cross-species Extrapolation, 209
Developmental Exposures, 209
Mixtures, 209
Validation, 210
Ethical, Legal, and Social Issues, 210
Education and Training in Toxicogenomics, 212

REFERENCES **213**

APPENDIXES

A BIOGRAPHIC INFORMATION ON THE COMMITTEE ON APPLICATIONS OF TOXICOGENOMIC TECHNOLOGIES TO PREDICTIVE TOXICOLOGY **255**

B MINIMUM INFORMATION ABOUT A MICROARRAY EXPERIMENT FOR TOXICOGENOMICS **262**

C OVERVIEW OF RISK ASSESSMENT **270**

BOXES, FIGURES, AND TABLES

BOXES

S-1 Human Toxicogenomics Initiative, 6
1-1 Toxicogenomics Definition, 12
1-2 Typical Information Gaps and Inconsistencies That Limit Conventional Risk Assessment, 14
1-3 Statement of Task, 18
1-4 Selected Examples of the Use of Toxicogenomic Technologies in Predictive Toxicology, 19
2-1 Experimental Details of Transcriptome Profiling with Microarrays, 29
3-1 Supervised and Unsupervised Analysis, 52
3-2 Defining Biomarkers, 55
6-1 Paraoxonase 1 Gene Polymorphisms and Occupational Exposure to Pesticide, 94
6-2 Multiple Genetic Factors Influence Response to Cigarette Smoke, 95
6-3 Summary of Research Issues and Potential Applications, 104
7-1 Types of Experimental Approaches in Toxicogenomics Studies to Evaluate Mechanisms of Toxicity, 109
9-1 Validation of Toxicogenomic Applications, 136

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

9-2 Microarray Gene Expression Data Society Workshop, September 2005, Bergen, Norway, 140
9-3 Clinical Validation of Transcriptome Profiling in Breast Cancer, 145
10-1 Possible Steps for Developing an Initial Database, 170
11-1 Toxicogenomic Research on Workers, 182
11-2 Genetic Screening for Susceptibility to Chronic Beryllium Disease, 184
11-3 BiDil and Race-Linked Pharmaceuticals, 186
11-4 Regulatory Agency Guidance Documents, 188

FIGURES

2-1 Hierarchical relationships of DNA, RNA, proteins, and metabolites, 23
2-2 Overview of commonly used methods and technologies for gene expression analysis, 28
2-3 An overview of DNA microarray analysis, 32
2-4 Schematic representation of 2D gel-based proteome analysis (upper) and shotgun proteome analysis (lower), 35
2-5 A hierarchy of proteomic data, 38
3-1 Overview of the workflow in a toxicogenomic experiment, 46
9-1 Human plasma proteome, 138

TABLES

3-1 Data Analysis Approaches, 48
3-2 Phases of Cancer Biomarker Development As Defined by Pepe et al. (2001), 57
5-1 Elements of the OECD SIDS, 75
9-1 Worldwide Regulatory Policies and Guidelines Relations to Toxicogenomics and Pharmacogenomics, 147

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Summary

The Human Genome Project, which set the goal of determining the complete nucleotide sequence of the human genome, was among the most important biologic research projects of all time. First envisioned in the late 1980s and considered by many to be technologically impossible at the time, it was the combination of adequate resources and strong scientific leadership of the project that fostered development of the requisite rapid DNA sequencing technologies. These new technologies were so successful that the genomic sequence of the bacterium *Haemophilus influenzae* was obtained only a few years later in 1995. Since then, the genomes of dozens of organisms have been elucidated and made available to the research community, and most important, the reference human genome sequence was made available by the year 2000, several years ahead of schedule.

To capitalize on the enormous potential of having access to genome-wide sequence information, scientists, clinicians, engineers, and information scientists combined forces to develop a battery of new molecular and bioinformatic tools that now make it possible to obtain and analyze biologic datasets of unprecedented magnitude and detail. Generally referred to as genomic technologies, these approaches permit sequence analysis—as well as gene transcript, protein, and metabolite profiling—on a genome-wide scale. As a result, the Human Genome Project and the technologic innovations and computational tools that it spawned are having profound effects on biologic research and understanding.

The application of these technologies to toxicology has ushered in an era when genotypes and toxicant-induced genome expression, protein, and metabolite patterns can be used to screen compounds for hazard identification, to monitor individuals' exposure to toxicants, to track cellular responses to different doses, to assess mechanisms of action, and to predict individual variability in sensitivity to toxicants.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

This potential has prompted a plethora of scientific reviews and commentaries about toxicogenomics written over the past several years that attest to the widely held expectation that toxicogenomics will enhance the ability of scientists to study and estimate the risks different chemicals pose to human health and the environment. However, there are limitations in the data that are currently available, and fully understanding what can be expected from the technologies will require a greater consolidation of useful data, tools, and analyses. Given the inherent complexity in generating, analyzing, and interpreting toxicogenomic data and the fact that toxicogenomics cannot address all aspects of toxicology testing, interested parties need to prepare in advance. This preparation will help them understand how best to use these new types of information for risk assessment and for implementing commensurate changes in regulations and public health, while preparing for the potential economic, ethical, legal, and social consequences.

COMMITTEE'S CHARGE

In anticipation of these questions, the National Institute of Environmental Health Sciences (NIEHS) of the U.S. Department of Health and Human Services, asked the National Academies to direct its investigative arm, the National Research Council (NRC), to examine the potential impacts of toxicogenomic technologies on predictive toxicology. NIEHS has invested significant resources in toxicogenomic research through establishment of the National Center for Toxicogenomics, funding of the National Toxicogenomics Research Consortium, development of the Chemical Effects in Biological Systems database for toxicogenomic data, and other collaborative ventures.

In response to the NIEHS request, the NRC assembled a panel of 16 experts with perspectives from academia, industry, environmental advocacy groups, and the legal community. The charge to the committee was to provide a broad overview for the public, government policy makers, and other interested and involved parties of the benefits potentially arising from toxicogenomic technologies; to identify the challenges in achieving them; and to suggest approaches that might be used to address the challenges.

COMMITTEE'S RESPONSE TO ITS CHARGE

The committee clarified its task by defining the terms "toxicogenomics" and "predictive toxicology" as follows:

- *Toxicogenomics* is defined as the application of genomic technologies (for example, genetics, genome sequence analysis, gene expression profiling, proteomics, metabolomics, and related approaches) to study the adverse effects of environmental and pharmaceutical chemicals on human health and the environment. Toxicogenomics combines toxicology with information-dense ge-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

nomic technologies to integrate toxicant-specific alterations in gene, protein, and metabolite expression patterns with phenotypic responses of cells, tissues, and organisms. Toxicogenomics can provide insight into gene-environment interactions and the response of biologic pathways and networks to perturbations. Toxicogenomics may lead to information that is more discriminating, predictive, and sensitive than that currently used to evaluate exposures to toxicants or to predict effects on human health.

- *Predictive toxicology* is used in this report to describe the study of how toxic effects observed in model systems or humans can be used to predict pathogenesis, assess risk, and prevent human disease.

Because of the belief that toxicogenomics has the potential to place toxicology on a more predictive footing, the committee describes the momentum channeling the field in this direction and some of the obstacles in its path. The committee approached its charge by identifying, defining, and describing several proposed applications of toxicogenomics to hazard-identification screening, mechanism-of-action studies, classification of compounds, exposure assessment, defining genetic susceptibility, and reducing the use of animal-based testing. Studies supporting each of these putative applications were then critically evaluated to define limitations, to enumerate remaining challenges, and to propose viable solutions whenever possible. Finally, the committee outlined realistic expectations of how these applications can be validated and how they can be used in risk assessment. The second part of this summary reviews these applications and what is needed for each of them.

In evaluating the putative applications, the committee recognized some overarching themes and steps necessary for the field to move forward.

OVERARCHING CONCLUSIONS AND RECOMMENDATIONS

Reproducibility, Data Analysis, Standards, and Validation

After evaluating the different applications, the committee concluded that, for the most part, the technologic hurdles that could have limited the reproducibility of data from toxicogenomic technologies have been resolved, representing an important step forward. To consolidate this advance, those who use these tools need to make a unified effort to establish objective standards for assessing quality and quality-control measures. Across the different applications, validation efforts are an important next step: actions should be taken to facilitate the technical and regulatory validation of toxicogenomics.

There is also a need for bioinformatic, statistical, and computational approaches and software to analyze data. Thus, the committee recommends the development of specialized bioinformatic, statistical, and computational tools and approaches to analyze toxicogenomic data.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Use of Toxicogenomics in Risk Assessment

Improving risk assessment is an essential aim of predictive toxicology, and toxicogenomic technologies present new opportunities to enhance it by potentially improving the understanding of dose-response relationships, cross-species extrapolations, exposure quantification, the underlying mechanisms of toxicity, and the basis of individual susceptibilities to particular compounds.

Although the applications of toxicogenomic technologies to risk assessment and the regulatory decision-making process have been exploratory to date, the potential to improve risk assessment has just begun to be tapped. Toxicogenomic technologies clearly have strong potential to affect decision making, but they are not currently ready to replace existing required testing regimes in risk assessment and regulatory toxicology. Toxicogenomic technologies are assuming an increasing role as adjuncts to and extensions of existing technologies for predictive toxicology. Toxicogenomics can provide additional molecular level information and tests that add to the "weight of the evidence" for refining judgments about the risks posed by environmental toxicants and drugs. Ultimately, however, they are envisioned to be more sensitive and informative than existing technologies and may supplant some approaches currently used or at least be a component of batteries that will replace certain tests.

To move forward, the committee recommends that regulatory agencies enhance efforts to incorporate toxicogenomic data into risk assessment. The following actions are needed: (1) substantially enhance agencies' capability to effectively integrate toxicogenomic approaches into risk assessment practice, focusing on the specific applications below; (2) invest in research and personnel within the infrastructure of regulatory agencies; and (3) develop and expand research programs dedicated to integrating toxicogenomics into challenging risk assessment problems, including the development of partnerships between the public and private sectors.

Need for a Human Toxicogenomics Initiative

Several themes emerged throughout evaluation of the different applications discussed below, including the need for more data, the need to broaden data collection, the need for a public database to facilitate sharing and use of the volumes of data, and the need for tools to mine this database to extract biologic knowledge.

Concerted efforts are necessary to address these needs and propel the field forward. Fully integrating toxicogenomic technologies into predictive toxicology will require a coordinated effort approaching the scale of the Human Genome Project. It will require funding and resources significantly greater than what is allocated to existing research programs and will benefit from public-private partnerships to achieve its goals. These types of investments and coordinated scientific leadership will be essential to develop toxicogenomic tools to

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

the point where many of the expected benefits for predicting the toxicity of compounds and related decision making can be realized.

To achieve this goal, NIEHS should cooperate with other stakeholders to explore the feasibility and objectives of a human toxicogenomics initiative (HTGI), as described in Box S-1. The HTGI would support the collection of toxicogenomic data and would coordinate the creation and management of a large-scale database that would use systems biology approaches and tools to integrate the results of toxicogenomic analyses with conventional toxicity testing data.

The information generated from toxicogenomic experiments is on a scale vastly exceeding DNA sequencing efforts like the Human Genome Project. The heft of these outputs, consisting of multidimensional datasets that include genotype, gene expression, metabolite, and protein information; design factors such as dose, time, and species information; and information on toxicologic effects warrant the creation of a public database. This database is needed to compile and analyze the information at a more complex level than the current "one disease is caused by one gene" approach. Curation, storage, and mining of these data will require developing and distributing specialized bioinformatic and computational tools. Current public databases are inadequate to manage the types or volumes of data to be generated by large-scale applications of toxicogenomic technologies and to facilitate the mining and interpretation of the data, which are just as important as their generation and storage.

Although the database and tools are important, the database itself is not sufficient. Data on a large number of compounds are needed so that comparisons can be made and data can be mined to identify important relationships. To collect and generate these toxicogenomic data, it will be important to leverage large publicly funded studies and facilitate the production and sharing of private sector data.

In addition to data, work is needed in the collection of physical samples appropriate for toxicogenomic research. Specifically, a national biorepository for human clinical and epidemiologic samples is needed so that toxicogenomic data can eventually be extracted from them. In addition, when possible and appropriate, the collection of samples and data should be incorporated into major human studies. The collection of human samples and their corresponding data raises a number of ethical, legal, and social issues of the type described below, which need to be addressed.

Because the realization of the goals articulated here will require significantly higher levels of funding, leadership, and commitment than are currently allocated to toxicogenomics, planning and organizing research should begin immediately. Collaborations among government, academia, and the private sector not only will expedite discovery but will ensure optimal use of samples and data; prevent unnecessary duplication or fragmentation of datasets; enhance the ability to address key ethical, legal, and social effects; reduce costs; and promote intellectual synergy.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX S-1 Human Toxicogenomics Initiative

NIEHS should cooperate with other stakeholders in exploring the feasibility and objectives of implementing a human toxicogenomics initiative (HTGI) dedicated to advancing toxicogenomics. Elements of the HTGI should include the following:

1. Creation and management of a large, public database for storing and integrating the results of toxicogenomic analyses with conventional toxicity-testing data.
2. Assembly of toxicogenomic and conventional toxicologic data on a large number (hundreds) of compounds into the single database. This includes the generation of new toxicogenomic data from humans and animals for a number of compounds for which other types of data already exist as well as the consolidation of existing data. Every effort should be made to leverage existing research studies and infrastructure (such as those of the National Toxicology Program) to collect samples and data that can be used for toxicogenomic analyses.
3. Creation of a centralized national biorepository for human clinical and epidemiologic samples, building on existing efforts.
4. Further development of bioinformatic tools, such as software, analysis, and statistical tools.
5. Consideration of the ethical, legal, and social implications of collecting and using toxicogenomic data and samples.
6. Coordinated subinitiatives to evaluate the application of toxicogenomic technologies to the assessment of risks associated with chemical exposures.

The resulting publicly accessible HTGI data resource would strengthen the utility of toxicogenomic technologies in toxicity assessment and thus enable more accurate prediction of health risks associated with existing and newly developed compounds and formulations.

SPECIFIC APPLICATIONS OF TOXICOGENOMICS

To address the expectation that toxicogenomics will revolutionize predictive toxicology, the committee explored several proposed applications of toxicogenomics, including hazard screening, the study of toxicologic mechanisms of action, exposure assessment, and characterizing variability in susceptibility. These and the other applications can be used in conjunction with risk assessment, although they are also important in predictive toxicology, which is removed from the risk assessment process. In the following sections, the committee reports findings from the evaluation of these topics that were assimilated into the conclusions of the report.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Exposure Assessment

The application of toxicogenomics for defining biomarkers of exposure will require consensus on what constitutes an exposure biomarker. Standardized toxicogenomic platforms that are appropriate for identifying signatures of environmental or drug exposures in target and surrogate tissues and fluids will also be required. Additional technical challenges include the individual variation in response to an environmental exposure and the persistence of a toxicogenomic signature after exposure.

Toxicogenomic technologies should be adapted and applied for the study of exposure assessment by developing signatures of exposure to individual chemicals and perhaps to chemical mixtures. To facilitate the development of exposure-assessment tools based on toxicogenomics, large human population studies should include a collection of samples that can be used for transcriptomic, proteomic, metabolomic, or other toxicogenomic analyses in addition to traditional epidemiologic measures of exposure.

Hazard Screening

Toxicogenomic technologies provide new and potentially useful indicators for use in toxicity screening. Near-term applications include current uses in drug development and the validation of categories of compounds for screening chemicals found in the environment. In contrast to applications in evaluating new drug candidates, screening approaches for environmental chemicals will need to address a broader range of exposure levels and a more comprehensive set of adverse health effects.

Toxicogenomic screening methods should be integrated into relevant current and future chemical regulatory and safety programs upon validation and development of adequate databases. To move toward this goal, it is important to improve the quantity and quality of data available for deriving screening profiles and to develop a database to organize this information. The process of creating such a database could be accelerated by addressing proprietary and legal hurdles so at least some of the toxicogenomic data currently in private databases could be made available, and by integrating toxicogenomic assays into ongoing chemical screening and testing initiatives such as those conducted by the National Toxicology Program. In addition, regulatory agencies should continue to develop and refine guidance documents for their staff on interpreting toxicogenomic data.

Variability in Susceptibility

People vary in their susceptibility to toxic effects of chemical exposures. Toxicogenomic technologies (including the analysis of gene sequences and epigenetic modifications) offer the opportunity to use genetic information in a pro-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

spective fashion to identify susceptible subpopulations and assess the distribution of differences in susceptibility in larger populations. Toxicogenomic technologies could also reduce the uncertainty surrounding assumptions used in regulatory processes to address population variability.

Toxicogenomic information should be used to prospectively identify, understand the mechanisms of, and characterize the extent of genetic and epigenetic influences on variations in human susceptibility to the toxic effects of chemicals. Animal models and genome-wide human studies should be used to identify genetic variations that influence sensitivity to chemicals, using existing large human studies when possible to investigate the effect of genetic variations on responses to a wide array of chemical exposures and pharmaceutical therapies. More attention should be focused on modeling effects involving multiple genes and the study of context-dependent genetic effects (that is, gene-gene interactions as well as the impact of developmental age, sex, and life course).

Mechanistic Information

Toxicogenomic studies are improving knowledge of the molecular level events that underlie toxicity and may thus advance the consideration of mechanistic information in risk assessment and decision making. Tools and approaches should continue to be developed to advance the ability of toxicogenomics to provide useful mechanistic information. Developing richer algorithms and models that can integrate complex and various types of toxicogenomic data (for example, metabolomic and proteomic data) may make it possible to shift the focus of mechanistic investigations of single genes to more integrated analyses. These will encompass more of the complexity of biologic systems as a whole as well as the multidimensionality of the dose- and time-related effects of toxicants.

Cross-Species Extrapolation

Toxicogenomic technologies offer the potential to significantly enhance confidence in animal-to-human toxicity extrapolations that constitute the foundation of risk evaluations. Using toxicogenomics to analyze species differences in toxicity will help explain the molecular basis for the differences and improve the translation of animal observations into estimates of potential human risk. In addition, by providing molecular level comparisons between humans and other species, toxicogenomics may assist in identifying those animal species and strains that are most relevant for specific assays.

Toxicogenomics should continue to be used to study differences in toxicant responses between animal models and humans, and genotyped and genetically altered animal model strains should continue to be used as experimental tools to better extrapolate results from animal tests to human health. Algorithms must be developed to facilitate accurate identification of genes and proteins that

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

serve the same function in different organisms and species—called orthologous genes and proteins—used in toxicologic research.

Dose-Response Relationships

Toxicogenomics has the potential to improve the understanding of dose-response relationships, particularly at low doses. Future toxicologic assessment should incorporate dose-response and time-course analyses appropriate to risk assessment. Analyses of toxic compounds that are well characterized could provide an intellectual framework for future studies.

Developmental Exposures

Although recognized to be important in a number of disorders, relatively little is known about the health impacts of fetal and early-life exposures to many chemicals in current use. Because of their sensitivity, toxicogenomic technologies are expected to reveal more than previously was possible about the molecules involved in development and the critical molecular level events that can be perturbed by toxicants. Toxicogenomics may also enable screening for chemicals that cause gene expression changes associated with adverse developmental effects. In short, toxicogenomic technologies should be used to investigate how exposure during early development conveys susceptibility to drug and chemical toxicities.

Mixtures

Although much toxicology focuses on the study of single chemicals, humans are frequently exposed to multiple chemicals. It is difficult to decipher how exposure to many chemicals will influence the effects of each one. It is unlikely that toxicogenomic signatures will be able to decipher all interactions among complex mixtures, but it should be possible to use mechanism-of-action data to design informative toxicogenomic experiments, including screening chemicals for potential points of biologic conversion (overlap) such as shared activation and detoxification pathways, enhancing identification and exploration of potential interactions, and moving beyond empirical experiments. Toxicogenomic approaches should be used to test the validity of methods for the ongoing challenge of estimating potential risks associated with mixtures of environmental chemicals.

ETHICAL, LEGAL, AND SOCIAL ISSUES

The committee evaluated ethical, legal, and social implications of toxicogenomics. As toxicogenomic data linked to clinical and epidemiologic informa-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

tion are collected, it is critical to ensure adequate protections of the privacy, confidentiality, and security of toxicogenomic information in health records and information used in studies. Safeguarding this information will further advance important individual and societal interests. It will also prevent individuals from being dissuaded from participating in research or undergoing the genetic testing that is the first step in individualized risk assessment and risk reduction.

Toxicogenomics is also likely to play a role in occupational, environmental, and pharmaceutical regulation and litigation. Regulatory agencies and courts should give appropriate weight to the validation, replication, consistency, sensitivity, and specificity of methods when deciding whether to rely on toxicogenomic data.

Ethical, legal, and social issues that affect the use of toxicogenomic data and the collection of data and samples needed for toxicogenomic research should be addressed. This could occur through legislative improvements to enhance individual protection, exploration of how to facilitate large-scale biorepository and database research while protecting individuals, and consideration by courts and regulatory agencies of appropriate factors when deciding how to consider toxicogenomic data. Finally, special efforts should be made to address the impact of toxicogenomic research and findings on vulnerable populations.

EDUCATION AND TRAINING IN TOXICOGENOMICS

Given the complexity of toxicogenomics, the generation, analysis, and interpretation of toxicogenomic information represents a challenge to the scientific community and requires the collaborative cross-disciplinary efforts of scientific teams of specialists. Therefore it is essential that education and training in toxicogenomics become a continuous, ongoing process that reflects the rapid developments in these new technologies. There is a need to develop education and training programs relevant to toxicogenomic applications to predictive toxicology. Specifically, programs are needed to reach the general public, susceptible subgroups, health professionals, government regulators, attorneys and judges, the media, scientists in training, scientists on the periphery of toxicogenomics, and institutions that participate in toxicogenomic research.

CONCLUSIONS

In summary, toxicogenomic technologies present a set of powerful tools for transforming current observation-based approaches into predictive science, thereby enhancing risk assessment and public health decision making. To leverage this potential will require more concerted efforts to generate data, make multiple uses of existing data sources, and develop tools to study data in new ways. Beyond the technical challenges and opportunities, other challenges in the communication, education, ethics, and legal arenas will need to be addressed to ensure that the potential of the field can be realized.

Copyright © National Academy of Sciences. All rights reserved.

1

Introduction

Scientists, regulators, and the public all desire efficient and accurate approaches to assess the toxicologic effects of chemical, physical, and biologic agents on living systems. Yet, no single approach exists to analyze toxicologic responses, a difficult task given the complexity of human and animal physiology and individual variations. The genomic knowledge and new technologies that have emerged in the post-genomic era promise to inform the understanding of many risks as well as enlighten current approaches and lead to novel predictive approaches for studying disease risk. As biologic knowledge progresses with the science of toxicology, "toxicogenomics" (see Box 1-1 for definition) has the potential to improve risk assessment and hazard screening.

BACKGROUND

Approaches to Assessing Toxicity

Detection of toxicity requires a means to observe (or measure) specific effects of exposures. Toxicology traditionally has focused on phenotypic changes in an organism that result from exposure to chemical, physical, or biologic agents. Such changes range from reversible effects, such as transient skin reactions, to chronic diseases, such as cancer, to the extreme end point of death. Typical whole-animal toxicology studies may range from single-dose acute to chronic lifetime exposures, and (after assessment of absorption, distribution, metabolism, and excretion properties) they include assessments of end points such as clinical signs of toxicity, body and organ weight changes, clinical chemistry, and histopathologic responses.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 1-1 Toxicogenomics Definition

Toxicogenomics: In this report, *toxicogenomics* is defined as the application of genomic technologies (for example, genetics, genome sequence analysis, gene expression profiling, proteomics, metabolomics, and related approaches) to study the adverse effects of environmental and pharmaceutical chemicals on human health and the environment. Toxicogenomics combines toxicology with information-dense[1] genomic technologies to integrate toxicant-specific alterations in gene, protein, and metabolite expression patterns with phenotypic[2] responses of cells, tissues, and organisms. Toxicogenomics can provide insight into gene-environment interactions and the response of biologic pathways and networks to perturbations. Toxicogenomics may lead to information that is more discriminating, predictive, and sensitive than that currently used to evaluate toxic exposure or predict effects on human health.

Toxicology studies generally use multiple doses that span the expected range from where no effects would be observed to where clinical or histopathologic changes would be evident. The highest dose at which no overt toxicity occurs in a 90-day study (the *maximum tolerated dose*), is generally used to establish animal dosing levels for chronic assays that provide insight into potential latent effects, including cancer, reproductive or developmental toxicity, and immunotoxicity. These studies constitute the mainstays of toxicologic practice.[3]

In addition to animal studies, efforts to identify and understand the effects of environmental chemicals, drugs, and other agents on human populations have used epidemiologic studies to examine the relationship between a dose and the response to exposures. In contrast to animal studies, in which exposures are experimentally controlled, epidemiologic studies describe exposure with an estimate of error, and they assess the relationship between exposure and disease distribution in *human* populations. These studies operate under the assumption that many years of chemical exposures or simple passage of time may be required before disease expression can be detected.

[1]Toxicogenomic approaches are often referred to as "high-throughput," a term that can refer to the density of information or the ability to analyze many subjects (or compounds) in a short time. Although the toxicogenomic techniques described here create information that is highly dense, the techniques do not all offer the ability to analyze many subjects or compounds at one time. Therefore, the term high-throughput is not used except in reference to gene sequencing technologies.

[2]Relating to "the observable properties of an organism that are produced by the interaction of the genotype and the environment" (Merriam-Webster's Online Dictionary, 10th Edition).

[3]For more information, see the National Research Council report Toxicity Testing for Assessment of Environmental Agents: Interim Report (NRC 2006a).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

As medical science has progressed, so have the tools used to assess animal toxicity. For example, more sensitive diagnostic and monitoring tools have been used to assess organ function, including tools to detect altered heart rhythms, brain activity, and changes in hormone levels as well as to analyze changes visible by electron microscopy. Most notable, however, are the contributions of chemistry, cell and molecular biology, and genetics in detecting adverse effects and identifying cellular and molecular targets of toxicants. It is now possible to observe potential adverse effects on molecules, subcellular structures, and organelles before they manifest at the organismal level. This ability has enhanced etiologic understanding of toxicity and made it possible to assess the relevance of molecular changes to toxicity.

These molecular and cellular changes have been assessed in studies of animals but have also been applied to study *human* populations ("molecular epidemiology") with some success. For example, our understanding of gene-environment interactions has benefited greatly from studies of lung, head, and neck cancer among tobacco users—studies that examined differences in genes (polymorphisms) that are related to carcinogen metabolism and DNA repair. Similarly, studies of UV sunlight exposure and human differences in DNA repair genes have clarified gene-environment interactions in skin cancer risk. Current technology now enables the role of multiple genes of cell signaling pathways to be examined in human population studies aimed at assessing the interplay between environmental exposures and cancer risk.

Although current practice in toxicology continues to strongly emphasize changes observable at the level of the whole organism as well as at the level of the organ, the use of cellular and molecular end points sets the stage for applying toxicogenomic technologies to a more robust examination of how complex molecular and cellular systems contribute to the expression of toxicity.

Predictive Toxicology

Predictive toxicology describes the study of how toxic effects observed in humans or model systems can be used to predict pathogenesis, assess risk, and prevent human disease. Predictive toxicology includes, but is not limited to, risk assessment, the practical facilitation of decision making with scientific information. Many of the concepts described in this report relate to approaches to risk assessment; key risk assessment concepts are reviewed in Appendix C. Typical information gaps and inconsistencies that limit conventional risk assessment are listed in Box 1-2. These gaps and inconsistencies present opportunities for toxicogenomics to provide useful information.

Although toxicogenomics includes effects on wildlife and other environmental effects, this report is limited to a discussion of toxicogenomics as it applies to the study of human health.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 1-2 Typical Information Gaps and Inconsistencies That Limit Conventional Risk Assessment (Modified from NRC 2005a)

- Lack of sufficient screening data—basic short-term in vitro or animal-bioassay data on toxicity or carcinogenicity of the compound.
- Lack of information or inconsistent information about effects on humans—epidemiologic studies.
- Paucity of accurate information on human exposure levels.
- Relevance of animal data to humans—quantitative or qualitative.
- Paucity of information on the relationship between dose and response, especially at low doses relevant to human environmental exposures.
- Inconsistent animal-bioassay data on different species—differential responses in varied animal test models.
- Paucity of information or inconsistencies in data on different exposures, particularly exposure during development and early-life exposures or by varied routes of exposure (inhalation, diet, drinking water).
- Lack of data on impacts of coexposures to other chemicals—current risk assessment practices are uncertain about "how to add" coexposures to the many and varied chemicals present in real-world environments.
- Paucity of data on the impact of human variability on susceptibility, including age, gender, race, disease, and other confounders.

Overview of Toxicogenomic Technologies

Toxicogenomic technologies comprise several different technology platforms for analysis of genomes, transcripts, proteins, and metabolites. These technologies are described briefly here and in more detail in Chapter 2. It is important to recognize two additional issues associated with the use of toxicogenomic technologies. First, the large quantity of information that a single experiment can generate, and the comprehensive nature of this information, is much greater than what traditional experiments generate. Second, the advancement of computing power and techniques enable these large amounts of information to be synthesized from different sources and experiments and to be analyzed in novel ways.

Genomic technologies encompass both genome sequencing technologies, which derive DNA sequences from genes and other regions of DNA, and genotype analysis, which detects sequence variations between individuals in individual genes. Whereas the sequencing of genomes was once an extraordinary undertaking, rapid evolution of sequencing technology has dramatically increased throughput and decreased cost, now outperforming the benchmark technology standard used for the Human Genome Project. The convergence of genome sequencing and genotyping technologies will eventually enable whole-genome sequences of individuals to be analyzed. Advances in genotyping technologies

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

(see Chapter 2) allow the simultaneous assessment of multiple variants across the whole genome in large populations rather than just single or several gene polymorphisms.

Transcriptomic technologies (or gene expression profiling) measure mRNA expression in a highly parallel assay system, usually using microarrays. As the first widely available method for global analysis of gene expression, DNA microarrays are the emblematic technology of the post-genomic era. Microarray technology for transcriptomics has enabled the analysis of complex, multigene systems and their responses to environmental perturbations.

Proteomics is the study of collections of proteins in living systems. Because the same proteins may exist in multiple modified and variant forms, proteomes are more complex than the genomes and transcriptomes that encode them. Proteomic technologies use mass spectrometry (MS) and microarray technologies to resolve and identify the components of complex protein mixtures, to identify and map protein modifications, to characterize protein functional associations, and to compare proteomic changes quantitatively in different biologic states.

Metabolomics is the study of small-molecule components of biologic systems, which are the products of metabolic processes. Because metabolites reflect the activities of RNAs, proteins, and the genes that encode them, metabolomics allows for functional assessment of diseases and drug and chemical toxicity. Metabolomics technologies, employing nuclear magnetic resonance spectroscopy and MS, are directed at simultaneously measuring dozens to thousands of compounds in biofluids (for example, urine) or in cell tissue extracts. A key strength of metabolomic approaches is that they can be used to noninvasively and repeatedly measure changes in living tissues and living animals and that they measure changes in the actual metabolic flow. As with proteomics, the major limitation of metabolomics is the difficulty of comprehensively measuring diverse metabolites in complex biologic systems.

Bioinformatics is a branch of computational biology focused on applying advanced computational techniques to the collection, management, and analysis of numerical biologic data. Elements of bioinformatics are essential to the practice of all genomic technologies. Bioinformatics also encompasses the integration of data across genomic technologies, the integration of genomic data with data from other observations and measurements, and the integration of all these data in databases and related information resources. It is helpful to think of bioinformatics not as a separate discipline but as the universal means of analyzing and integrating information in biology.

Policy Context

Regulatory agencies with the biggest stake in predictive toxicology include the EPA, the Occupational Safety and Health Administration (OSHA), and the Food and Drug Administration (FDA). The EPA and OSHA are concerned

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

with potentially toxic exposures in the community and in the workplace. The mission of the EPA is to protect human health and the environment and to safeguard the nation's air, water, and land. OSHA's mission is to ensure the safety and health of America's workers by setting and enforcing standards; providing training, outreach, and education; establishing partnerships; and encouraging continual improvement in workplace safety and health. The FDA is responsible for protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biologic products, medical devices, the nation's food supply, cosmetics, and products that emit radiation. The FDA is also responsible for advancing public health by facilitating innovations that make medicines and foods more effective, safer, and more affordable. Finally, the FDA is responsible for helping the public receive the accurate, science-based information it needs to use these regulated products to improve health.

Working in parallel with these agencies and providing additional scientific underpinning to regulatory agency efforts is the Department of Health and Human Services (DHHS), National Institutes of Health (NIH). The NIH (2007) mission is "science in pursuit of fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability," including research on "causes, diagnosis, prevention, and cure of human disease."

The NIH National Institute of Environmental Health Sciences (NIEHS) strives to use environmental sciences to understand human disease and improve human health, including "how environmental exposures fundamentally alter human biology" and why some people develop disease in response to toxicant exposure and others do not.[4]

In sum, NIH, regulatory agencies, the chemical and pharmaceutical industries, health professionals, attorneys, the media, and the general public are all interested in knowing how new genomic technologies developed in the aftermath of the Human Genome Project can improve our understanding of toxicity and ultimately protect public health and the environment. Although the FDA and the EPA have developed planning documents on toxicogenomic policies (see Chapter 9), specific policies have not yet emerged, and it is clear that stakeholders are grappling with similar questions:

- Where is the science of toxicogenomics going?
- What are the potential benefits and drawbacks of using toxicogenomics information for regulatory agencies, industry, and the public?
- What are the challenges in implementing toxicogenomic technologies, collecting and using the data, and communicating the results?
- Can genomic technologies predict health effects?
- How will government agencies, industry, academics, and others know when a particular technology is ready to be used for regulatory purposes?

[4]See http://www.niehs.nih.gov/od/fromdir.htm (accessed April 2, 2007).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

- Will regulatory requirements have to be changed to reap the benefits of the new technologies to protect public health and the environment?

COMMITTEE CHARGE AND RESPONSE

In April 2004, NIEHS asked the National Academies to direct its investigative arm, the National Research Council (NRC), to examine the impact of toxicogenomic technologies on predictive toxicology (see Box 1-3 for the complete statement of task).

In response, the NRC formed the Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology, a panel of 16 members that included experts in toxicology, molecular and cellular biology, epidemiology, law and ethics, bioinformatics (including database development and maintenance), statistics, public health, risk communication, and risk assessment (see Appendix A for committee details). The committee held two public meetings in Washington, DC, to collect information, meet with researchers and decision makers, and accept testimony from the public. The committee met five additional times, in executive session, to deliberate on findings and complete its report. The remaining chapters of this report constitute the findings of the NRC committee.

The committee approached its charge by focusing on potential uses of toxicogenomics often discussed in the broad toxicology community, with a focus on human health issues and not environmental impact. The committee determined that the applications described in Chapters 4-8 capture much of the often-cited potential value of toxicogenomics. After identifying potential toxicogenomic applications, the committee searched the scientific literature for case examples that demonstrate useful implementations of toxicogenomic technologies in these areas.

This report is not intended to be a compendium of all studies that used toxicogenomic technologies and does not attempt to highlight the full range of papers published. Peer-reviewed and published papers in the public domain were selected to illustrate applications the committee identified as worthy of consideration. For example, Box 1-4 contains brief summaries of selected studies where toxicogenomic technologies have shown promise in predictive toxicology. New studies using toxicogenomic technologies are published almost daily. Likewise, approaches to analyzing and interpreting data are rapidly evolving, resulting in changes in attitudes toward various approaches.[5] Even while this report was being prepared, such changes were observed, and therefore the committee has attempted to provide a snapshot of this rapidly evolving field.

This report is the product of the efforts of the entire NRC committee. The report underwent extensive, independent external review overseen by the NRC

[5]The value of platforms and software is also evolving rapidly and any discussion of platforms or software should not be considered an endorsement by the committee.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Report Review Committee. It specifically addresses, and is limited to, the statement of task as agreed upon by the NRC and the DHHS.

This report consists of chapters on existing or potential "applications" and chapters that deal with broader issues. The technologies encompassed in toxicogenomics are described in Chapter 2. This is followed by a discussion of experimental design and data analysis in Chapter 3. Chapter 4 discusses the applications of toxicogenomic technologies to assess exposure: "Can toxicogenomic technologies determine whether an individual has been exposed to a substance and, if so, to how much?" Chapter 5 asks "Can toxicogenomic data be used to detect potential toxicity of an unknown compound quickly, reliably, and at a reasonable cost?" Chapter 6 addresses the assessment of individual variability in humans. The question in this context is "Can toxicogenomic technologies detect variability in response to exposures and provide a means to explain variability between individuals?" Chapter 7 addresses the question "What can toxicogenomic technologies teach us about the mechanisms by which toxicants produce adverse effects in biologic systems?" Considerations for risk assessment not covered in these four application chapters are discussed in Chapter 8. Chapter 9 focuses on validation issues that are relevant to most of the applications. In Chapter 10, sample and data collection and analysis are discussed, as are database needs. The ethical, legal, and social issues raised by use of toxicogenomics are considered in Chapter 11. Finally, Chapter 12 summarizes the recommendations from the other chapters and identifies several overarching recommendations.

BOX 1-3 Statement of Task

A committee of the NRC will examine the impact of "toxicogenomic" technologies on predictive toxicology. These approaches include studying gene and protein activity and other biologic processes to begin to characterize toxic substances and their potential risks. For the promise of these technologies to be realized, significant challenges must be recognized and addressed. This study will provide a broad overview for the public, senior government policy makers, and other interested and involved parties of the benefits potentially arising from these technologies, identify the challenges to achieving them, and suggest approaches and incentives that may be used to address the challenges.

Potential scientific benefits might include identifying susceptible populations and mechanisms of action and making better use of animal toxicity testing. Potential challenges might include scientific issues such as correlating gene expression with adverse effects; conflicting, nonexistent, or inadequate regulatory requirements; legal, social, and ethical issues; coordination between regulators and the regulated communities; organizational infrastructure for handling large volumes of data, new analytic tools, and innovative ways to synthesize and interpret results; communication with appropriate audiences about scientific and nonscien-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

tific information; and the need for scientific standards in conducting and interpreting toxicogenomic experiments.

This study will highlight major new or anticipated uses of these technologies and identify the challenges and possible solutions to implementing them to improve the protection of public health and the environment.

BOX 1-4 Selected Examples of the Use of Toxicogenomic Technologies in Predictive Toxicology

Predictive toxicology is predicated on the hypothesis that similar treatments leading to the same end point will share comparable changes in gene expression. Examples where toxicogenomic technologies have such shown promise in predictive toxicology are presented below. Explanations of the technologies, methodologies, and concepts are presented in greater detail in later chapters of the report.

1. Steiner et al. (2004) evaluated different classes of toxicants by transcript profiling in male rats treated with various model compounds or the appropriate vehicle controls. The results of this study demonstrated the feasibility of compound classification based on gene expression profile data. Most of the compounds evaluated were either well-known hepatotoxicants or showed hepatatotoxicity during preclinical testing. These compounds included acetaminophen, amiodarone, aflatoxin B1, carbon tetrachloride, coumarin, hydrazine, 1,2-dichlorobenzene, and 18 others. The aim was to determine whether biologic samples from rats treated with these various compounds can be classified based on gene expression profiles. Hepatic gene expression profiles were analyzed using a supervised learning method (support vector machines [SVM]) to generate classification rules and combine this with recursive feature elimination to identify a compact set of probe sets with potential use as biomarkers. In all studies, more than 150 genes were expressed above background and showed at least a 2-fold modulation with a p value of <0.05 (two-tailed, unpaired t test).

The predictive models were able to discriminate between hepatotoxic and non-hepatotoxic compounds. Furthermore, they predicted the correct class of hepatotoxicant in most cases. As described in this report (Chapter 5), the predictive model produced virtually no false-positive outcomes but at the cost of some false-negative results—amiodarone, glibenclamide, and chlorpromazine were not recognized as toxic. This example is also important in that the it shows that a predictive model based on transcript profiles from the Wistar rat strain can successfully classify profiles from another rat strain (Sprague-Dawley) for the peroxisome proliferator WY14643. In addition, the model identified non-responding animals (those treated with a toxicant but not exhibiting conventional toxicologic effects).

2. Ruepp et al. (2002) performed both genomic and proteomic analyses of acetaminophen toxicity in mouse liver. Acetaminophen overdose causes severe

(Continued on next page)

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 1-4 Continued

centrilobular hepatic necrosis in humans and in experimental animals. In this case, the researchers explored the mechanism of toxicity using sub toxic and toxic doses to overnight fasted mice. Animals were sacrificed at different time points from 15 minutes to 4 hours postinjection. Liver toxicity was assessed by plasma ALT activity (a liver enzyme marker) and by electron microscopy. Using RT-PCR, genomic expression analysis was performed. In addition, proteomic analysis on liver mitochondrial subfractions using a quantitative fluorescent 2D-DIGE method was done. The results showed Kupffer cell-derived GM-CSF mRNA (GM-CSF is a granulocyte specific gene) induction at both doses acutely, and chaperone proteins Hsp10 and Hsp60 decreased in mitochondria at both doses, most likely by leaking into the cytoplasm. All of these perturbations occurred before morphologic changes.

Other genomic studies of acetaminophen have shown that its hepatotoxicity can be reproduced. Liver diseases that induce nonuniform lesions often give rise to greatly varying histopathology results in needle biopsy samples from the same patient. Heinloth et al. (2007) examined whether gene expression analysis of such biopsies utilizing acetaminophen as a model hepatotoxicant that gives a multifocal pattern of necrosis following toxic doses. Rats were treated with a single toxic or subtoxic dose of acetaminophen and sacrificed 6, 24, or 48 hours after exposure. Left liver lobes were harvested, and both gene expression and histopathologic analysis were performed on the same biopsy-sized samples. While histopathologic evaluation of such small samples revealed significant sample to sample differences after toxic doses of acetaminophen, gene expression analysis provided a very homogeneous picture and allowed clear distinction between subtoxic and toxic doses. The results show that the use of genomic analysis of biopsy samples together with histopathologic analyses could provide a more precise representation of the overall condition of a patient's liver than histopathologic evaluation alone.

3. Haugen et al. (2004) examined arsenic-response networks in *Saccharomyces cerevisiae* by employing global gene expression and sensitivity phenotype data in a metabolic network composed of all known biochemical reactions in yeast, as well as the yeast network of 20,985 protein-protein/protein-DNA interactions. Arsenic is a nonmutagenic carcinogen to which millions of people are exposed. The phenotypic-profiling data mapped to the metabolic network. The two significant metabolic networks unveiled were shikimate and serine, threonine and glutamate biosynthesis. Transcriptional profiling of specific deletion strains confirmed that the several transcription factors strongly mediate the cell's adaptation to arsenic-induced stress. By integrating phenotypic and transcriptional profiling and mapping the data onto the metabolic and regulatory networks, the researchers demostrated that arsenic is likely to channel sulfur into glutathione for detoxification; that leads to indirect oxidative stress by depleting glutathione pools and alters protein turnover via arsenation of sulfhydryl groups on proteins. As described by Haugen et al., "Our data show that many of the most sensitive genes . . . are involved in serine and threonine metabolism, glutamate, aspartate and arginine metabolism, or shikimate metabolism, which are pathways upstream of the differentially expressed sulfur, methionine and homocysteine metabolic pathways,

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

respectively. These downstream pathways are important for the conversion to glutathione, necessary for the cell's defense from arsenic. . . . This overlap of sensitive upstream pathways and differentially expressed downstream pathways provides the link between transcriptional and phenotypic profiling data.”

Copyright © National Academy of Sciences. All rights reserved.

2

Toxicogenomic Technologies

All the information needed for life is encoded in an organism's genome. In humans, approximately 25,000 genes encode protein products, which carry out diverse biologic functions. The DNA segments that compose genes are transcribed to create messenger RNA (mRNA), each of which in turn is translated to proteins (Figure 2-1). Other DNA sequences encode short RNA molecules that regulate the expression of genes and the stability and processing of mRNA and proteins. The integration of genomics, transcriptomics, proteomics, and metabolomics provides unprecedented opportunities to understand biologic networks that control responses to environmental insults. Toxicogenomics uses these new technologies to analyze genes, genetic polymorphisms, mRNA transcripts, proteins, and metabolites. The foundation of this field is the rapid sequencing of the human genome and the genomes of dozens of other organisms, including animals used as models in toxicology studies. Whereas the human genome project was a multiyear effort of a consortium of laboratories, new rapid, high-density, and high-efficiency sequencing approaches allow a single laboratory to sequence genomes in days or weeks. Gene sequencing technologies have also enabled rapid analysis of sequence variation in individual genes, which underlies the diversity in responses to chemicals and other environmental factors.

Perhaps the most emblematic technology of the post-genomic era is the microarray, which makes it possible to simultaneously analyze many elements of a complex system. The integration of microarray methods with the polymerase chain reaction (PCR) has made the analysis of mRNAs the first and most technologically comprehensive of all the -omics technologies. Rapid advances in mass spectrometry (MS) and nuclear magnetic resonance (NMR) have driven the development of proteomics and metabolomics, which complement trans-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

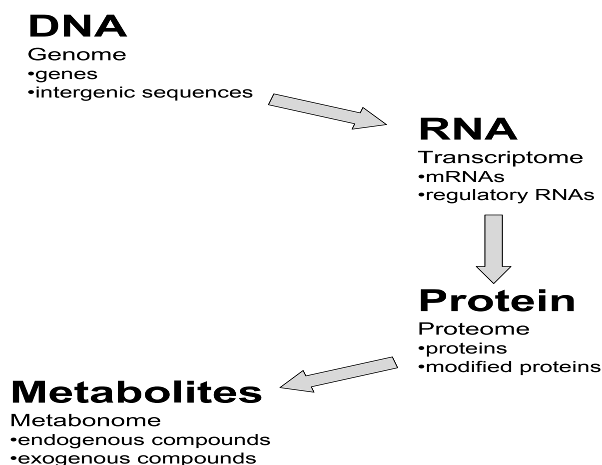

FIGURE 2-1 Hierarchical relationships of DNA, RNA, proteins, and metabolites.

criptomics. In contrast to microarrays for transcriptomics, technologies for proteomics and metabolomics are limited by the huge range of analyte concentrations involved (at least six orders of magnitude), because all existing instrumentation favors the detection of more abundant species over those that are less abundant. This fundamental problem limits essentially all proteomic and metabolomic analyses to subsets of the complete collections of proteins and metabolites, respectively. Despite this limitation, proteomic and metabolomic approaches have fundamentally advanced the understanding of the mechanisms of toxicity and adaptation to stress and injury.

The core technologies for toxicogenomics are evolving rapidly, and this is making toxicogenomic approaches increasingly powerful and cost-effective. Several key aspects of technology development will drive toxicogenomics in the next decade:

1. New sequencing technologies offer the prospect of cost-effective individual whole-genome sequencing and comprehensive genotype analysis.
2. Array-based whole-genome scanning for variations in individual genes, known as single nucleotide polymorphisms (SNPs), will dramatically increase throughput for genotyping in population studies.
3. Advances in NMR and MS instrumentation will enable high-sensitivity analyses of complex collections of metabolites and proteins and quantitative metabolomics and proteomics.
4. New bioinformatic tools, database resources, and statistical methods will integrate data across technology platforms and link phenotypes and toxicogenomic data.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

The interplay of technology development and application continues to drive the evolution of toxicogenomic technologies in parallel with advances in their applications to other fields of biology and medicine. This dynamic will extend the longstanding paradigm of toxicology as a means to understand both fundamental physiology and environmentally induced disease.

The basic types of technologies relevant to this report are described in this chapter. However, it is not feasible for the report to describe comprehensively the various tools and developments in this rapidly evolving field.

GENOMIC TECHNOLOGIES

The dramatic evolution of rapid DNA sequencing technologies during the 1980s and 1990s culminated in sequences of the human genome and the genomes of dozens of other organisms, including those used as animal models for chemical toxicity. Complete genome sequences provide catalogs of genes, their locations, and their chromosomal context. They serve as general reference maps but do not yet reflect individual variations, including SNPs and groups of SNP alleles (haplotypes), that account for the individual genetic variation underlying different responses to toxic chemical and environmental factors. This section describes technologies for sequencing, for analyzing genotype variation, and for analyzing epigenetic modifications. As technology continues to advance, so will the analyses of genome variations and their roles in humans' different responses to chemicals and other environmental factors.

Genome Sequencing Technologies

High-throughput gene sequencing began when Sanger and colleagues introduced dideoxynucleotide sequencing in 1977 (Sanger et al. 1977). Technical innovations over the next 18 years led to the development of current generation automated instruments, which use fluorescent tagging and capillary electrophoresis to sequence up to 1.6 million base pairs (bp) per day (Chan 2005). These instruments have become the principal engines of genome sequencing projects. Despite the impressive advances in technology, the cost of genome sequencing remains a barrier to the application of whole-genome sequencing to study individual variation in toxic responses and susceptibility. Current costs for whole-genome sequencing in a modern genome sequencing center using Sanger sequencing technology were estimated at approximately $0.004 per bp, which translates to approximately $12M for an entire human genome (Chan 2005).

Fundamental technology limitations of automated Sanger sequencing have spawned the development of new technologies that promise significant cost and throughput improvements. A cycle extension approach using a highly parallel picoliter reactor system and a pyrosequencing protocol sequenced 25 million bp at 99% accuracy in a 4-h analysis (Margulies et al. 2005).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

This represents a 100-fold increase in throughput over the best available Sanger sequencing instrumentation and is now marketed commercially (454 Life Sciences 2006). Church and colleagues (Shendure et al. 2005) recently reported the use of a microscopy-based sequencing approach that employs "off-the-shelf" reagents and equipment and is approximately one-ninth as costly as high-speed Sanger sequencing (Shendure et al. 2005). Nanotechnology approaches may yield even more efficient DNA sequencing instrumentation, but these approaches have not yet been implemented for high-throughput sequencing in a research laboratory setting (Chan et al. 2004; Chan 2005).

The dominant genome sequencing strategy to emerge from the human genome project is the whole-genome "shotgun" sequencing approach (Venter et al. 1996), which entails breaking up chromosomal DNA into fragments of 500-1,000 DNA bases, which then are subjected to automated, repetitive sequencing and subsequent data analyses to reassemble the fragment sequences into their original chromosomal context. With this technique, most DNA bases on a chromosome are sequenced three to seven times, resulting in cost- and time-efficient generation of high-fidelity sequence information.

Genotype Analysis

Analysis of genetic variation between individuals involves first the discovery of SNPs and then the analysis of these variations in populations. SNPs occur on approximately every 2 kilobases in the human genome and have been commonly detected by automated Sanger sequencing as discussed above. The Human DNA Polymorphism Discovery Program of the National Institute of Environmental Health Sciences Environmental Genome Project (EGP) is one example of the application of automated DNA sequencing technologies to identify SNPs in human genes (Livingston et al. 2004). The EGP selected 293 "candidate genes" for sequencing based on their known or anticipated roles in the metabolism of xenobiotics. The major limitation of this candidate gene approach is that it enables discovery of polymorphisms only in those genes targeted for analysis.

This analysis at the population level requires high-throughput, cost-effective tools to analyze specific SNPs. The analysis generally involves PCR-based amplification of a target gene and allele-specific detection. The dominant approaches in current use are homogeneous systems using TaqMan (Livak 1999) and molecular beacons (Marras et al. 1999), which offer high throughput in a multiwell format. However, these systems are not multiplexed (multiple SNPs cannot be analyzed simultaneously in the same sample), and this ultimately limits throughput. Other approaches may provide the ability to analyze multiple SNPs simultaneously. For example, matrix-assisted laser desorption ionization MS (MALDI-MS) has been used to simultaneously analyze multiple products of primer-extension reactions (Gut 2004). A more highly multiplexed

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

approach is the combination of tag-based array technology with primer extension for genotyping (Phillips et al. 2003; Zhang et al. 2004).

The ultimate approach for SNP discovery is individual whole-genome sequencing. Although not yet feasible, rapidly advancing technology makes this approach the logical objective for SNP discovery in the future. In the meantime, rapid advances in microarray-based SNP analysis technology have redefined the scope of SNP discovery, mapping, and genotyping. New microarray-based genotyping technology enables "whole-genome association" analyses of SNPs between individuals or between strains of laboratory animals (Syvanen 2005). In contrast to the candidate gene approach described above, whole-genome association identifies hundreds to thousands of SNPs in multiple genes. Arrays used for these analyses can represent a million or more polymorphic sequences mapped across a genome (Gunderson et al. 2005; Hinds et al. 2005; Klein et al. 2005). This approach makes it possible to identify SNPs associated with disease and susceptibility to toxic insult. As more SNPs are identified and incorporated into microarrays, the approach samples a greater fraction of the genome and becomes increasingly powerful. The strength of this technology is that it puts a massive amount of easily measurable genetic variation in the hands of researchers in a cost-effective manner ($500-$1,000 per chip). Criteria for including the selected SNPs on the arrays are a critical consideration, as these affect inferences that can be drawn from the use of these platforms.

Epigenetics

Analysis of genes and genomes is not strictly confined to sequences, as modifications of DNA also influence the expression of genes. *Epigenetics* refers to the study of reversible heritable changes in gene function that occur without a change in the sequence of nuclear DNA. These changes are major regulators of gene expression. Differences in gene expression due to epigenetic factors are increasingly recognized as an important basis for individual variation in susceptibility and disease (Scarano et al 2005), as discussed in Chapter 6.

Epigenetic phenomena include DNA methylation, imprinting, and histone modifications. Of all the different types of epigenetic modifications, DNA methylation is the most easily measured and amenable to the efficient analysis characteristic of toxicogenomic technologies. DNA methylation refers to addition of a methyl group to the 5-carbon of cytosine in an area of DNA where there are many cytosines and guanines (a CpG island) by the enzyme DNA methyltransferase. Many different methods for analyzing DNA methylation have been developed and fall into two main types—global and gene-specific. *Global* methylation analysis methods measure the overall level of methyl cytosines in the genome using chromatographic methods (Ramsahoye 2002) or methyl accepting capacity assays. *Gene-specific* methylation analysis methods originally used methylation-sensitive restriction enzymes to digest DNA before it was analyzed by Southern blot analysis or PCR amplification. Sites that were methy-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

lated were identified by their resistance to the enzymes. Recently, methylation-sensitive primers or bisulfite conversion of unmethylated cytosine to other bases have been used in methods such as methylation specific PCR and bisulfite genomic sequencing. These methods give a precise map of the pattern of DNA methylation in a particular genomic region or gene and are fast and cost effective (e.g., Yang et al. 2006).

To identify unknown methylation hot-spots within a larger genomic context, techniques such as Restriction Landmark Genomic Scanning for Methylation (Ando and Hayashizaki 2006) and CpG island microarrays (e.g., Y. Wang et al. 2005a; Bibikova et al. 2006; Hayashi et al. 2007) have also been developed. Restriction landmark genomic scanning (RLGS) is a method to detect large numbers of methylated cytosines sites in a single experiment using direct end-labeling of the genomic DNA digested with a restriction enzyme and separated by high-resolution two-dimensional electrophoresis.

Several array-based methods have also been developed recently. Bibikova et al. (2006) developed a rapid method for analyzing the methylation status of hundreds of preselected genes simultaneously through an adaptation of a genotyping assay. For example, the methylation state of 1536 specific CpG sites in 371 genes was measured in a single reaction by multiplexed genotyping of bisulfite-treated genomic DNA. The efficient nature of this quantitative assay could be useful for DNA methylation studies in large epidemiologic samples. Hayashi et al. (2007) describe a method for analysis of DNA methylation using oligonucleotide microarrays that involves separating methylated DNA immunoprecipitated with anti-methylcytosine antibodies.

Most of these gene-specific methods work consistently at various genomic locations (that is, they have minimal bias by genomic location) and will be useful for high-resolution analysis of the epigenetic modifications to the genome in toxicogenomic studies.

TRANSCRIPTOMIC TECHNOLOGIES

Transcriptomics describes the global measurement of mRNA transcripts in a biologic system. This collection of mRNA transcripts represents the transcription of all genes at a point in time. Technologies that allow the simultaneous analysis of thousands of transcripts have made it possible to analyze transcriptomes.

Technologic Approaches

Technologies for assaying gene, protein, and metabolic expression profiles are not new inventions. Measurements of gene expression have evolved from the single measures of steady-state mRNA using Northern blot analysis to the more global analysis of thousands of genes using DNA microarrays and serial analysis of gene expression (SAGE), the two dominant technologies (Figure 2-2). The

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

advantage of global approaches is the ability of a single investigation to query the behavior of hundreds, thousands, or tens of thousands of biologic molecules in a single assay. For example, in profiling gene expression, one might use technologies such as Northern blot analysis to look at expression of a single gene. Quantitative real-time reverse transcriptase PCR (qRT-PCR), often used with subtractive cloning or differential display, can easily be used to study the expression of 10 or more genes.

Techniques such as SAGE allow the entire collection of transcripts to be catalogued without assumptions about what is actually expressed (unlike microarrays, where one needs to select probes from a catalogue of genes). SAGE is a technology based on sequencing strings of short expressed sequence tags representing both the identity and the frequency of occurrence of specific sequences within the transcriptome. SAGE is costly and relatively low throughput, because each sample to be analyzed requires a SAGE Tag library to be constructed and sequenced. Massively parallel signature sequencing speeds up the SAGE process with a bead-based approach that simultaneously sequences multiple tags, but it is costly.

DNA microarray technology can be used to generate large amounts of data at moderate cost but is limited to surveys of genes that are included in the microarray. In this technology (see Box 2-1 and Figure 2-3), a solid matrix surface supports thousands of different, surface-bound DNAs, which are hybridized against a pool of RNA to measure gene expression. A systematic comparison indicates that gene expression measured by oligonucleotide microarrays correlates well with SAGE in transcriptional profiling, particularly for genes expressed at high levels (Kim 2003).

Gene Expression Analysis Methods

Method	Throughput	Comments
•Northern blot	1 gene	•Standard procedure; low throughput
•Substractive cloning	Increasing Data Density ↓	•Not always comprehensive
•Differential display		•Follow up full-length cloning required; potential to identify rare mRNAs
•EST/SAGE		•"Expensive" and requires a dedicated sequencing facility
•Gridded filters		•Cannot multiplex probes derived from two different tissue samples
•High density arrays	10^4 gene	•Identification of differentially expressed genes dependent on arrayed elements

FIGURE 2-2 Overview of commonly used methods and technologies for gene expression analysis.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 2-1 Experimental Details of Transcriptome Profiling with Microarrays

TmRNA extracted from cell or tissue samples is prepared for microarray analysis by PCR-based amplification (Hardiman 2004). A fluorescent dye (or biotin for Affymetrix microarrays) is incorporated into the amplified RNA sequences. Two-color arrays involve fluorescently labeling paired samples (control versus experimental) with different dyes (see Figure 2-3). The amplified, labeled sequences, termed “targets,” are then hybridized to the microarrays.

After hybridization and washing, the arrays are imaged with a confocal laser scanner and the relative fluorescence intensity (or streptavidin-conjugated phycoerythrin) for each gene-specific probe represents the expression level for that gene. The actual value reported depends on the microarray technology platform used and the experimental design. For Affymetrix GeneChips, in which each sample is hybridized to an individual array, expression for each gene is measured as an “average difference” that represents an estimated expression level, less nonspecific background. For two-color arrays, assays typically compare paired samples and report expression as the logarithm of the ratio of the experimental sample to the control sample. Regardless of the approach or technology, the fundamental data used in all subsequent analyses are the expression measures for each gene in each experiment. These expression data are typically represented as an “expression matrix” in which each row represents a particular gene and each column represents a specific biologic sample (Figure 2-3). In this representation, each row is a “gene expression vector,” where the individual entries are its expression levels in the samples assayed and each column is a “sample expression vector” that records the expression of all genes in that sample.

The data are normalized to compensate for differences in labeling, hybridization, and detection efficiencies. Approaches to data normalization depend on the platform and the assumptions made about biases in the data (Brazma et al. 2001; Schadt et al. 2001; I.V. Yang et al. 2002; Y.H. Yang et al. 2002; Sidransky et al. 2003). Filtering transformations are often applied to the data by using statistical approaches that, for example, eliminate genes that have minimal variance across the collection of samples or those that fail to provide data in most of the experiments. These filtering transformations reduce dataset complexity by eliminating genes unlikely to contribute to the experimental goal. The choice of normalization and filtering transformations can have a profound effect on the results (Hoffmann et al. 2002). Normalization adjusts the fluorescence intensities on each array and therefore can change the relative difference observed among samples. Normalization is generally necessary to compensate for systematic errors introduced during measurement, but overnormalizing can distort the data. Similarly, different methods of data filtering can produce very different results. All statistical tests that are applied rely on assumptions about the nature of the variance in the measurements. Different statistical tests applied to the very same dataset can often produce different (but generally overlapping) sets of significant genes. Dealing with these “high-dimensional” datasets in which there are often more measurements (genes) than samples is an area of active research and debate.

(Continued on next page)

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 2-1 Continued

Standardization of protocols for transcriptional profiling experiments has contributed to validation and verification strategies that ensure the quality of data. In large measure, progress was facilitated by creation of Minimum Information About a Microarray Experiment (MIAME) guidelines by the Microarray Gene Expression Data Society. MIAME was designed as a set of recommendations to address issues related to data quality and exchange (Brazma et al. 2001; Ball et al. 2002a,b, 2004a,b). The scientific community has endorsed the guidelines (MIAME 2005), and most scientific journals now require adherence to the MIAME recommendations for publishing toxicogenomic studies. MIAME guidelines encompass parameters such as degree of signal linearity, hybridization specificity, normalization strategy, and use of exogenous and internal controls.

In principle, it should be possible to mine datasets generated by multiple laboratories with different microarray platforms. There is tremendous value in making gene expression datasets publicly available and being able to mine the datasets. Besides serving as a source of independent data that can be used as a means of validating results, larger and more diverse sample populations can provide more robust datasets for "meta-analysis" designed to find patterns of gene expression that can be associated with specific biologic states and responses (Malek et al. 2002; Stuart et al. 2003). However, a number of published studies have failed to find concordance between microarray platforms designed to assay expression patterns, in part because of observed disparities between results obtained by different groups analyzing similar samples (calling the validity of microarray assays into question) (Kuo et al. 2002; Maitra et al. 2003; Rogojina et al. 2003; Mah et al. 2004; Park et al. 2004; Shippy et al. 2004; Ulrich et al. 2004; Yauk et al. 2004). In many instances, it appears that this failure to find concordance is a failure not of the platform or the biologic system but of the metrics used to evaluate concordance. For example, other meta-analyses focused on lists of significant genes, neglecting the fact that in many instances these lists of genes are derived not only from different platforms but also from vastly different approaches to data analysis (Tan et al. 2003; Jarvinen et al. 2004; Mah et al. 2004). This effect can be seen even in looking at a single dataset generated on a single platform. When results from the same array platforms are compared, the results generally show good concordance among different laboratories (Kane et al. 2000; Hughes et al. 2001; Yuen et al. 2002; Barczak et al. 2003; Carter et al. 2003; H.T. Wang et al. 2003). The data analysis effect can be seen even in looking at a single dataset generated on a single platform. When results from the same array platforms are compared, the results generally show good concordance among different laboratories (Kane et al. 2000; Hughes et al. 2001; Yuen et al. 2002; Barczak et al. 2003; Carter et al. 2003; H.T. Wang et al. 2003).

A series of papers that appeared in the May 2005 issue of *Nature Methods* systematically dealt with the problem of platform and laboratory comparison (Bammler et al. 2005; Irizarry et al. 2005; Larkin et al. 2005). Larkin et al. (2005) analyzed gene expression in a mouse model of hypertension and compared results obtained using spotted cDNA arrays and Affymetrix GeneChips. For the genes that could be compared, 88% showed expression patterns that appeared to be driven by

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

the underlying biology rather than the platform, and these genes also correlated well with qRT-PCR. Surprisingly, the 12% of genes that showed platform-specific effects also correlated poorly with qRT-PCR. Comparison of these platform discrepant genes with the platform concordant genes showed that the discrepant genes were much more likely to map to poorly annotated regions of the genome and consequently were more likely to represent different forms of MRNA (different splice forms). Irizarry and colleagues (2005) compared gene expression using a pair of defined RNA samples and looked at a variety of platforms with data generated by a number of laboratories using a variety of microarray platforms. This study showed that one can estimate the "lab effect," which encompasses differences in sites, platforms, and protocols and, in doing so, arrive at estimates of gene expression that can be compared among laboratories. Finally, the Toxicogenomics Research Consortium (Bammler et al. 2005) reported that standardization of laboratory and data analysis protocols resulted in a dramatic increase in concordance among the results different laboratories obtained. Independently, these three groups arrived at the general conclusion that, if experiments are done and analyzed carefully and systematically, the results are quite reproducible and provide insight into the underlying biology driving the systems being analyzed.

Although toxicogenomic studies typically rely on technologies that generate large amounts of data, results are often confirmed and replicated with lower throughput assays. For example, differential gene expression detected with more global approaches is often verified by qRT-PCR analysis. The utility of these lower throughput approaches goes beyond validation. A subset of genes analyzed by qRT-PCR may exhibit sensitivity and specificity comparable to global transcriptomic analyses with microarrays. Relatively small sets of marker genes that represent more complex gene expression patterns may be of considerable value in toxicogenomics.

DNA Microarray Technology

Microarray technology (Figure 2-3) fundamentally advanced biology by enabling the simultaneous analysis of all transcripts in a system. This capability for simultaneous, global analysis is emblematic of the new biology in the genomic era and has become the standard against which other global analysis technologies are judged. DNA microarrays contain collections of oligonucleotide sequences located in precise locations in a high-density format. Two complementary DNA (cDNA) microarray formats have come to dominate the field. *Spotted microarrays* are prepared from synthesized cDNAs or oligonucleotide probes that are printed on a treated glass slide surface in a high-density format. These spotted arrays were the first widely used DNA microarrays (Schena et al. 1995, 1996) and were originally printed in individual investigators' laboratories

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

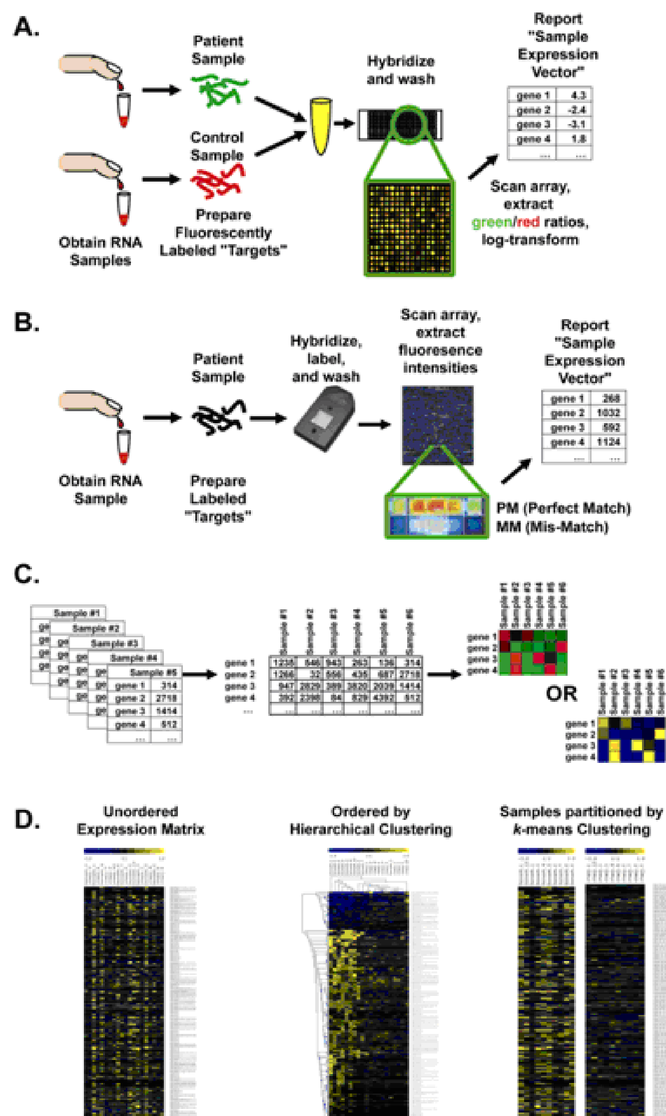

FIGURE 2-3 An overview of DNA microarray analysis. (A) In two-color analysis approaches, RNA samples from patient and control samples are individually labeled with distinguishable fluorescent dyes and cohybridized to a single DNA microarray consisting of individual gene-specific probes. Relative gene expression levels in the two samples are estimated by measuring the fluorescence intensities for each arrayed probe; a sample expression vector summarizing the expression level of each gene in the patient sample (relative to the control) is reported. (B) Single-color analysis, such as that using the Affymetrix GeneChip, hybridize labeled RNA from each biologic sample to a single array

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

in which a series of perfectly matched (PM) gene-specific probes are arrayed. Gene expression levels are estimated by measuring hybridization intensities for each probe, and background is measured by using a corresponding set of mismatched (MM) probes. For each array technology, gene expression levels are reported for each sample as a "sample expression vector" summarizing the difference between signal and background for each gene. (C) The data from each gene in each sample are collected and these "sample expression vectors" are assembled into a single "expression matrix." Each column in the expression matrix represents an individual sample and its measured expression levels for each gene (the sample expression vector); each row represents a gene and its expression levels across all samples (a "gene expression vector"). The expression matrix is often visualized by presenting a colored matrix (typically red/green, although other combinations such as blue/yellow are now common). (D) An unordered dataset, subjected to average linkage hierarchical clustering or k-mean clustering, reveals underlying patterns that can help identify classes in the dataset.

---

from collections of clones. Complications in characterizing, managing, and standardizing these collections led to substantial variability in performance. Commercially produced oligonucleotide microarrays, in which oligonucleotides are synthesized in situ using inkjet printing, have largely replaced cDNA microarrays (Hughes et al. 2001). Whole-genome microarrays for human and mouse genomes contain 40,000-45,000 features corresponding to unique genes and transcripts. The probes range from 20 to 60 bp and individual microarrays typically contain between 5,000 and 50,000 features. The longer probes provide improved sensitivity and tolerance of polymorphic sequence mismatches. Several commercial vendors provide spotted arrays or variants of this technology and development in this area continues (Hardiman 2004).

The alternative technology uses photolithographic synthesis of oligonucleotide probes on a quartz surface and was developed by Affymetrix (Fodor et al. 1993; Pease et al. 1994; Lipshutz et al. 1999). These GeneChip arrays are characterized by very high probe densities (up to 1.3 million probes per chip) and typically consist of up to 25-mer probes (probes with 25-base residues). Each "gene" may be represented by as many as 20 overlapping probe sequences and paired mismatch probes, which contain sequence substitutions that enable quantitative evaluation of nonspecific hybridization. Elaboration of this mismatch strategy also allows analysis of SNPs by microarray analysis (see SNP discussion in section above). Considerable research into probe design has contributed to the improvement of microarray performance and has facilitated the standardization of transcriptome analysis.

Other array formats have been developed. Nylon membranes and plastic microarrays have been used with varying degrees of success (Qian et al. 2005). Nylon membranes produce low- to medium-density cDNA microarrays, whereas plastic retains the advantages of glass for producing high-density microarrays that are somewhat cheaper than glass slide arrays. The probes for nylon arrays are typically labeled with radioactive phosphorus isotopes (^{32}P or ^{33}P) to afford

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

increases in sensitivity, but this approach is not favored because of problems associated with the use of radioactivity and efficiency of analysis.

Affymetrix and other major commercial vendors (Agilent, GE Healthcare [formerly Amersham], and Applied Biosystems) currently offer several different microarrays corresponding to essentially all known genes and transcripts for human as well as similar microarray products for model organisms used in toxicity studies. In addition, Affymetrix also offers whole-genome microarrays for application to SNP mapping and detection (see above).

PROTEOMIC TECHNOLOGIES

Proteomics is the study of proteomes, which are collections of proteins in living systems. Because proteins carry out most functions encoded by genes, analysis of the protein complement of the genome provides insights into biology that cannot be drawn from studies of genes and genomes. MS, gene and protein sequence databases, protein and peptide separation techniques, and novel bioinformatics tools are integrated to provide the technology platform for proteomics (Yates 2000; Smith 2002; Aebersold and Mann 2003). In contrast to "the genome," there is no single, static proteome in any organism; instead, there are dynamic collections of proteins in different cells and tissues that display moment-to-moment variations in response to diet, stress, disease processes, and chemical exposures. There is no technology analogous to PCR amplification of nucleic acids that can amplify proteins, so they must be analyzed at their native abundances, which span more than six orders of magnitude. Each protein may be present in multiple modified forms; indeed, variations in modification status may be more critical to function than absolute levels of the protein per se (Mann and Jensen 2003). A related problem is the formation of protein adducts by reactive chemical intermediates generated from toxic chemicals and endogenous oxidative stress (Liebler et al. 2003). Protein damage by reactive chemical intermediates may also perturb endogenous regulatory protein modifications. All these characteristics add to the challenge of proteome analysis. In contrast to the microarray technologies applied to gene expression, most analytical proteomic methods represent elaborate serial analyses rather than truly parallel technologies.

Gel-Based Proteomics

Two major approaches used are gel-based proteomics and "shotgun" proteomics (see Figure 2-4). In the gel-based approach, proteins are resolved by electrophoresis or another separation method and protein features of interest are selected for analysis. This approach is best represented by the use of two-dimensional sodium dodecylsulfate polyacrylamide gel electrophoresis (2D-SDS-PAGE) to separate protein mixtures, followed by selection of spots, and identification of the proteins by digestion to peptides, MS analysis, and database

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

searching. Gel-based analyses generate an observable "map" of the proteome analyzed, although liquid separations and software can be coupled to achieve analogous results. Reproducibility of 2D gel separations has dramatically improved with the introduction of commercially available immobilized pH gradient strips and precast gel systems (Righetti and Bossi 1997). Comparative 2D-SDS-PAGE with differential fluorescent labeling (for example, differential gel electrophoresis, DIGE) offers powerful quantitative comparisons of proteomes (Tonge et al. 2001; Von Eggeling et al. 2001). Moreover, modified protein forms are often resolved from unmodified forms, which enable separate characterization and quantitative analysis of each. Although 2D gels have been applied most commonly to global analyses of complex proteomes, they have great potential for comparative analyses of smaller subproteomes (for example, multiprotein complexes). Problems with gel-based analyses stem from the poor separation characteristics of proteins with extreme physical characteristics, such as hydrophobic membrane proteins. A major problem is the limited dynamic range for protein detection by staining (200- to 500-fold), whereas protein abundances vary more than a million fold (Gygi et al. 2000). This means that abundant proteins tend to preclude the detection of less abundant proteins in complex mixtures. This problem is not unique to gel-based approaches.

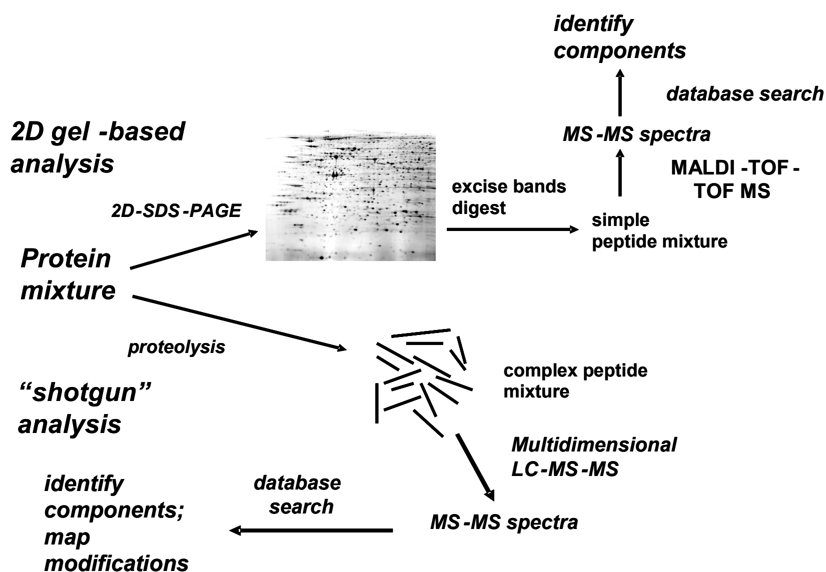

FIGURE 2-4 Schematic representation of 2D gel-based proteome analysis (upper) and shotgun proteome analysis (lower). LC, liquid chromatography; TOF, time of flight; MS-MS, tandem mass spectrometry.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Shotgun Proteomics

Shotgun proteomic analysis is somewhat analogous to the genome sequencing strategy of the same name. Shotgun analyses begin with direct digestion of protein mixtures to complex mixtures of peptides, which then are analyzed by liquid-chromatography-coupled mass spectrometry (LC-MS) (Yates 1998). The resulting collection of peptide tandem mass spectrometry (MS-MS) spectra is searched against databases to identify corresponding peptide sequences and then the collection of sequences is reassembled using computer software to provide an inventory of the proteins in the original sample mixture. A key advantage of shotgun proteomics is its unsurpassed performance in the analysis of complex peptide mixtures (Wolters et al. 2001; Washburn et al. 2002). Peptide MS-MS spectra are acquired by automated LC-MS-MS analyses in which ions corresponding to intact peptides are automatically selected for fragmentation to produce MS-MS spectra that encode peptide sequences (Stahl et al. 1996). This approach enables automated analyses of complex mixtures without user intervention. However, selection of peptide ions for MS-MS fragmentation is based on the intensity of the peptide ion signals, which favors acquisition of MS-MS spectra from the most abundant peptides in a mixture. Thus, detection of low-abundance peptides in complex mixtures is somewhat random. Application of multidimensional chromatographic separations (for example, ion exchange and then reverse-phase high-performance liquid chromatography) “spreads out” the peptide mixture and greatly increases the number of peptides for which MS-MS spectra are acquired (Link et al. 1999; Washburn et al. 2001; Wolters et al. 2001). This increases the detection of less abundant proteins and modified protein forms (MacCoss et al. 2002) (see below). New hybrid linear ion trap-tandem MS instruments offer more rapid acquisition of MS-MS spectra and more accurate identification of peptide ion mass-to-charge ratio values, which provides more identifications with greater reliability. Nevertheless, a continuing challenge of shotgun proteome analyses is the identification of less abundant proteins and modified protein forms.

Quantitative Proteomics

The application of quantitative analyses has become a critical element of proteome analyses. Quantitative methods have been developed for application to both gel-based and shotgun proteomic analyses. The most effective quantitative approach for gel-based analyses is DIGE (see above), which involves using amine- or thiol-reactive fluorescent dyes that tag protein samples with different fluorophores for analysis on the same gel. This approach eliminates gel-to-gel variations inherent in comparing spots from individual samples run on different gels. The use of a separate dye and mixed internal standards allows gel-to-gel comparisons of DIGE analyses for larger studies and enables reliable statistical comparisons (Alban et al. 2003). Quantitative shotgun proteome analyses have

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

been done with stable isotope tags, which are used to derivatize functional groups on proteins (Julka and Regnier 2004). Stable isotope tagging is usually used in paired experimental designs, in which the relative amounts of a protein or protein form are measured rather than the absolute amount in a sample. The first of these to be introduced were the thiol-reactive isotope-coded affinity tag reagents (Gygi et al. 1999), which have been further developed to incorporate solid-phase capture and labeling (Zhou et al. 2002). These reagents are available in "heavy" (for example, ^{2}H- or ^{13}C-labeled) and "light" (for example, ^{1}H- or ^{12}C-labeled) forms. Analysis of paired samples labeled with the light and heavy tags allows relative quantitation by comparing the signals for the corresponding light and heavy ions. Other tag chemistries that target peptide N and C termini have been developed and have been widely applied (Julka and Regnier 2004). An alternative approach to tagging proteins and peptides is to incorporate stable isotope labels through metabolic labeling of proteins during synthesis in cell culture (Ong et al. 2002). Quantitative proteomic approaches are applicable not only to comparing amounts of proteins in samples but also to kinetic studies of protein modifications and abundance changes as well as to identification of protein components of multiprotein complexes as a function of specific experimental variables (Ranish et al. 2003).

Major limitations of the isotope-tagging approaches described above include the requirement for chemical induction of changes in the samples (derivatization) or metabolic incorporation of isotope label and the need to perform quantitative analyses by pairwise comparison. Recent work has demonstrated that quantitative data from LC-MS full-scan analyses of intact peptides (W. Wang et al. 2003) and data from MS-MS spectra acquired from peptides are proportional to the peptide concentration in mixtures (Gao et al. 2003; Liu et al. 2004). This suggests that survey-level quantitative comparisons between any samples analyzed under similar conditions may be possible.

Finally, the use of stable isotope dilution LC-MS-MS analysis provides a method for absolute quantification of individual proteins in complex samples (Gerber et al. 2003). Use of stable-isotope-labeled standard peptides that uniquely correspond to proteins or protein forms of interest are spiked into proteolytic digests from complex mixtures, and the levels of the target protein are measured relative to the labeled standard. This approach holds great potential for targeted quantitative analysis of candidate biomarkers in biologic fluids (Anderson et al. 2004).

Bioinformatic Tools for Proteomics

A hierarchy of proteomic data is rooted in MS and MS-MS spectra (Figure 2-5) and includes identification and quantitation of proteins and peptides and their modified forms, including comparisons across multiple experiments, analyses, and datasets. A key element of MS-based proteomic platforms is the identification of peptide and protein sequences from MS data. This task is accom-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

plished with a variety of algorithms and software ("bioinformatics" tools) that search protein and nucleotide sequence databases (Fenyo 2000; Sadygov et al. 2004; MacCoss 2005). Measured peptide masses from MALDI-MS spectra of tryptic peptide digests can be searched against databases to identify the corresponding proteins (Perkins et al. 1999). This peptide mass fingerprinting approach works best with relatively pure protein samples. The most widely used and most effective approach is to search uninterpreted MS-MS spectra against database sequences with algorithms and software, such as Sequest, Mascot, and X!Tandem (Eng et al. 1994; Perkins et al. 1999; Craig and Beavis 2004). These algorithms match all spectra to some sequence and provide scores or probability assessments of the quality of the matches. Nevertheless, the balance between sensitivity and specificity in these analyses amounts to a trade-off between missed identifications (low sensitivity) and false-positive identifications (poor specificity) (Nesvizhskii and Aebersold 2004). A second tier of bioinformatic tools evaluates outputs from database search algorithms and estimates probabilities of correct protein identifications (Keller et al. 2002; Nesvizhskii et al. 2003). Other software applications have been developed to enable the identification of modified peptide forms from MS-MS data, even when the chemical nature and amino acid specificity of the modification cannot be predicted (Hansen et al. 2001, 2005; Liebler et al. 2002, 2003).

A key issue in proteomics is the standardization of data analysis methods and data representation and reporting formats. A fundamental problem is the variety of MS instruments, data analysis algorithms, and software used in proteomics. These generate a variety of different data types that describe proteins

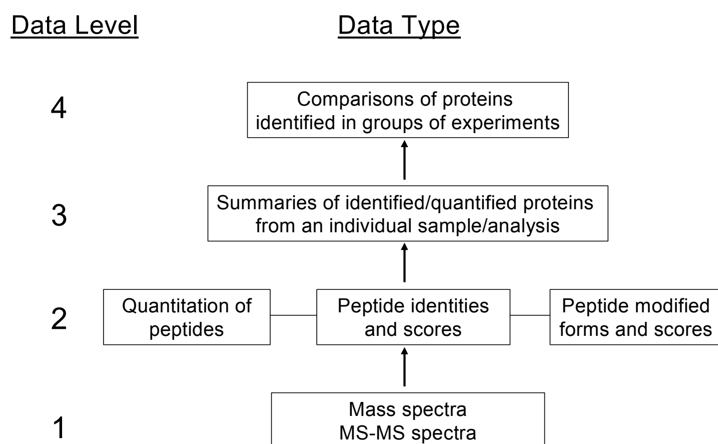

FIGURE 2-5 A hierarchy of proteomic data is rooted in MS and MS-MS spectra (level 1) and includes outputs of database search analyses and related data reduction (level 2), integrated information about single proteins (level 3), and information about groups of proteins or proteomes across multiple experiments (level 4).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

and peptides and their modifications. Proposals for common representations of MS and proteomic data have been published recently (Taylor et al. 2003; Craig et al. 2004; Pedrioli et al. 2004). In addition, draft criteria for data reporting standards for publication in proteomic journals are under consideration (e.g., Bradshaw 2005a). Another useful development is the emerging collection of databases of matched peptide and protein sequences and corresponding spectral data that define them (Craig et al. 2004; Desiere et al. 2005, 2006).

Another important, but unresolved, issue concerns the differences in protein and peptide identifications attributable to the use of different database search algorithms. Because different database search software packages are sold with different MS instruments (for example, Sequest is licensed with Thermo ion trap MS instruments), differences in performance of the algorithms are difficult to separate from differences in characteristics of the instruments. Another issue in comparing the performance of different database searching software is wide variation in identifications due to variation in criteria used to filter the search results (Peng et al. 2003; Elias et al. 2005). This situation will be improved somewhat by adopting standards for reporting false-positive identification rates (Bradshaw 2005b).

Although the efforts described above represent useful steps in information sharing and management, the diversity of instrumentation, analytical approaches, and available data analysis tools will make standardization of informatics an ongoing challenge.

Proteome Profiling

Another type of proteome analysis that has attracted widespread interest is proteome profiling, in which MALDI time-of-flight (MALDI-TOF) MS is used to acquire a spectral profile of a tissue or biofluid sample (for example, serum) (Chaurand et al. 1999; Petricoin et al. 2002a,b; Villanueva et al. 2004). The signals in these spectra represent intact proteins or protein fragments and collectively reflect the biologic state of the system but, in profiling (compared with approaches described above), the overall pattern rather than identification of specific proteins or protein fragments is the focus. Analyses with high-performance MALDI-TOF instruments can generate spectral profiles containing hundreds to thousands of signals. These typically correspond to the most abundant, lower molecular weight (<25 kilodaltons) components of proteomes. Machine learning approaches have been used to identify sets of spectral features that can classify samples based on spectral patterns (Baggerly et al. 2004, 2005; Conrads et al. 2004). This approach has attracted considerable interest as a potential means of biomarker discovery for early detection of diseases, particularly cancers, as well as drug toxicity. Despite intense interest, proteome profiling studies have created considerable controversy due to problems with lab-to-lab reproducibility of the marker sets identified, a lack of identification of the proteins corresponding to the marker signals, and artifacts in data generation and

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

analysis (Diamandis 2004; Baggerly et al. 2005). In addition, studies that have identified some of the marker species have shown that they typically are proteolysis products of abundant blood proteins (Marshall et al. 2003), which raises questions about the biologic relationship of the markers to the disease processes under study. The general utility of biofluid proteome profiling for biomarker discovery remains an attractive, if unproven, approach. Nevertheless, methods of instrumental and data analysis are rapidly evolving in this field, and the applicability of this approach should be better substantiated within the next 2-3 years.

New MS Instrumentation and Related Technology for Proteomics

Despite impressive advances over the past 15 years, MS instrumentation for proteomics is limited in the numbers of peptides or proteins that can be identified and in the quality of the data generated. New hybrid tandem MS instruments that couple rapid-scanning linear ion tray analyzers with Fourier transform ion cyclotron resonance (FTICR), high-resolution ion trap, and TOF mass analyzers offer both high mass accuracy measurements of peptide ions and rapid scanning acquisition of MS-MS spectra (Syka et al. 2004a; Hu et al. 2005). This improves the fidelity of identification and the mapping of modifications (Wu et al. 2005). New methods for generating peptide sequence data, such as electron transfer dissociation (Syka et al. 2004b), can improve the mapping of posttranslational modifications and chemical adducts. An important emerging application of FTICR instrumentation is the tandem MS analysis of intact proteins, which is referred to as "top-down" MS analysis (Ge et al. 2002; Kelleher 2004). This approach can generate near-comprehensive sequence analysis of individual protein molecular forms, thus enabling sequence-specific annotation of individual modification variants (Pesavento et al. 2004; Coon et al. 2005). A limitation of the approach is the requirement for relatively purified proteins and larger amounts of samples than are used in shotgun analyses. However, rapid technology development will make top-down methods increasingly useful for targeted analyses of individual proteins and their modified forms.

Non-MS-Based Proteomic Approaches

Non-MS-based technologies have been applied to proteome analyses, but they have not proven to be as robust and versatile as MS-based methods. Microarray technology approaches include antibody microarrays, in which immobilized antibodies recognize proteins in complex mixtures (de Wildt et al. 2000; Miller et al. 2003; Olle et al. 2005). Although straightforward in principle, this approach has not proven robust and reliable for several reasons. Monospecific antibodies with high affinity for their targets are difficult to obtain and they often lose activity when immobilized. Because arrays must be probed under native conditions, antibodies may capture multiprotein complexes as well as individual proteins, which complicates interpretation. Short strands of chemically synthe-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

sized nucleic acid (aptamers) have been studied as potential monospecific recognition molecules for microarrays, and this technology may eventually overcome some of the problems with antibody arrays (Smith et al. 2003; Kirby et al. 2004). "Reversed-phase" microarrays, which consist of multiple samples of protein mixtures (for example, tissues, cell lysates), are probed with individual antibodies (Paweletz et al. 2001; Janzi et al. 2005). This establishes the presence of the target protein in multiple samples rather than the presence of multiple proteins in any sample. As with antibody microarrays, the main limitations of this approach stem from the quality and availability of antibodies for the targets of interest. Microarrays of expressed proteins or protein domain substructures have been probed with tagged proteins or small molecules to identify protein binding partners or small molecule ligand sites or to conduct surveys of substrates for enzymes (for example, kinases) (Zhu et al. 2001; Ramachandran et al. 2004). This approach is directed to functional analysis of known proteins as opposed to identification and analysis of the components of complex mixtures. A related technique directed at the study of protein-protein interactions is surface plasmon resonance (SPR) (Liedberg et al. 1995; Homola 2003; Yuk and Ha 2005). This technology allows real-time measurements of protein binding affinities and interactions. In common usage, SPR is used to study single pairs of interacting species. However, recent adaptations of SPR allow direct analysis of protein-protein interactions in microarray format (Yuk et al. 2004).

METABOLOMIC TECHNOLOGIES

Metabolomics[1] is the analysis of collections of small molecule intermediates and products of diverse biologic processes. Metabolic intermediates reflect the actions of proteins in biochemical pathways and thus represent biologic states in a way analogous to proteomes. As with proteomes, metabolomes are dynamic and change in response to nutrition, stress, disease states, and even diurnal variations in metabolism. Unlike genomes, transcriptomes, and proteomes, metabolomes comprise a chemically diverse collection of compounds, which range from small peptide, lipid, and nucleic acid precursors and degradation products to chemical intermediates in biosynthesis and catabolism as well as metabolites of exogenous compounds derived from the diet, environmental exposures, and therapeutic interventions. A consequence of the chemical diversity of metabolome components is the difficulty of comprehensive analysis with any single analytical technology.

[1]Although some scientists attempt to distinctly define the terms metabolomics and metabonomics, the committee uses the term metabolomics throughout the report simply because it is used more frequently in the literature.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

NMR-Based Metabolomics

The principal technology platforms for metabolomics are NMR spectroscopy and gas chromatography MS (GC-MS) or LC-MS. NMR has been the dominant technology for metabolomic studies of biofluids ex vivo. High-field (600 mHz) ^{1}H-NMR spectra of urine contain thousands of signals representing hundreds to thousands of metabolites (Nicholson et al. 2002). Hundreds of analytes have been identified in such spectra and collectively represent a plurality of metabolic processes and networks from multiple organs and tissues. Although NMR has been most commonly applied to urine samples, similar analyses of intact solid tissues were accomplished with the use of magic angle spinning ^{1}H-NMR (Waters et al. 2000; Nicholson et al. 2002; Y. Wang et al. 2003).

Although it is possible to establish the identity of many, but not all, of the peaks in NMR spectra of urine and biofluids, the value of the data has been in the analyses of collections of spectral signals. These pattern recognition approaches have been used to identify distinguishing characteristics of samples or sample sets. Unsupervised[2] analyses of the data, such as principal components analysis (PCA), have proven useful for grouping samples based on sets of similar features (Beckwith-Hall et al. 1998; Holmes et al. 1998). These similar features frequently reflect chemical similarity in metabolite composition and thus similar courses of response to toxicants. Supervised analyses allow the use of data from biochemically or toxicologically defined samples to establish models capable of classifying samples based on multiple features in the spectra (Stoyanova et al. 2004).

NMR-based metabolomics of urine measure global metabolic changes that have occurred throughout an organism. However, metabolite profiles in urine can also indicate tissue-specific toxicities. PCA of urinary NMR data have shown that the development and resolution of chemically induced tissue injury can be followed by plotting trajectories of PCA-derived parameters (Azmi et al. 2002). Although the patterns themselves provide a basis for analyses, some specific metabolites have also been identified (based on their resonances in the NMR spectra). Mapping these metabolites onto known metabolic pathways makes it possible to draw inferences about the biochemical and cellular consequences and mechanisms of injury (Griffin et al. 2004). An interesting and important consequence was the identification of endogenous bacterial metabolites as key elements of diagnostic metabonomic profiles (Nicholls et al. 2003; Wilson and Nicholson 2003; Robosky et al. 2005). Although the interplay of gut bacteria with drug and chemical metabolism had been known previously, recent NMR metabolomic studies indicate that interactions between host tissues and gut microbes have a much more pronounced effect on susceptibility to injury than had been appreciated previously (Nicholson et al. 2005).

[2]Unsupervised analysis methods look for patterns in the data without using previous knowledge about the data; information about treatment or classification supervised methods use this knowledge. See Chapter 3 for more details.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

A critical issue in the application of metabolomics is the standardization of methods, data analysis, and reporting across laboratories. A recent cooperative study by the Consortium for Metabonomic Toxicology indicated that NMR-based technology is robust and reproducible in laboratories that follow similar analytical protocols (Lindon et al. 2003). Investigators in the field recently have agreed on consensus standards for analytical standardization and data representation in metabonomic analyses (Lindon et al. 2005a).

MS-Based Metabolomics

MS-based analyses offer an important alternative approach to metabolomics. The greatest potential advantage of MS-based methods is sensitivity. MS analyses can detect molecules at levels up to 10,000-fold lower than does NMR (Brown et al. 2005; Wilson et al. 2005a). Both GC-MS and LC-MS approaches have been used, although limits of volatility of many metabolites reduce the range of compounds that can be analyzed successfully with GC-MS. LC-MS analyses are done with both positive and negative ion electrospray ionization and positive and negative chemical ionization. These four ionization methods provide complementary coverage of diverse small molecule chemistries. The principal mode of analysis is via "full scan" LC-MS, in which the mass range of the instrument is repeatedly scanned (Plumb et al. 2002; Wilson et al. 2005b). This analysis records the mass-to-charge ratios and retention times of metabolites. Because most small molecules produce singly charged ions, the analyses provide molecular weights of the metabolites. Analysis of standards in the same system and the use of MS-MS analysis can establish the identity of the components of interest. However, apparent molecular weight measurement alone is often insufficient to generate candidate metabolite identifications; frequently, hundreds or thousands of molecules are being analyzed. Nevertheless, accurate information about molecular weight, where possible, is of great value in identification. For this reason LC-MS metabolomic analyses are most commonly done with higher mass accuracy MS instruments, such as LC TOF, quadruple TOF, and FTICR MS instruments (Wilson et al. 2005a,b). NMR- and MS-based approaches provide complementary platforms for metabolomic studies and an integration of these platforms will be needed to provide capabilities that are most comprehensive. Clearly, either platform can detect metabolite profile differences sufficient to distinguish different toxicities. What is not yet clear is the degree to which either approach can resolve subtly different phenotypes.

TECHNOLOGY ADVANCEMENT AND ECONOMY OF SCALE

A major determinant of success in genome sequencing projects was achieving economy of scale for genome sequencing technologies (see above). The successful implementation of large-scale toxicogenomic initiatives will require advances in standardization and economy of scale. Most proteomic and

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

metabolomic analyses in academic and industry laboratories are done on a small scale to address specific research questions. The evolution of transcriptome profiling and proteomic and metabolomic technology platforms to increase standardization and reduce costs will be essential to maximize their impact.

Copyright © National Academy of Sciences. All rights reserved.

3

Experimental Design and Data Analysis

The greatest challenge of toxicogenomics is no longer data generation but effective collection, management, analysis, and interpretation of data. Although genome sequencing projects have managed large quantities of data, genome sequencing deals with producing a reference sequence that is relatively static in the sense that it is largely independent of the tissue type analyzed or a particular stimulation. In contrast, transcriptomes, proteomes, and metabolomes are dynamic and their analysis must be linked to the state of the biologic samples under analysis. Further, genetic variation influences the response of an organism to a stimulus. Although the various toxicogenomic technologies (genomics, transcriptomics, proteomics, and metabolomics) survey different aspects of cellular responses, the approaches to experimental design and high-level data analysis are universal.

This chapter describes the essential elements of experimental design and data analysis for toxicogenomic experiments (see Figure 3-1) and reviews issues associated with experimental design and data analysis. The discussion focuses on transcriptome profiling using DNA microarrays. However, the approaches and issues discussed here apply to various toxicogenomic technologies and their applications. This chapter also describes the term biomarker.

EXPERIMENTAL DESIGN

The types of biologic inferences that can be drawn from toxicogenomic experiments are fundamentally dependent on experimental design. The design must reflect the question that is being asked, the limitations of the experimental system, and the methods that will be used to analyze the data. Many experiments using global profiling approaches have been compromised by inadequate consideration of experimental design issues. Although experimental design for toxi-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

cogenomics remains an area of active research, a number of universal principles have emerged. First and foremost is the value of broad sampling of biologic variation (Churchill 2002; Simon et al. 2002; Dobbin and Simon 2005). Many early experiments used far too few samples to draw firm conclusions, possibly because of the cost of individual microarrays. As the cost of using microarrays and other toxicogenomic technologies has declined, experiments have begun to include sampling protocols that provide better estimates of biologic and systematic variation within the data. Still, high costs remain an obstacle to large, population-based studies. It would be desirable to introduce power calculations into the design of toxicogenomic experiments (Simon et al. 2002). However, uncertainties about the variability inherent in the assays and in the study populations, as well as interdependencies among the genes and their levels of expression, limit the utility of power calculations.

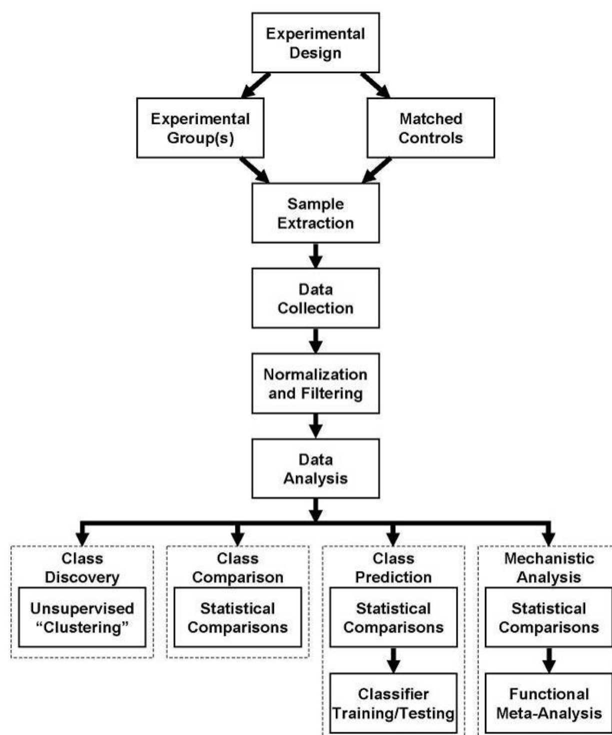

FIGURE 3-1 Overview of the workflow in a toxicogenomic experiment. Regardless of the goal of the analysis, all share some common elements. However, the underlying experimental hypothesis, reflected in the ultimate goal of the analysis, should dictate the details of every step in the process, starting from the experimental design and extending beyond what is presented here to the methods used for validation.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

A second lesson that has emerged is the need for carefully matched controls and randomization in any experiment. Because microarrays and other toxicogenomic technologies are extremely sensitive, they can pick up subtle variations in gene, protein, or metabolite expression that are induced by differences in how samples are collected and handled. The use of matched controls and randomization can minimize potential sources of systematic bias and improve the quality of inferences drawn from toxicogenomic datasets.

A related question in designing toxicogenomic experiments is whether samples should be pooled to improve population sampling without increasing the number of assays (Dobbin and Simon 2005; Jolly et al. 2005; Kendziorski et al. 2005). Pooling averages variations but may also disguise biologically relevant outliers—for example, individuals sensitive to a particular toxicant. Although individual assays are valuable for gaining a more robust estimate of gene expression in the population under study, pooling can be helpful if experimental conditions limit the number of assays that can be performed. However, the relative costs and benefits of pooling should be analyzed carefully, particularly with respect to the goals of the experiment and plans for follow-up validation of results. Generally, the greatest power in any experiment is gained when as many biologically independent samples are analyzed as is feasible.

Universal guidelines cannot be specified for all toxicogenomic experiments, but careful design focused on the goals of the experiment and adequate sampling are needed to assess both the effect and the biologic variation in a system. These lessons are not unique to toxicogenomics. Inadequate experimental designs driven by cost cutting have forced many studies to sample small populations, which ultimately compromises the quality of inferences that can be drawn.

TYPES OF EXPERIMENTS

DNA microarray experiments can be categorized into four types: class discovery, class comparison, class prediction, and mechanistic studies. Each type addresses a different goal and uses a different experimental design and analysis. Table 3-1 summarizes the broad classes of experiments and representative examples of the data analysis tools that are useful for such analyses. These data analysis approaches are discussed in more detail below.

Class Discovery

Class discovery analysis is generally the first step in any toxicogenomic experiment because it takes an unbiased approach to looking for new group classes in the data. A class discovery experiment asks “Are there unexpected, but biologically interesting, patterns that exist in the data?” For example, one might consider an experiment in which all nephrotoxic compounds are used individually to treat rats, and gene expression data are collected from the kidneys

Copyright © National Academy of Sciences. All rights reserved.

TABLE 3-1 Data Analysis Approaches

Application[a]	Algorithm[b]	Representative References[c]
Class Discovery	Hierarchical clustering	Weinstein et al. 1997; Eisen et al. 1998; Wen et al. 1998
	k-means clustering	Soukas et al. 2000
	Self-organizing maps	Tamayo et al. 1999; Toronen et al. 1999; Wang et al. 2002
	Self-organizing trees	Herrero et al. 2001
	Relevance networks	Butte and Kohane 1999
	Force-directed layouts	Kim et al. 2001
	Principal component analysis	Raychaudhuri et al. 2000
Class Comparison	*t*-test	Baggerly et al. 2001
	Significance analysis of microarrays	Tusher et al. 2001
	Analysis of variance	Long et al. 2001
Class Prediction	*k*-nearest neighbors	Theilhaber et al. 2002
	Weighted voting	Golub et al. 1999
	Artificial neural networks	Bloom et al. 2004; Ellis et al. 2002
	Discriminant analysis	Nguyen and Rocke 2002; Orr and Scherf 2002; Antoniadis et al. 2003; Le et al. 2003
	Classification and regression trees	Boulesteix et al. 2003
	Support vector machines	Brown et al. 2000; Ramaswamy et al. 2001
Functional and Network Inference for Mechanistic Analysis	EASE	Hosack et al. 2003
	MAPPFinder	Doniger et al. 2003

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

	GOMiner	Zeeberg et al. 2003
	Cytoscape	Shannon et al. 2003
	Boolean networks	Akutsu et al. 2000; Savoie et al. 2003; Soinov 2003
	Probabilistic boolean networks	Shmulevich et al. 2002a,b; Datta et al. 2004; Hashimoto et al. 2004
	Bayesian networks	Friedman et al. 2000; Imoto et al. 2003; Savoie et al. 2003; Tamada et al. 2003; Zou and Conzen 2005

[a]Application of these analytical tools is not limited to individual datasets but can be applied across studies if the data and relevant ancillary information (for example, about treatment, phenotype) are available.
[b]A wide range of algorithms has been developed to facilitate analysis of toxicogenomic datasets. Although most approaches have been applied in the context of gene expression microarray data, the algorithms are generally applicable to any toxicogenomic-based data. A representative sample is presented here; many similar approaches are being developed.
[c]In general, the citations represent the first published use of a particular method or those that are most widely cited.

of these rats after they begin to experience renal failure (Amin et al. 2004). Evaluation of the gene expression data may indicate the nephrotoxic compounds can be grouped based on the cell type affected, the mechanism responsible for renal failure, or other common factors. This analysis may also suggest a new subgroup of nephrotoxic compounds that either affects a different tissue type or represents a new toxicity mechanism.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

of these rats after they begin to experience renal failure (Amin et al. 2004). Evaluation of the gene expression data may indicate the nephrotoxic compounds can be grouped based on the cell type affected, the mechanism responsible for renal failure, or other common factors. This analysis may also suggest a new subgroup of nephrotoxic compounds that either affects a different tissue type or represents a new toxicity mechanism.

Class discovery analyses rely on *unsupervised* data analysis methods (algorithms) to explore expression patterns in the data (see Box 3-1). Unsupervised data analysis methods are often among the first techniques used to analyze a microarray dataset. Unsupervised methods do not use the sample classification as input; for example, they do not consider the treatment groups to which the samples belong. They simply group samples together based on some measure of similarity. Two of the most widely used unsupervised approaches are hierarchical clustering (Weinstein et al. 1997; Eisen et al. 1998; Wen et al. 1998) and *k*-means clustering (Soukas et al. 2000). Other approaches have been applied to unsupervised analysis, including self-organizing maps (Tamayo et al. 1999; Toronen et al. 1999; Wang et al. 2002), self-organizing trees (Herrero et al. 2001), relevance networks (Butte and Kohane 1999), force-directed layouts (Kim et al. 2001), and principal component analysis (Raychaudhuri et al. 2000). Fundamentally, each of these methods uses some feature of the data and a rule for determining relationships to group genes (or samples) that share similar patterns of expression. In the context of disease analysis, all the methods can be extremely useful for identifying new subclasses of disease, provided that the subclasses are reproducible and can be related to other clinical data. All these methods will divide data into "clusters," but determining whether the clusters are meaningful requires expert input and analysis. Critical assessment of the results is essential. There are anecdotal reports of clusters being found that separate data based on the hospital where the sample was collected, the technician who ran the microarray assay, or the day of the week the array was run. Clearly, microarrays can be very sensitive. However, recent reports suggest that adhering to standard laboratory practices and carefully analyzing data can lead to high-quality, reproducible results that reflect the biology of the system (Bammler et al. 2005; Dobbin et al. 2005; Irizarry et al. 2005; Larkin et al. 2005).

In the context of toxicogenomics, class discovery methods can be applied to understand the cellular processes involved in responding to specific agents. For example, in animals exposed to a range of compounds, analysis of gene expression profiles with unsupervised "clustering" methods can be used to discover groups of genes that may be involved in cellular responses and suggest hypotheses about the modes of action of the compounds. Subsequent experiments can confirm the gene expression effects, confirm or refute the hypotheses, and identify the cell types and mode of action associated with the response. A goal of this type of research would be to build a database of gene expression profiles of sufficiently high quality to enable a gene expression profile to be used to classify compounds based on their mode of action.

Copyright © National Academy of Sciences. All rights reserved.

Class Comparison

Class comparison experiments compare gene expression profiles of different phenotypic groups (such as treated and control groups) to discover genes and gene expression patterns that best distinguish the groups. The starting point in such an experiment is the assumption that one knows the classes represented in the data. A logical approach to data analysis is to use information about the various classes in a *supervised* fashion to identify those genes that distinguish the groups. One starts by assigning samples to particular biological classes based on objective criteria. For example, the data may represent samples from treatment with neurotoxic and hepatotoxic compounds. The first question would be, "Which genes best distinguish the two classes in the data?" At this stage, the goal is to find the genes that are most informative for distinguishing the samples based on class.

A wide variety of statistical tools can be brought to bear on this question, including *t*-tests (for two classes) and analysis of variance (for three or more classes) that assign *p* values to genes based on their ability to distinguish among groups. One concern with these statistical approaches is the problem of multiple testing. Simply put, in a microarray with 10,000 genes, applying a 95% confidence limit on gene selection ($p \leq 0.05$) means that, by chance, one would expect to find 500 genes as significant. Stringent gene selection can minimize but not eliminate this problem; consequently, one must keep in mind that the greatest value of statistical methods is that they provide a way to prioritize genes for further analysis. Other approaches are widely used, such as significance analysis of microarrays (Tusher et al. 2001), which uses an adjusted *t* statistic (or *F* statistic) modified to correct for overestimates arising from small values in the denominator, along with permutation testing to estimate the false discovery rate in any selected significant gene set. Other methods attempt to correct for multiple testing, such as the well-known Bonferroni correction, but these methods assume independence between the measurements, a constraint that is violated in gene analysis as many genes and gene products operate together in pathways and networks and so are co-regulated. Further confounding robust statistical analysis of toxicogenomic studies is the "n < p problem," which means the number of samples analyzed is typically much smaller than the number of genes, proteins, or metabolites assayed. For these reasons, statistical analysis of higher-dimensional datasets produced by toxicogenomic technologies remains an area of active research.

As described above, class comparison analyses provide collections of genes that the data indicate are useful in distinguishing among the various experimental groups being studied. These collections of genes can be used either as a starting point for mechanistic studies or in an attempt to classify new compounds as to their mode of action.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 3-1 Supervised and Unsupervised Analysis

Analysis methods for toxicogenomic data can be divided into two broad classes depending on how much prior knowledge, such as information about the samples, is used.

Unsupervised methods examine the data without benefit of information about the samples. Unsupervised methods, such as hierarchical clustering, are particularly useful in class discovery studies as they group samples without prior bias and may allow new classes to be found in samples previously thought to be "identical." These methods are also useful for quality control in large experiments as they can verify similarity among related samples or identify outliers (for example, failed assays).

Supervised methods use information about the samples being analyzed to find features in the data. Most statistical approaches are supervised; once samples are assigned to groups, the data for each gene, protein, or metabolite are compared across groups to find those that can distinguish between groups. Class comparison and classification studies use supervised methods in the early stages of analysis.

Class Prediction

Class prediction experiments attempt to predict biologic effects based on the gene expression profile associated with exposure to a compound. Such an experiment asks "Can a particular pattern of gene expression be combined with a mathematical rule to predict the effects of a new compound?" The underlying assumption is that compounds eliciting similar effects will elicit similar effects on gene expression. Typically, one starts with a well-characterized set of compounds and associated phenotypes (a database of phenotype and gene expression data) and through a careful comparison of the expression profiles finds genes whose patterns of expression can be used to distinguish the various phenotypic groups under analysis. Class prediction approaches then attempt to use sets of informative genes (generally selected using statistical approaches in class comparison) to develop mathematical rules (or computational algorithms) that use gene expression profiling data to assign a compound to its phenotype group (class). The goal is not merely to separate the samples but to create a rule (or algorithm) that can predict phenotypic effects for new compounds based solely on gene expression profiling data.

For example, to test a new compound for possible neurotoxicity, gene expression data for that compound would be compared with gene expression data for other neurotoxic compounds in a database and a prediction would be made about the new compound's toxicity. (The accuracy of the prediction depends on the quality of the databases and datasets.)

When developing a classification approach, the mathematical rules for analyzing new samples are encoded in a *classification algorithm*. A wide range

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

of algorithms have been used for this purpose, including weighted voting (Golub et al. 1999), artificial neural networks (Ellis et al. 2002; Bloom et al. 2004), discriminant analysis (Nguyen and Rocke 2002; Orr and Scherf 2002; Antoniadis et al. 2003; Le et al. 2003), classification and regression trees (Boulesteix et al. 2003), support vector machines (Brown et al. 2000; Ramaswamy et al. 2001), and *k*-nearest neighbors (Theilhaber et al. 2002). Each of these uses an original set of samples, or training set, to develop a rule that uses the gene expression data (trimmed to a previously identified set of informative genes) for a new compound to place this new compound into the context of the original sample set, thus identifying its class.

Functional and Network Inference for Mechanistic Analysis

Although class prediction analysis may tell us what response a particular compound is likely to produce, it does not necessarily shed light on the underlying mechanism of action. Moving from class prediction to mechanistic understanding often relies on additional work to translate toxicogenomic-based hypotheses to validated findings. Bioinformatic tools play a key role in developing those hypotheses by integrating information that can facilitate interpretation—including gene ontology terms, which describe gene products (proteins), functions, processes, and cellular locations; pathway database information; genetic mapping data; structure-activity relationships; dose-response curves; phenotypic or clinical information; genome sequence and annotation; and other published literature. Software developed to facilitate this analysis includes MAPPFinder (Doniger et al. 2003), GOMiner (Zeeberg et al. 2003), and EASE (Hosack et al. 2003), although they may only provide hints about possible mechanisms. There is no universally accepted way to connect the expression of genes, proteins, or metabolites to functionally relevant pathways leading to particular phenotypic end points, so a good deal of user interaction and creativity is currently required.

New approaches to predict networks of interacting genes based on gene expression profiles use several modeling techniques, including boolean networks (Akutsu et al. 2000; Savoie et al. 2003; Soinov 2003), probabilistic boolean networks (Shmulevich et al. 2002a,b; Datta et al. 2004; Hashimoto et al. 2004;), and Bayesian networks (Friedman et al. 2000; Imoto et al. 2003; Savoie et al. 2003; Tamada et al. 2003; Zou and Conzen 2005). These models treat individual objects, such as genes and proteins, as "nodes" in a graph, with "edges" connecting the nodes representing their interactions. A set of rules for each edge determines the strength of the interaction and whether a particular response will be induced. These approaches have met with some success, but additional work is necessary to convert the models from descriptive to predictive. In metabolic profiling, techniques that monitor metabolic flux and its modeling (Wiback et al. 2004; Famili et al. 2005) also may provide predictive models.

The advent of global toxicogenomic technologies, and the data they provide, offers the possibility of developing quantitative, predictive models of bio-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

logic systems. This approach, dubbed "systems biology," attempts to bring together data from many different domains, such as gene expression data and metabolic flux analysis, and to synthesize them to produce a more complete understanding of the biologic response of a cell, organ, or individual to a particular stimulus and create predictive biomathematical models. Whereas toxicogenomic data are valuable even when not used in a systems biology mode, achieving this systems-level understanding of organismal response and its relationship to the development of a particular phenotype is a long-term goal of toxicogenomics and other fields. The best efforts to date have allowed the prediction of networks of potentially interacting genes. However, these network models, while possibly predictive, lack the complexity of the biochemical or signal transduction pathways that mediate cellular responses. Attempts to model metabolic flux, even in simpler organisms like yeast and bacteria, provide only rough approximations of system function responses and then only under carefully controlled conditions. However, significant progress in the ability to model complex systems is likely and additional toxicogenomic research will continue to benefit from and help advance systems biology approaches and their applications.

TOXICOGENOMICS AND BIOMARKER DISCOVERY

An opinion paper by Bailey and Ulrich (2004) outlined the use of microarrays and related technologies for identifying new biomarkers; see Box 3-2 for definitions. Within the drug industry, there is an acute need for effective biomarkers that predict adverse events earlier than otherwise could be done in every phase of drug development from discovery through clinical trials, including a need for noninvasive biomarkers for clinical monitoring. There is a widespread expectation that, with toxicogenomics, biomarker discovery for assessing toxicity will advance at an accelerated rate. Each transcriptional "fingerprint" reflects a cumulative response representing complex interactions within the organism that include pharmacologic and toxicologic effects. If these interactions can be significantly correlated to an end point, and shown to be reproducible, the molecular fingerprint potentially can be qualified as a predictive biomarker. Several review articles explore issues related to biomarker assay development and provide examples of the biomarker development process (Wagner 2002; Colburn 2003; Frank and Hargreaves 2003).

The utility of gene expression-based biomarkers was clearly illustrated by van Leeuwen and colleagues' 1986 identification of putative transcriptional biomarkers for early effects of smoking using peripheral blood cell profiling (van Leeuwen et al. 1986). Kim and coworkers also demonstrated a putative transcriptional biomarker that can identify genotoxic effects but not carcinogenesis using lymphoma cells but noted that the single marker presented no clear advantage over existing in vitro or in vivo assays (Kim et al. 2005). Sawada et al. discovered a putative transcriptional biomarker predicting phospholipidosis in the HepG2 cell line, but they too saw no clear advantage over exist-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

ing assays (Sawada et al. 2005). In 2004, a consortium effort based at the International Life Sciences Institute's Health and Environmental Sciences Institute identified putative gene-based markers of renal injury and toxicity (Amin et al. 2004). As has been the case for transcriptional markers, protein-based expression assays have also shown their value as predictive biomarkers. For example, Searfoss and coworkers used a toxicogenomic approach to identify a protein biomarker for intestinal toxicity (Searfoss et al. 2003).

Exposure biomarker examples also exist. Koskinen and coworkers developed an interesting model system in rainbow trout, using trout gene expression microarrays to develop biomarkers for assessing the presence of environmental contaminants (Koskinen et al. 2004). Gray and colleagues used gene expression in a mouse hepatocyte cell line to identify the presence of aromatic hydrocarbon receptor ligands in an environmental sample (Gray et al. 2003).

BOX 3-2 Defining Biomarkers

Throughout this chapter, a wide range of applications of gene expression microarray and other toxicogenomic technologies has been discussed. Many of the most promising applications involve using gene, protein, or metabolic expression profiles as diagnostic or prognostic indicators and refer to them as biomarkers. However, use of this term has been rather imprecise, in part because the term has developed a rather broad range of interpretations and associations with detection of a range of measurable end points.

To resolve some of the potential confusion about the term's use, the National Institutes of Health (NIH) formed a committee to provide working definitions for specific terms and a conceptual model of how biomarkers could be used (BDW Group 2001). According to the NIH Initiative on Biomarkers and Surrogate Endpoints a biologic marker (biomarker) is defined as follows: "A characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." A biomarker is distinguished from a clinical end point, which is defined as "a characteristic or variable that reflects how a patient feels, functions, or survives" and is distinguished from a surrogate end point, which is defined as "a biomarker that is intended to substitute for a clinical end point. A surrogate end point is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence."

In the terminology used in this report, this NIH definition is consistent with a "biomarker of effect," whereas the phrase "biomarker of exposure" is more consistent with the following definition from the National Research Council's 2006 report *Human Biomonitoring for Environmental Chemicals:* "A biomarker of exposure is a chemical, its metabolite, or the product of an interaction between a chemical or some target molecule or cell that is measured in humans." (NRC 2006d, p.4).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Ultimately, toxic response is likely to be mediated by changes at various levels of biologic organization: gene expression, protein expression, and altered metabolic profiles. Whereas most work to date has focused on developing biomarkers based on the output of single toxicogenomic technologies (for example, transcriptomics, proteomics, metabolomics), an integrated approach using multiple technologies provides the opportunity to develop multidomain biomarkers that are more highly predictive than those derived from any single technology. Further, existing predictive phenotypic (and genotypic) measures should not be ignored in deriving biomarkers.

Finally, particular attention must be paid to developing toxicogenomic-based biomarkers, especially those that are not tied mechanistically to a particular end point. In 2001, Pepe and colleagues outlined the stages of cancer biomarker development (Pepe et al. 2001) (see Table 3-2), suggesting that a substantial effort involving large populations would be required to fully validate a new biomarker for widespread clinical application. The Netherlands breast cancer study discussed in Chapter 9 (validation) is an example of a biomarker that has reached Phase 4, with a prospective analysis in which 6,000 women will be recruited and screened at an estimated cost of $54 million (Bogaerts et al. 2006; Buyse et al. 2006). Most toxicogenomic studies have reached only Phase 1 or Phase 2 and significant additional work and funding are necessary if toxicogenomic biomarkers are to achieve the same level of validation.

CONCLUSIONS

This chapter focused largely on questions of experimental design and the associated analytical approaches that can be used to draw biologic inferences. The published examples have largely drawn on individual studies in which datasets have been analyzed in isolation. However, any of these methods can be applied more generally to larger collections of data than those in individual studies, provided that the primary data and the information needed to interpret them are available.

Clearly, a carefully designed database containing toxicogenomic data along with other information (such as structure-activity relationships and information about dose-response and phenotypic outcome for exposure) would allow many of the unanswered questions about the applicability of genomic technologies to toxicology to be addressed. In fact, a more extensive analysis would allow scientists to more fully address questions about reproducibility, reliability, generalizability, population effects, and potential experimental biases that might exist and that would drive the development of standards and new analytical methods.

A distinction must be drawn between datasets and a database. A database compiles individual datasets and provides a structure for storing the data that captures various relationships between elements, and it facilitates our ability to

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

TABLE 3-2 Phases of Cancer Biomarker Development As Defined by Pepe et al. (2001)

Phase 1	Preclinical exploratory	Promising directions identified
Phase 2	Clinical assay and validation	Clinical assay detects established disease
Phase 3	Retrospective longitudinal	Biomarker detects disease before it becomes clinical and a "screen-positive" rule is defined
Phase 4	Prospective screening	Extent and characteristics of disease detected by the test and the false referral rate are identified
Phase 5	Cancer control	Impact of screening on reducing the burden of disease on the population is quantified

investigate associations among various elements. Such a database must go beyond individual measurements and provide information about, for example, how the individual experiments are designed, the chemical properties of the individual compound tested, the phenotypes that result, and the genetic background of the animals profiled. Many considerations must go into designing such a database and populating it with relevant data; a more detailed discussion is provided in Chapter 10. However, creating such a database that captures relevant information would allow more extensive data mining and exploration and would provide opportunities currently not available. Making full use of such a database would also require a commitment to develop new analytical methods and to develop software tools to make these analytical methods available to the research and regulatory communities.

Although assembling a central toxicogenomic database would be a massive undertaking, creating such a resource, with a focus not only on data production but also on delivery of protocols, databases, software, and other tools to the community, should serve as a catalyst to encourage others to contribute to building a more comprehensive database. Mechanisms should be investigated that would facilitate the growth of such a database by using data from academic and industrial partners. When possible and feasible, attention should be paid to integrating development of such a database and related standards with the work of parallel efforts such as caBIG (NCI 2006d) at the National Cancer Institute. The success of any toxicogenomic enterprise depends on data and information and the National Institute of Environmental Health Sciences and other federal agencies must make an investment to produce and provide those to the research community.

RECOMMENDATIONS

1. Develop specialized bioinformatics, statistical, and computational tools to analyze toxicogenomic data. This will require a significant body of carefully collected controlled data, suggesting the creation of a national data resource

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

open to the research community. Specific tools that are needed include the following:

a. Algorithms that facilitate accurate identification of orthologous genes and proteins in species used in toxicologic research,
b. Tools to integrate data across multiple analytical platforms (for example, gene sequences, transcriptomics, proteomics, and metabolomics), and
c. Computational models to enable the study of network responses and systems-level analyses of toxic responses.

2. Continue to improve genome annotation for all relevant species and elucidation of orthologous genes and pathways.

3. Emphasize the development of standards to ensure data quality and to assist in validation.

Copyright © National Academy of Sciences. All rights reserved.

4

Application to Exposure Assessment

People are exposed daily to diverse chemical and physical agents in the environment, some of which may adversely affect human health. These agents are present across an enormous range of concentrations, which results in a wide variety of human exposures.

In this context, it is useful to consider the distinction between exposure and dose. Exposure is the amount of a given agent presented by the immediate environment, whereas dose is the amount of an agent absorbed by the individual. Thus, dose has an implicit time component, which depends on the duration of exposure. Individuals in the same environment for the same amount of time may experience different doses of a given compound because of differences in absorption or metabolism. Human behavior adds to the variables. For example, individuals consuming arsenic-contaminated drinking water from the same source may absorb different doses simply because they drink different amounts of the water. Individual variations in physiology or metabolism can further modify the dose by affecting the rate of clearance or enzymatic processing. Finally, the dose to specific organs often depends on metabolic pathways, as in the concentration of radioactive iodine in the thyroid.

Despite these challenges, environmental toxicologists have long sought to use the modifications or expression changes of specific biomolecules to determine actual exposures in individuals and in populations. This chapter describes the current status of exposure assessment with biologic molecules as indicators, focusing on the application of toxicogenomic technologies to assess human exposure to environmental agents. The hope and expectation is that these new toxicogenomic approaches will allow more refined and sensitive assessment of exposure through the measurement of more subtle and broader changes in gene and protein expression and through identification and quantification of greater varieties of metabolites. The greater detail provided by toxicogenomic information might also enable the distinction of individual components of exposure in

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

complex mixtures or the detection of exposures after time has elapsed, providing "fingerprints" of biologic responses to agents that are not retained in tissues, such as ionizing radiation.

This chapter focuses on applications of toxicogenomic technologies to measure environmental exposures in populations exposed to occupational and environmental agents, an important task in risk assessment and toxicology research. The related problem of exposure to pharmaceutical agents is not considered here. The relevance of toxicogenomics to risk assessment modeling is discussed after a description of the state of the art.

CONVENTIONAL BIOMARKERS OF EXPOSURE AND RESPONSE

The term biomarker has been widely applied to describe quantifiable molecular species that reflect biologic states, exposures, and disease. Under this broad heading, two biomarker subcategories can be distinguished: biomarkers of exposure, reflecting the occurrence of an exposure; and biomarkers of response, which indicate the response of an organism to an exposure.

Biomarkers of exposure may be specific modifications of specific molecules, such as the adducts formed on hemoglobin due to exposure to benzene or pyrolysis products (Skipper et al. 1994; Medeiros et al. 1997; Alexander et al. 2002), the DNA adducts produced by exposure to vinyl chloride or urethane (Skipper et al. 1994), or the polycyclic aromatic hydrocarbons from cigarette smoke (Shugart et al. 1983; Perera et al. 1986). Although xenobiotics and the metabolites that persist in tissues (for example, polychlorinated biphenyls, dioxins) are used as exposure biomarkers, adducts are the most commonly used biomarkers of exposure. Adducts can persist detectably in the organism, constituting biomarkers of a past exposure. Exposure biomarkers such as hemoglobin adducts might not be involved in toxic effects, but others (for example, mutagenic DNA damage) may be.

Single-molecule species have been most commonly used as biomarkers, and traditionally they have been measured by conventional approaches such as gas chromatography and high-performance liquid chromatography. However, a collection of biomarkers may also be used as a fingerprint of exposure. Monitoring multiple biomarker molecules could boost the sensitivity of detection as well as its specificity in reporting a particular exposure type. One reason is that different agents can produce overlapping profiles of adducts, so that measuring multiple types of adducts could distinguish related but different exposures. Such distinctions can help define exposures associated with different disease risks. For example, weakly carcinogenic methylating agents such as methyl methanesulfonate produce N^7-methylguanine as the predominant DNA adduct, which is also the main DNA lesion formed by potent carcinogens such as *N*-methyl-*N*'-nitro-*N*-nitrosourea. The key mutational and carcinogenic effects of these compounds, however, are mainly due to differences in the levels of relatively minor lesions: the strongly mutational adduct O^6-methylguanine accounts for only

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

about 0.1% of DNA damage caused by the weak mutagen methyl methanesulfonate, but it accounts for about 7% of the damage caused by the potent mutagen *N*-methyl-*N*'-nitro-*N*-nitrosourea (Montesano et al. 1980; Pegg 2000).

Adducts and persistent metabolites are just one measure of exposure. Cells and tissues also alter their metabolism or gene expression in response to exposure. Well-known examples include the induction of specific groups of genes in response to heat shock, hypoxia, or osmotic stress (Finkel and Holbrook 2000; Chellappan 2001; Berra et al. 2006). These responses can produce patterns of specific changes in gene expression, proteins, or metabolic profiles that report the exposure to a particular agent or class of agent. Complex sets of gene expression changes and corresponding changes in protein networks and metabolite profiles have been beyond the reach of older technologies. Toxicogenomic technologies offer new opportunities to detect and quantify these changes and apply them as new classes of biomarkers. Some of the initial work along these lines is described in the next section.

STATE OF THE ART: TOXICOGENOMIC APPROACHES TO THE DEVELOPMENT OF EXPOSURE BIOMARKERS

Transcriptomics

As discussed in Chapter 2, transcriptional profiling seeks to catalog, at the level of messenger RNA (mRNA), the changes in gene expression that are provoked by exposure to chemical and physical agents or to compare differences among tissue types, developmental stages, genetic variants, and so forth. The approaches most commonly rely on microarray technologies.

The use of mRNA molecules to report cellular exposure to toxicants is not new. For example, Northern blot analysis (see Chapter 2) demonstrated that genotoxic exposure (for example, to ultraviolet (UV) light) activates the expression of *GADD* (growth arrest and DNA damage) genes (Fornace et al. 1988). More recently, conventional analysis of the transcriptional function of the tumor suppressor protein p53 revealed a number of genes whose expression is activated by p53 in response to DNA damage (Lakin and Jackson 1999). Thus, changes in individual p53-regulated transcripts, such as the *GADD* genes or the gene encoding the cell cycle arrest protein p21, are now often used to report cellular exposure to DNA-damaging agents.

The transcriptomic approach extends information of this type to most or all of the expressed genome. In principle, transcriptomics could exploit the overall pattern of gene expression moving beyond changes in just a few genes to generate more specific signatures of exposure. Greater specificity in tying patterns of change in gene expression to a particular exposure in this way is an important potential advantage of transcriptional profiling and of toxicogenomic technologies in general. For example, whereas many DNA-damaging agents activate p21 expression, other genes respond to UV light, and still others re-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

spond to X-rays (Lu and Lane 1993). To be used as biomarkers, altered levels of a given transcript need not be connected to a specific biologic end point, nor do the specific functions of all the mRNA molecules have to be known, although such information would be valuable.

Application to Human Exposure

Although applying such toxicogenomic technologies to determining human exposure lies in the future, experiments with model organisms or human cells in culture support the transcriptomic approach in its broad outline. In baker's yeast *Saccharomyces cerevisiae*, exposures to an oxidant (*t*-butyl hydroperoxide), UV light, or a DNA-alkylating agent led to transcriptional profiles with both common and agent-specific components (Begley and Samson 2004). Similar approaches in mammalian cell lines (Dickinson et al. 2004; Newton et al. 2004; Kim et al. 2005) also gave signatures for different agents that could be distinguished by their transcriptional profiles (van Delft et al. 2005). Of course, such carefully controlled exposures, using genetically uniform cell populations, are a far cry from the highly variable and complex exposures that occur in the human population.

Unlike microorganisms or cell lines in culture, additional complexity in studying human responses results from the existence of multiple organs and tissue types. Not all tissues or cell types will respond equally to a given agent, nor will all tissues be equally exposed as a function of the organism's exposure. This may constitute a significant challenge because of the practical limitations on which human tissues can be readily sampled. However, at least some modifications that are informative about exposure are likely to be available through analysis of DNA damage in circulating lymphocytes. For example, single-transcript studies showed that changes in CYP1A1 expression could be detected in the peripheral lymphocytes of railroad workers exposed to creosote, although no strong correlation with exposure could be established (Cosma et al. 1992). Additional complexity will come from genetic variation and perhaps individual life history that enhance the idiotypic nature of transcriptional responses in a human population. Nevertheless, the high level of information inherent to the transcriptomic approach may confer the ability to distinguish various types of exposure.

Quantitative Assessment

Applying transcriptional profiling to determine the dose of a given agent presents another level of challenge. All the processes affecting mRNA levels—transcription, processing, and turnover—are subject to thresholds for a response and maximum changes. Researchers often refer to the linear range for transcriptional (or other) responses, and the correlation of exposure and transcriptional induction is usually most accurate in this range (Baggerly et al. 2001; Larkin et

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

al. 2005). Thus, for any given gene, the extent of the linear response in an individual basically determines the range of assignable exposures. This limitation may be partly overcome by toxicogenomic approaches because the overall pattern of gene expression can be more informative than the behavior of an individual gene. On the other hand, significant changes in biologic mechanism may occur with increasing dose, so that the toxicogenomic readout—in metabolomics, for example—is altered qualitatively as well as quantitatively (Keun et al. 2004).

Persistence

Further issues concern the persistence of a detectable change after an exposure and whether changes in gene expression or other indicators can be detected after very brief exposures or after long-term exposures to very low levels of agents. Even large changes in gene or protein expression, or in metabolites, will eventually return to baseline levels with increasing time after an exposure. The high sensitivity of toxicogenomic approaches may extend the interval during which a brief or low-level exposure can be detected (exposures that are currently out of reach with conventional approaches), but the degree to which this is true will have to be determined by fundamental studies.

Proteomics

Whereas transcriptional profiling is a powerful method, proteins are generally the functional "business end" of genomic expression affecting cellular metabolism and regulation. Chemical exposures are manifested by two types of proteome changes. In the first, altered gene expression, mRNA stability, protein stability, or some combination of the three alters protein expression levels. In the second type of response, there is a change in the distribution of a protein between two or more modified forms. Such redistribution may result from physical modification of proteins by xenobiotic metabolites, directly or indirectly, or from effects of chemical exposure on the cellular systems that modify proteins (for example, protein kinases and phosphatases, protein cleaving activities, ubiquitylation).

New technologies enable changes in protein levels and protein modification to be assessed on a proteome-wide scale, with both types of changes referred to as proteomics. Analysis by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is one approach for detecting such changes. As of this writing, approximately a dozen studies describe the application of 2D-PAGE-based proteome analysis to identify protein biomarkers of chemical exposure. Xiao et al. (2003) examined the responses of a macrophage cell line exposed to diesel exhaust particles and identified expression changes for 32 proteins, which were linked functionally by their known roles in oxidative stress and antioxidant response networks (but see comment at the end of this paragraph). 2D-PAGE

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

analysis of liver tissue has been used to assess proteome changes in animal models of chemical exposure. An earlier study noted dozens of proteome changes that distinguish a toxic dose of acetaminophen from an equivalent dose of a nontoxic compound of a closely related structure (Myers et al. 1995). Changes in protein expression levels were detected in rat liver after treatment with bromobenzene, but few of the protein species involved were identifiable (Heijne et al. 2003). A more recent study detected 45 2D-PAGE features that were differentially altered in rat liver treated with the carcinogen *N*-nitrosomorpholine (Fella et al. 2005). Of the proteins identified (for example, antioxidant enzymes, heat shock and chaperone proteins), several were already known as characteristic of stress responses to multiple chemicals, some of which were found in earlier studies. The utility of these approaches hinges on whether the observed changes represent generalized responses to stress versus truly chemical-exposure-specific changes. This issue will be resolved only with a much larger body of studies with diverse chemicals in relevant biologic models.

Although the examples described below are interesting, they represent only a few published studies describing proteomic approaches to identify biomarkers of exposure. A key question to be answered about proteomic approaches is whether agent-specific changes can be identified. Satisfactory resolution of this issue will be possible only after a much larger body of proteomic data on exposed cells, tissues, and biofluids is collected and analyzed.

Accessibility

The foregoing discussion shows the detail that can accrue in studying a model organism, but it is not expected that tissues such as liver will be routinely available for analysis in exposure studies of human populations. Monitoring changes in blood-borne cells such as macrophages seems more realistic. Nevertheless, it is also expected that some changes in gene expression and metabolism in "inaccessible" organs will indirectly produce altered profiles in "accessible" tissues and fluids, and these in turn may make valuable contributions to the development of exposure indicators.

A potentially useful development is the proteomic analysis of biofluids that can be obtained in a noninvasive manner to identify biomarkers of exposure. An advantage of this approach is that certain biofluids (for example, bile, urine, nasal or bronchoalveolar lavage) are in direct contact with target tissues of interest, and changes in the proteomes of these fluids may closely reflect the tissue changes that result from exposure. Moreover, protein profiles of these biofluids may be less complex than the proteomic expression of tissues, which could make it easier to detect biomarkers. Lindahl and colleagues used 2D-PAGE analyses of bronchoalveolar and nasal lavage fluids to identify biomarkers of exposure to cigarette smoke or the lung irritant dimethylbenzylamine (Lindahl et al. 1995, 1999). These analyses identified changes in the levels of several proteins as well as changes in the distribution of apparently modified

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

protein forms. Most of these protein changes could be explained on the basis of known functions in airway inflammation and protection. However, these studies also identified a novel protein, palate lung nasal epithelium clone, which is modulated by airway irritants and specifically binds lipopolysaccharides, thus suggesting a role in innate immunity in airways (Weston et al. 1999).

"Shotgun" Proteome Analysis

Shotgun proteome analyses (described in Chapter 2) present an attractive alternative to 2D-PAGE because this approach allows more proteins to be analyzed. For example, Welch et al. (2005) used this approach to analyze proteome changes in mouse liver. They identified and quantified more than 100 proteins that changed more than 2-fold upon treatment with acetaminophen, almost 10 times the number of protein changes observed with 2D gel analyses. In another example, shotgun proteome analysis revealed proteome changes in the bile of rats given two hepatotoxicants that differ in the severity of the injury they produce (Jones et al. 2003). Although the observed changes clearly represented toxic injury, follow-up studies can help determine whether such changes can lead to specific biomarkers of exposure.

Protein Adducts

Some chemical exposures directly modify proteins (for example, through the formation of adducts), while others perturb endogenous posttranslational modifications. Although much previous work on exposure markers focused on specific adducts in the accessible proteins in blood (Skipper et al. 1994; Ehrenberg et al. 1996), protein adducts have not been measured in proteomes more complex than blood. The main problem, even in relatively simple proteomes, is limited sensitivity for detection of adducted proteins in the presence of excess unmodified protein. Analyses of protein posttranslational modifications often use methods that enrich the modified protein or peptide forms in the samples—for example, with specific antibodies or by chromatography methods. Unfortunately, analogous affinity tools for enrichment of xenobiotic-modified proteins are not generally available. The placement of immunoreactive or other tags on proteins in cell model systems provides an alternative means of enriching samples for analysis. For example, the addition of such small tags to ubiquitin and small ubiquitin-related modifier (SUMO) proteins enabled the capture of cellular proteins to which ubiquitin or SUMO had become attached as a result of environmental and chemical stresses (Manza et al. 2004; Zhou et al. 2004; Kirkpatrick et al. 2005). Although this approach is not directly applicable to human populations, cell culture models can indicate which exposure-responsive stress-responsive proteome changes might be analyzed by other means.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Metabolomics

Because diverse exposures sometimes produce similar changes in gene or protein expression, additional approaches may be needed to allow more precise identification of the exposure agent. As noted above, changes in gene and protein expression can alter metabolism in particular ways that can provide distinct signatures. Metabolomics involves measuring collections of small compounds in cells or biologic fluids, providing considerable biochemical detail in that the measured molecules include both metabolized products of environmental chemicals and endogenous metabolites. Endogenous metabolites may also be persistently changed by exposure, enabling chemical exposures to be detected for longer periods of time than the environmental compounds or their metabolites persist. The nuclear magnetic resonance- (NMR) and mass-spectrometry-based methods used for metabolomics in principle can specifically identify metabolites, but such identification may not be necessary if one is interested only in the use of metabolomic patterns as biomarkers.

Whereas combining information from metabolomics with other toxicogenomic data may turn out to be especially powerful for analyzing subtle or complex exposures (Griffin 2004), metabolomics is likely to offer several advantages over other toxicogenomic technologies (for example, transcriptomics and proteomics). Metabolomic analysis can be conducted on biofluids collected noninvasively (for example, urine, saliva), which would greatly facilitate sampling large populations of humans. Moreover, detection of continuing changes by metabolomic technologies may yield a timeline that allows extrapolation back to an estimate of the time of the exposure. Metabolomics may also have some advantages over proteomics because metabolism is often conserved across species—for example, the similarities and differences in the species-dependent pathophysiology and metabolomic profiles noted in mice and rats treated with hydrazine (Keun et al. 2004; Bollard et al. 2005).

Global metabolomic profiling might also be a simpler task than transcriptomic or proteomic profiling and thus more amenable to high-throughput screening. This was suggested by studies showing that the main yeast metabolome consists of fewer than 600 low-molecular-weight compounds (Oliver et al. 1998), but significant metabolomic analysis of exposure effects will be needed to address the possibility of high-throughput screening. Global metabolomic analysis in exposure studies in animals is just beginning and has not yet addressed the low-level exposures of interest in large populations. As noted above, metabolomic studies can readily be performed with noninvasive samples such as biofluids and breath condensate as well as on tissues in vivo. Profiles within a tissue or cell can be compared with profiles in biologic fluids or with cell secretion products to understand the metabolic consequences of xenobiotic-induced toxicity. Excellent examples of this approach are the studies by Waters and colleagues that used NMR and pattern recognition analysis to investigate time-related metabolic effects of α-naphthylisothiocyanate on liver, urine, and plasma in the rat (Waters et al. 2001, 2002). One way this has been investigated is with

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

high-resolution "magic angle spinning" NMR spectroscopy of intact tissue, which made it possible to link hepatic and renal histopathology to urinary and plasma metabolites (Garrod et al. 2001; Waters et al. 2001, 2002, 2005). Waters and colleagues found that hepatic lipidosis was associated with the increased urinary excretion of taurine and creatine. In addition, there was reduced urinary excretion of intermediates in the tricarboxylic acid cycle and increased excretion of plasma ketone bodies. These studies enabled a clearer understanding of key metabolic effects during development of and recovery from a toxic lesion. As with other approaches, however, development of this technology to generate biomarkers will require performing studies at subtoxic exposures—the low-level exposure of interest for assessing human exposures.

Larger scale work in metabolomics comes from the Consortium for Metabonomic Toxicology, which involves six pharmaceutical companies and the Imperial College of Science, Technology and Medicine, London, U.K. They are applying NMR-based metabolomic analysis of urine and blood serum for preclinical toxicologic screening and have completed studies of more than 80 candidate hepatic and renal toxicants (Lindon et al. 2003). This effort has yielded initial recommendations for standardization and reporting of metabolic analyses (Lindon et al. 2005a).

EVALUATION OF EXPOSURE IN RISK ASSESSMENT

The text to this point describes how toxicogenomic technologies may help improve exposure assessment, which is a critical element in risk assessment. However, it is possible that toxicogenomic technologies will be useful in risk assessment beyond directly improving exposure assessment. Because direct measurement of exposure to environmental contaminants (such as through biomonitoring of body fluids or personal or area monitors) is extremely rare, even in occupational settings, the U.S. Environmental Protection Agency (EPA) relies on exposure models that consider factors such as the quantities of water typically consumed in a day, respiratory rates, and activity patterns that would affect exposures. In the past, exposure analysis relied primarily on point estimates developed for illustrative populations (such as "reasonable worst-case exposures"). EPA now generally relies on probabilistic models (such as Monte Carlo methods) for exposure calculations that generate a distribution of population exposures. From this distribution, the risk assessor can choose a meaningful percentile value—such as the 90th, 95th, or 99th percentile—to assess risk to the population from a compound under evaluation.

Exposure assessment can also be modified by the availability of data on the mode of action (MOA) of toxicants, and toxicogenomic data may be able to inform this process. Typically, MOA approaches consider pharmacokinetic and pharmacodynamic variables that affect dose- and species-dependent responses. The emergence of physiologically based pharmacokinetic (PBPK) models has sometimes improved the understanding of dose as affected by kinetics (Meek et

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

al. 2003a). For those chemicals whose toxicity is mediated by rate-controlled activation or detoxification processes, information provided by PBPK models has enabled exploration of the relevance of high-dose and route-of-administration specific animal toxicity responses to potential human risks. This type of MOA data refinement is well illustrated with the example of chloroform, a water disinfection by-product, which produced liver tumors in mice when administered as boluses in high-dose oral gavage cancer bioassays but not when administered in comparable doses in the animals' drinking water. PBPK dosimetry assessments examining critical rates of chloroform metabolism to its toxic intermediate and subsequent detoxification demonstrated that the results of bolus gavage bioassays did not appropriately predict risks of chloroform carcinogenicity under conditions of long-term, low-level drinking water exposure. Information from these assessments resulted in significant modification of estimated human risks (Meek et al. 2003a).

PBPK models are not without their limitations. There can be considerable uncertainty using these models because their predictions depend on selection from the multiple modeling and parameter assumptions that are consistent with the data. Toxicogenomics may be able to improve such modeling by informing selection of assumptions and parameters. For example, the kinetic behavior and toxicity of chemicals often depend on metabolism, and toxicogenomic approaches can be used to rapidly identify potential key and rate-limiting enzyme targets whose activities, with further characterization, can be incorporated into PBPK models that better refine dose-dependent toxicity observations. Thus, toxicogenomic data provide mechanistic insights that can help refine the PBPK models used to predict target organ doses and hence provide a more accurate assessment of dose response.

SCIENTIFIC CHALLENGES

Establishing Toxicogenomic Profiles

The toxicogenomic approaches are distinguished by their ability to reveal patterns of change involving many individual molecules. The resolving power of such patterns, when they can be recognized, will likely be much greater than that provided by individual molecules. The key scientific challenge is to identify reliable patterns (signatures) that report specific exposures and their intensities. There will be significant statistical challenges in establishing criteria for recognizing transcriptomic, proteomic, and metabolomic signatures of exposure. One challenge is that human exposure to most toxicants occurs at rather low levels and the resulting signals might be obscured by greater noise than occurs in experimental models. Another important challenge will be to develop the means to detect exposures at relatively long times after they occur so that long-term health

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

effects might be assessed. Considerably more work is needed in transcriptomics, proteomics, and metabolomics to establish toxicogenomic signatures of exposure.

Coexposure Studies

Experimental studies typically have focused on assessing exposure to a single agent (Yang 1994), but, in real life, humans are exposed to combinations of substances. For example, chlorinated drinking water contains hundreds of chlorinated organic compounds. The air we breathe, particularly in urban areas, contains a suite of toxic ingredients, including combustion products, oxides of nitrogen and sulfur, ozone, particulates, and more, in addition to living organisms (bacteria, fungi, spores). The latter will surely produce their own transcriptional, proteomic, and metabolomic signatures, which must be distinguished from those generated by chemical and physical agents. Beyond this, dietary and pharmaceutical components are likely to add to the signals and variation among individual humans.

The committee was able to identify only two studies that had generated microarray data to evaluate multicomponent exposures. Although these studies focused on the biologic effects of the agents, it can readily be seen that the changes in both gene expression and DNA modification might contribute to exposure assessment.

Working with MCF-7 (human breast carcinoma) cells in culture, Mahadevan et al. (2005) compared the effects of a standard reference particulate material (SRM 1649a) alone or in combination with two well-studied carcinogenic compounds, benzo[*a*]pyrene (BP) and dibenzo[*a*,*l*] pyrene (DBP), on gene expression, metabolic activation, and the formation of DNA adducts. Such polycyclic aromatic hydrocarbons (PAHs) are already known to be involved in activation, detoxification, and DNA repair, and thus they can alter genomic integrity. Researchers focused on the impact on two PAH metabolism genes (*CYP1A1* and *CYP1B1*, encoding two cytochrome P450s), because the PAH compounds in the mixture are known to affect and be affected by these enzymes. Global analyses of the gene expression data revealed a clear additive induction of *CYP1A1* and *CYP1B1* upon cotreatment with SRM 1694a and BP as well as other effects. DBP had no effect alone or in combination with SRM 1649a. SRM 1649a decreased the total level of BP DNA adducts.

Bae et al. (2002) compared gene expression changes in human keratinocytes exposed to *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine (a DNA-alkylating carcinogen), arsenic, or a metal mixture containing arsenic, cadmium, chromium, and lead. Arsenic alone induced DNA-protective genes in the exposed cells (consistent with an anticarcinogenic effect). However, of the DNA repair genes activated in cells treated with arsenic alone, only *hNTH1* was induced in the cells treated with the arsenic-containing mixture. In fact, the

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

metals mixture actually suppressed expression of four DNA repair transcripts. Two metallothionein genes showed increased expression in the mixture compared with arsenic alone, perhaps because the mixture contained cadmium (Bae et al. 2002).

Human Genetic Variation

Genetic diversity across human populations constitutes its own challenge in the application of toxicogenomic approaches to exposure assessment. A hint of this problem can be seen even in the study of individual gene expression. One study sought to test whether DNA sequence differences in the promoter regions of the metallothionein IIA gene could be associated with differences in the inducibility of this defense protein in response to metals (Wu et al. 2000). Instead, the individual variation in zinc inducibility was so great that it precluded definitive identification of effects due to genetic variation in the promoter. Presumably, the overshadowing differences in response lay elsewhere—in the regulatory proteins governing the response or the cellular uptake of zinc. Determining whether agent-specific profiles can be recognized with toxicogenomic approaches will require a considerable amount of human study, perhaps beginning with workplace exposures, which may be high enough to produce substantial toxicogenomic signals.

There is also the issue of epigenetic variation—changes in gene expression that are due not to DNA sequence differences but rather to alterations such as DNA cytosine methylation, changes in chromatin structure, or effects arising from differences in the genes of mitochondria or their expression. It is already clear that such effects can contribute to disease development in humans (notably cancers), but little information is available to relate such effects to environmental exposure. One intriguing study in mice (Waterland and Jirtle 2004) showed that early nutritional differences could alter imprinting and hence expression of a specific gene. It is not hard to imagine that epigenetic variation, like genetic sources of variation, could overshadow differences in response attributable to a chemical exposure.

Human Studies

One study evaluated exposure to metallic fumes among welders by assessing transcriptomic profiles in whole blood (total RNA extracted) before and after acute exposures to metal fumes and in nonexposed controls (Z. Wang et al. 2005). A self-controlled study design, which involved taking measurements before and after individual exposure, was used to overcome the problems of large interindividual variations compared with the small changes caused by environmental exposure. The group was stratified according to smoking status (which profoundly affected the whole blood expression profiles), and nonsmokers exhibited altered gene expression in 35 genes from eight functional pathways, in-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

cluding processes related to oxidative stress, proinflammatory responses, phosphate metabolism, cell proliferation, and apoptosis.

Another study (Forrest et al. 2005) used microarrays to analyze transcriptomic profiles of blood lymphocytes of Chinese shoe factory workers chronically exposed to relatively high levels of benzene (mean = 47 parts per million) and matched nonexposed controls. This analysis identified 29 genes with highly altered expression in the benzene-exposed group. Real-time polymerase chain reaction confirmed changes in four of six selected genes. These transcripts may provide reliable molecular indicators of benzene exposure.

CONCLUSIONS

The ultimate goal of using toxicogenomic technologies as exposure indicators is to develop more sensitive, specific, and practically implementable exposure assessment methods. The application of toxicogenomic technologies to exposure assessment is new, and its use for human exposure assessment is in its infancy. A large gap exists between the current state of the art, with only a few studies to date having used gene expression analysis successfully to determine occupational exposure. Although high-sensitivity and data-rich toxicogenomic approaches may already be feasible for analyzing human exposure in some settings, applying these technologies to human populations exposed to low-level environmental contaminants will require considerably more development. Major challenges include the following:

- A lack of suitable information to define what constitutes a transcriptomic, proteomic, or metabolomic profile as a signature of specific toxicant exposure. Especially important with regard to "real-world" exposures is the question of sensitivity: what are the minimum exposures that such signatures can detect?
- Determining which accessible fluids and tissues display changes related to low-level exposures that will typify the human condition.
- A lack of information on how exposure to multiple agents affects toxicogenomic signatures and minimum thresholds that can be detected.
- A lack of information on how the time interval after an exposure affects the ability to recognize a toxicogenomic signature. Some long-term changes in gene expression may occur, but this has not been well studied.

These challenges should be addressed in research on the application of transcriptomic, proteomic, and metabolomic profiling to exposure assessment. In this context, the integration of information from multiple toxicogenomic approaches may provide considerably more analytical power than any one approach alone. On the question of sensitivity, a useful beginning would be a focus on occupational exposures, in which the toxicants are often known and the levels are higher than exposures to the population in general.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

RECOMMENDATIONS

Short Term

1. Use transcriptomic, proteomic, and metabolomic technologies to identify signatures of environmental exposures in target and surrogate tissues and fluids, primarily with animal models.
2. Begin testing complex mixtures for possible identification of distinct exposure signatures.
3. Examine the time course of chemical versus toxicogenomic signature persistence after initial chemical exposures.

Intermediate

4. To enable the further development of toxicogenomic measures of exposure, include transcriptomic, metabolomic, and/or proteomic analysis of samples in large human population studies and studies designed to assess exposures at toxicant levels commonly encountered in the workplace and certain communities1. This would be especially useful for chemicals whose toxicity across a range of exposures is well established. Use these studies to begin addressing issues of interindividual variability, background noise, confounding effects of combined exposures, the ability of toxicogenomic approaches to report exposures quantitatively, and determining ranges of quantifiable responses.
5. To enable the further development of toxicogenomic measures of exposure, include toxicogenomic analysis of samples in relevant case-control, cohort and panel studies that involve repeated measurements over time, as well as in clinical trials when possible and appropriate.
6. Use the information collected from studies to help develop and populate a database that will support further development of toxicogenomic exposure assessment.

Long Term

7. Adapt and apply toxicogenomic measurements to assess exposure by developing signatures of exposure to single compounds and complex mixtures that can be used in animal and human population studies.

---

[1]See issues raised about the protection of humans in Chapter 11.

Copyright © National Academy of Sciences. All rights reserved.

5

Application to Hazard Screening

This chapter addresses the application of toxicogenomics to screening chemical compounds for *hazard*, or the ability to cause harm. A screening test can be defined as one designed to detect a state or property more quickly and cheaply than more elaborate tests for that state or property. In predictive toxicology, the property being detected by screening tests is generally hazard. Screening tests may not give complete information on toxicity, such as the time course, chronic effects, or dose-response characteristics. Therefore, in the context of this chapter, screening data provide an input to the hazard identification step in risk assessment but do not allow full determination of risk.[1]

The chemical and pharmaceutical industries also use screening tests to detect desirable properties, such as the ability to bind to specific target receptors. The use of genomic techniques to screen for desirable, pharmacologic properties may be analogous to hazard screening (Lum et al. 2004) but is not the principal focus of this chapter. Toxicogenomic technologies may be incorporated directly into existing, more traditional hazard screening tests; they may be the basis of new tests that substitute for more traditional tests; or they may generate mechanistic insights that enable more basic tests to be conducted for screening compounds, such as receptor binding or other physicochemical assays. This chapter primarily discusses the first two applications of toxicogenomic technologies.

DESCRIPTION AND PROPERTIES OF CURRENT HAZARD SCREENING METHODS

Hazard screening can be comprehensive, intended to detect all potential hazards, or it can be more limited, detecting only a specific type of hazard. A

[1]Hazard identification is one of four elements of a full risk assessment as described in the National Research Council report (NRC 1983). The other three elements are exposure assessment, dose-response assessment, and risk characterization.

Copyright © National Academy of Sciences. All rights reserved.

comprehensive hazard assessment for a chemical substance generally requires a variety of in vitro and in vivo toxicologic assays as well as evaluations of physical properties.

The selection of individual screening tests depends greatly on the setting and specific regulatory requirements. For example, the current practice of the U.S. Environmental Protection Agency (EPA) under the Toxic Substances Control Act (TSCA), in the absence of more extensive preexisting data, is to screen new chemicals based solely on physicochemical data using quantitative structure-activity relationship models. In this setting, chemical tests may be limited to properties such as boiling point, octanol-water partition coefficient, vapor pressure, and solubility. If environmental fate and transport of substances are not primary concerns, short-term in vivo rodent assays may be used, such as a 28-day feeding study, which examines histopathology in most critical target organs. More comprehensive screening programs have adopted batteries of tests that provide information on different types of toxicity but remain insufficient to fully assess chemical risks. As one example, the Organization of Economic Cooperation and Development (OECD) has developed the Screening Information Data Set (SIDS), which consists of the 21 data elements shown in Table 5-1. Each toxicity test involves administering a measured amount of a compound to whole organisms or to cells in culture and then measuring indicators of toxic outcomes.

Compared with more extensive tests, screening tests tend to use higher and fewer doses of the compound being studied, fewer test subjects, a shorter time period of observation, and less extensive evaluation of the toxic outcomes. To reduce the use of mammals for laboratory testing, there is a strong impetus to develop and validate screening tests that use cultured cells or lower order animals, such as worms.

The incorporation of toxicogenomics into screening tests involves measuring gene, protein, or metabolite changes in response to specific doses of an administered test compound at specific time points, with or without the parallel measurement of more traditional markers of toxicity. The critical question about new toxicogenomic techniques is whether they can improve hazard screening by making tests faster, more comprehensive, less reliant on higher order animals, and more predictive and accurate without being prohibitively expensive.

For a screening test to be useful, it must be capable of detecting the property or state being tested when it truly exists. This is the definition of the "sensitivity" of a screening test. In many cases, screening tests are designed to be highly sensitive, sometimes at the expense of the specificity of the test or the ability of the test to return a negative result when the property or state of concern does not exist. Another way to describe this quality is that hazard screening tests often accept a higher rate of false-positive results to avoid not detecting a hazard because of a high rate of false-negative results.

When the data generated by screening tests are continuous, as is the case with gene and protein expression and metabolite assays, the selection of thresholds for positive and negative results plays a dominant role in determining the

Copyright © National Academy of Sciences. All rights reserved.

TABLE 5-1 Elements of the OECD SIDS

Data Elements	Comments
Physical-chemical properties	
Melting point	
Boiling point	
Relative density	Required for inorganic chemicals and should be provided if readily available for organic chemicals.
Vapor pressure	
Partition coefficient: *n*-octanol/water	
Water solubility	
Dissociation constant	For substances normally capable of dissociation.
Oxidation-reduction potential	Required for inorganic chemicals; may be required for certain organic chemicals.
Environmental fate	
Photodegradation	
Stability in water	Not required for classes of chemicals whose molecular structure does not possess functional groups subject to hydrolysis or that are generally recognized to be resistant to hydrolysis. In these cases, a qualitative statement can be provided.
Transport and distribution between environmental compartments including distribution pathways	Including Henry's law constant, aerosolization, volatilization, soil adsorption, and desorption, based on experimental data or, if not available or appropriate, calculated using structure-activity relationships.
Aerobic biodegradability	
Environmental toxicology	
Acute toxicity to fish	
Acute toxicity to daphnia	
Toxicity to algae	
Chronic toxicity	Necessity determined based on physical-chemical properties of the chemical. Any new data required should be collected using the most sensitive species (fish, daphnia, or algae) within limitations of the chemical properties.
Terrestrial toxicity	The need for testing will normally be addressed at the post-SIDS stage. However, if significant exposure is expected or identified in the terrestrial environment (soil), appropriate terrestrial toxicity tests should be considered at the SIDS level. Taking into account animal

(Continued on next page)

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

TABLE 5-1 Continued

Data Elements	Comments
	welfare considerations, the need for avian toxicity testing should be considered only at the post-SIDS stage.
Mammalian toxicology	
Acute toxicity	By oral route, dermal route, or inhalation; required only on the most relevant route of exposure
Repeated dose toxicity	The protocol for new studies should specify the use of the most relevant route of exposure.
Genetic toxicity	Two end points required, generally point mutation and chromosomal aberrations.
Reproductive toxicity	Requires data to assess fertility and developmental toxicity.
Experience with human exposure	If available.

Source: Adapted from OECD 2006. Reprinted with permission; 2006, Organisation for Economic Co-operation and Development.

sensitivity and specificity of the test. When larger values of the test are more likely to indicate the presence of a particular hazard, selection of a relatively low value as the threshold for a positive result will lead to greater sensitivity and lower specificity (that is, fewer false-negatives and more false-positives). Conversely, a high threshold for a positive result will lead to lower sensitivity and higher specificity (that is, more false-negatives and fewer false-positives). A critical challenge in designing and validating toxicogenomic screening tests is to identify and define a "gold standard" for hazard screening—the indication of the true state of toxicity against which the sensitivity and specificity of the screening test can be measured.

IN VITRO VERSUS IN VIVO TESTING

Current screening tests are done with whole animals in vivo and with tissue slices, cultured cells, or artificial biologic systems, including individual receptor molecules, in vitro. Because toxicogenomic technologies measure gene, protein, and metabolite responses, they require whole animals, intact tissues, or cell cultures. Other in vitro tests with receptor binding or other molecular end points may be used in combination with toxicogenomic-based tests but are not discussed in detail in this chapter. In vitro tests provide many benefits in a screening setting but also pose serious challenges. They are generally less expensive, more readily automated, and therefore faster than in vivo testing. However, questions about their relevance to in vivo toxicity limit their use in decision-making processes without additional supporting data (Boess et al. 2003; Jessen et al. 2003). Toxicogenomic technologies may offer in vitro methods for

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

answering questions that currently require in vivo tests, when results can be extrapolated from one species to another. In the following discussion, current practices and published studies describe the state of the art with respect to the development of toxicogenomic applications for both in vitro and in vivo screening tests.

DEVELOPMENT OF PREDICTIVE ALGORITHMS FOR SCREENING PURPOSES

Development of screening tools involves both identifying signatures that can distinguish between various toxic and nontoxic end points and developing predictive algorithms that can use these data effectively to make predictions. Both represent significant challenges, particularly given that most published studies involve far fewer samples than the number of measurements on each sample (the "n < p" problem discussed in Chapter 3) and the well-known problem of multiple testing. Given the broad interest in developing toxicogenomic predictors, these remain areas of active research in statistics, bioinformatics, and computer science. Although several studies have compared various combinations of predictive algorithms (Natsoulis et al. 2005; Van Delft et al. 2005), no single method has emerged as being superior for all applications. This result is not surprising given the work of Wolpert and McReady, who present a series of "no-free-lunch" theorems showing that, although a particular algorithm might be optimal for one class of problems, it might be greatly surpassed by another algorithm when applied to a slightly different class (Wolpert and MacReady 1997). As most toxicogenomic studies presented to date have looked at small numbers of samples and have not arrived at optimal solutions for predictive classification, these results suggest that significant additional work is necessary to fully address these issues.

STATE OF THE ART

Current Practices for Pharmaceutical Hazard Screening

With the advent of combinatorial chemistry, pharmaceutical companies are now able to consider thousands of drug candidates at once in the drug development process. Drug companies are seeking a variety of in vitro and in silico methods to screen drug candidates for effectiveness and toxicity before more expensive, and slower, in vivo testing (Johnson and Wolfgang 2000; Ulrich and Friend 2002; van de Waterbeemd and Gifford 2003; Suter et al. 2004; Butcher 2005). Toxicogenomic technologies are being explored in combination with more traditional in vitro cytotoxicity assays to improve the information obtained from this more rapid screening. The goals are to reduce attrition of compounds during more costly phases of drug development (that is, Phase 2 and Phase 3

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

clinical trials) by identifying potential adverse drug reactions and other possible hazards at much earlier stages.

One study has reported that current use of toxicogenomic technologies for screening drug candidates varied widely among leading pharmaceutical companies (Branca and Peck 2004). Of the 12 companies surveyed, 7 had made or were planning to soon make extensive use of toxicogenomics for drug candidate screening, and the other 5 limited the use of toxicogenomics to analysis of failed candidates or reference compounds. Most companies applied toxicogenomic technologies in the discovery and preclinical phases, and, whereas most preferred using in vivo assays, several said they were developing in vitro screening tests. Among the principal barriers to broader use and acceptance of toxicogenomic technologies in drug screening were uncertainty about biologic relevance, resource limitations, and regulatory consequences.

In addition to major pharmaceutical companies, smaller biotechnology companies specializing in toxicogenomic technologies provide screening services on a contractual basis. Two such companies, Gene Logic and Iconix, purport to identify compounds that are most likely to demonstrate unacceptable toxicity based on a molecular profile evaluated by using their proprietary toxicology signature models. These companies derived their signature models from analyzing databases containing hundreds or thousands of profiles from reference compounds tested in both in vivo and in vitro systems.

Gene Logic uses short-term studies of investigative compounds and their proprietary models to provide predictive profiles of liver, kidney, and heart toxicity (Gene Logic Inc. 2005). Iconix conducts short-term studies on investigative compounds using up to five time points and at least two doses to sample 12 different tissues in laboratory animals. Gene expression data are combined with an assortment of in vitro receptor binding and enzyme tests on the investigative compounds and the results are compared with Iconix's proprietary database of gene expression, molecular pharmacology, literature review, and histopathology to predict potential toxicities (Iconix Pharmaceuticals 2005). Both companies highlight the ability of gene expression arrays to provide mechanistic insight into any potential toxicities that are identified. Neither company claims the ability to predict outcomes of chronic toxicity, such as cancer or neurodegenerative diseases.

The Japanese government, through its Ministry of Health, Labour and Welfare and National Institute of Health Sciences (NIHS), has teamed up with a consortium of 17 pharmaceutical companies on the project "Construction of a Forecasting System for Drug Safety Based on the Toxicogenomics Technique and Related Basic Studies." The long-term goals of this 5 billion yen (approximately U.S. $43 million) project are to facilitate drug development and improve toxicity prediction by linking this database with others, including one being developed at NIHS for environmental chemicals (Urushidani and Nagao 2005). For each of the initial 150 compounds investigated, the following tests are conducted: toxicogenomic, biochemical, and histopathologic studies of acutely and

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

repeatedly dosed rats; in vitro tests on rat hepatocytes; and in vitro tests on human hepatocytes.

Current Practices for Environmental Chemicals

Whereas pharmaceutical companies are clearly conducting toxicogenomic-based screening of compounds, there is little documentation of such screening going on in the chemical industry. Publicly available information on the application of toxicogenomic technologies for screening environmental chemicals has come from government institutions rather than from industry.

EPA and NIHS are conducting two projects of interest. Through a contract with Iconix, EPA is assessing the usefulness of gene expression signatures for categorizing environmental chemicals according to hazard. The Iconix contract involves five chemicals and to date, because of funding constraints, has assessed only hepatic toxicity (R. Kavlock, EPA, personal communication, 2006). EPA scientists are investigating how to adjust the statistical thresholds to optimize sensitivity and specificity for environmental chemicals as opposed to drug candidates, upon which the Iconix system was originally designed (R. Kavlock, EPA, personal communication, 2006). The Japanese NIHS is creating a database of mouse gene expression array data corresponding to four doses and four time points for roughly 100 environmental and industrial chemicals. The platform for this data system, which they have dubbed "millefeuille" (French for thousand leaves), is based on Affymetrix microarrays. Data obtained are normalized on a per cell basis (Kanno et al. 2005; Urushudani and Nagao 2005).

Proof-of-Concept Studies from the Published Literature

Although some of the efforts to incorporate toxicogenomic technologies into hazard screens suggest the development of robust predictive models, the published literature does not yet provide details on these systems. Early proof-of-concept studies established that gene arrays can be used to classify toxicants with different mechanisms (Thomas et al. 2001; Waring et al. 2001; Hamadeh et al. 2002a,b). More recent studies, discussed below, apply data classification algorithms to gene expression data to predict the toxicity of unknown compounds or to demonstrate the use of proteomic methods and in vitro models to classify toxicants.

Steiner at al. (2004) at Hoffman La Roche Ltd. reported an early attempt at predicting the hazard of unknown compounds. Their study used a class prediction approach with support vector machines (SVMs), a class of supervised learning algorithms that recognize informative patterns within an input dataset and then classify previously unseen samples. The training set consisted of microarray gene expression data from 28 known hepatotoxicants and 3 compounds with no hepatotoxicity. The data were correlated with histopathology findings and a variety of clinical chemistry markers associated with liver toxicity. To

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

reduce the number of genes analyzed by the SVM algorithms, the authors used a recursive feature elimination approach that excluded genes that did not contribute significantly to the classification of compounds as hepatotoxicants. This resulted in a much smaller set of informative genes being identified. This approach correctly predicted 63/63 vehicle-treated animals as belonging to the untreated control group. Although this may reflect overfitting of the parameters in the classifier, it is worth noting that approximately 90% of test compounds with known hepatotoxicity, but not previously tested by the algorithm, were correctly predicted to have toxic effects on the liver, suggesting that the approach has identified a hepatotoxicity signature. Although there were no false-positive predictions of toxicity, there was clearly a significant (10%) false-negative classification of toxicants. These may be compounds with mechanisms of action not represented in the original training data, but it does suggest that additional work is necessary before these results can be broadly applied.

Steiner et al. (2004) next asked if the SVM approach could be used to predict the subtype of liver histopathology that a test compound would induce. They first classified compounds in the training set based on their mechanisms of toxicity: direct-acting hepatotoxicants, steatosis inducers, cholestasis inducers, or PPAR-α agonists (peroxisome proliferators). They based classification on the most prevalent type of histopathology induced at specified doses. They then generated five different SVM algorithms to distinguish among the controls and treatment classes. Not surprisingly, the number of genes required, as well as the accuracy of prediction, differed for each type of toxicity; in some cases, the same genes were included in classifying the different types of toxicity. The authors evaluated data from each new microarray experiment with each algorithm and predicted the mechanism of toxicity with a calculated discriminant value. This approach did not misclassify any of the controls. Although only three test compounds were analyzed, the results attested to the ability of the algorithm to accurately classify hepatotoxicants on the basis of their mechanism of action. Moreover, this approach seemed to be capable of discerning compounds with mixed modes of action.

Scientists from Hoffman-LaRoche (Suter et al. 2003; Ruepp et al. 2005) recently described another example of an early proof-of-concept study. The study provides a rare insight into the value of toxicogenomic approaches when investigators have access to a densely populated, high-quality database—in this case, the Roche toxicogenomic database, which classifies compounds into subcategories based on their histopathologic manifestations.

This study used a retrospective approach to determine whether gene expression profiling could have predicted the toxicity of a failed candidate compound before it was evaluated by more time-consuming methodologies of classic toxicology. The compound in question was Ro-Cmp-A, an antagonist of the brain-specific, 5-hydroxytryptamine (5-HT_6) receptor with predicted value in improving memory deficit in patients with Alzheimer's disease.

Toxicologic studies in rats and dogs found significant hepatotoxicity, characterized by steatosis, despite the absence of the 5-HT_6 receptor in liver

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

cells. The gene expression profiles induced in rat liver in vivo by Ro-Comp-A and a nonhepatotoxic homolog, Ro-Cmp-B, were determined as a function of dose and time after exposure.

The authors then compared the profiles with those in the reference gene expression database. Gene expression profiles induced at subchronic doses administered over 7 days made it possible to identify Ro-Cmp-A as a steatotic hepatotoxicant. Remarkably, gene expression profiles induced by acutely toxic doses also identified Ro-Cmp-A as a steatotic hepatotoxicant as early as 24 hours after exposure, despite a lack of characteristic histopathology. If specific marker genes with mechanistic links to toxicity were selected, steatotic activity was predicted by expression profiling as soon as 6 hours after an acutely toxic dose. In contrast, the nonhepatotoxic analog Ro-Comp-B induced gene expression patterns in rat liver that more closely resembled those of the untreated control animals. In addition to the in vivo studies, exposure of rat hepatocytes to Ro-Cmp-A in vitro at doses that also did not induce significant toxicity induced a subset of the predictive genes.

This example speaks to several of the issues related to the initial use of toxicogenomic data for screening or hazard identification, at least for application to pharmaceuticals. First and foremost is the necessity for access to a high-quality database that is densely populated with high-quality toxicogenomic data that represent many compounds and that are phenotypically anchored (associated with more traditional indicators of clinical pathology). This was a key factor in classification of the candidate compound as a steatotic agent across doses and exposure times. In fact, the authors indicated that, when using this database, essentially the same set of differentially expressed genes could be used for classification regardless of the type of statistical or computational approaches used to mine the data. Another significant finding of this study was that a subset of the expression profile that predicted steatosis could be replicated in a short term in vitro cell-based assay, at least at carefully selected doses.

Another example of a toxicogenomic screen for a specific mode of action also used a short-term in vitro cell-based assay (Sawada et al. 2005). The authors correlated electron microscopic findings of phospholipidosis at the cellular level (development of lamellar myelin-like bodies in lysosomes) with changes in gene expression measured with microarrays to define a subset of genes that were predictive of phospholipidosis. They developed an mRNA assay based on polymerase chain reaction (PCR) with this subset of genes and demonstrated its predictive ability with a different set of compounds with a known tendency to cause phospholipidosis. Whether this screening algorithm is generalizable to other model systems remains to be demonstrated. In addition, the use of a PCR-based assay for screening new compounds (after using microarrays to select the genes) makes it more expensive than other testing methods to detect phospholipidosis. The authors noted that converting the PCR-based assay to a gene-chip-based assay would be likely to lower costs and improve throughput for use as a screen.

Another example of toxicogenomic hazard screening involving mechanistic information is the effort to develop profiles to identify specific carcinogenic

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

modes of action. Toxicogenomic profiles for cancer mode of action would also have profound implications for risk assessment methodology. In an in vivo short-term study, Ellinger-Zeiglebauer et al. (2004) tested the hypothesis that genotoxic liver carcinogens deregulate a common set of genes and that those deregulated genes represent defined biologic pathways implicated in early events in tumorigenesis. Although neither a single gene nor pathway sufficiently discriminated genotoxic from non-genotoxic carcinogens, their findings suggested that, with further understanding of mechanisms of carcinogenesis as well as further development of data and analytic tools, combinations of pathway-associated gene expression profiles may ultimately be able to predict genotoxic or nongenotoxic carcinogenic potential of compounds in short-term studies (Ellinger-Ziegelbauer et al. 2004). Among nongenotoxic carcinogens, however, it may be difficult to generalize responses to derive simple profiles indicative of nongenotoxic mechanisms. Iida and coworkers demonstrated that mouse liver tumor transcriptional response to nongenotoxic carcinogens involved more differences than similarities in changes in early gene expression (Iida et al. 2003). Further analyses revealed that early gene expression changes appeared to be carcinogen specific and involved apoptosis and cell-cycle-related genes. Thus, the ability of early changes in gene expression to predict carcinogenesis requires further evaluation.

Much of the published literature has focused on hepatotoxicity as an end point, because it is one of the best characterized adverse effects and is frequently responsible for candidate drug failure. However, one recent study from Iconix (Fielden et al. 2005) describes the development of a screening algorithm for nephrotoxicity. The authors used a training set of 64 compounds (15 known renal tubular toxicants and 49 non-nephrotoxicants, with gene expression measured after 5 days of treatment) and a sparse linear programming algorithm (SPLP) to derive a set of 35 gene signatures for detecting preclinical nephrotoxicity. The SPLP algorithm is similar to the support vector machine approach described earlier, but it produces short lists of discriminant genes to allow more intuitive understanding of the mechanistic roles of individual genes in the signature. These 35 gene signatures were then tested on 21 compounds whose nephrotoxicity was already characterized but that were not in the original training set. Seven of the 9 (78%) known tubular toxicants and 9 of the 12 (75%) known non-nephrotoxicants were correctly identified. Although this moderate degree of sensitivity and specificity is not ideal for a screening test, the authors suggest that the test provides a promising alternative to 28-day in vivo studies with histopathology, which are the only tests now available to screen candidate drugs for nephrotoxicity.

Another nonhepatotoxicity example assessed the utility of transcriptome profiling in screening for endocrine disruptors (Daston and Naciff 2005). The investigators examined gene expression profiles in the uterus and ovaries of fetal and prepubertal female rats. They administered three prototypical estrogenic compounds—17-α-ethynylestradiol, genistein, and bisphenol A—at several doses. Transcriptome profiles indicated a similar pattern common to all three

Copyright © National Academy of Sciences. All rights reserved.

estrogenic compounds. Moreover, there was overlap in the responses in hormone-sensitive organs, including genes known to be regulated by estrogen.

Most efforts to apply toxicogenomic techniques to hazard screening have used gene expression (transcriptome) profiling, but a small number of studies have demonstrated the potential usefulness of proteomic and metabonomic studies to classify putative drugs or other compounds into categories of toxicity or biologic activity. Fella et al. (2005) reported the use of proteomic techniques to discover early markers of liver cancer in a 25-week study of Wistar rats exposed to *N*-nitrosomorpholine. The authors retrospectively identified several proteins characteristic of active cancers that were also deregulated at 3 weeks of exposure, suggesting that they might prove to be efficient indicators of liver carcinogenesis in shorter studies than standard cancer bioassays.

Researchers at Pfizer (Kikkawa et al. 2005; Yamamoto et al. 2005) added proteomic analyses to in vitro cell culture assays in two proof-of-concept studies using the hepatotoxicants acetaminophen, amiodarone, and tetracycline. In the first study, they identified 31 candidate marker proteins together with more standard measures of hepatotoxicity, such as lactate dehydrogenase release and altered mitochondrial respiration. They concluded that proteomic methods were superior to lactate dehydrogenase release at detecting toxic changes at earlier time points. In the second study, the authors identified three markers of oxidative stress that might be more reliable and earlier indicators of toxicity than measures of secondary cell membrane damage. These studies do not demonstrate the performance of these in vitro proteomic methods for actual screening purposes, but they do suggest that in vitro proteomic methods have potential for classifying responses to toxic drug candidates and chemicals.

The Consortium for Metabonomic Toxicology published a proof-of-concept paper demonstrating the use of metabonomic profiles for drug and chemical screening (Lindon et al. 2005b). The authors reported using a set of 147 compounds to develop an expert system for predicting toxicity based on nuclear magnetic resonance (NMR) spectra from urine specimens. The system correctly predicted histopathologic organ changes in approximately 50% of cases; in 5% of cases, the system made incorrect predictions. The remaining samples did not display conclusive histopathologic changes, metabolites obscured critical portions of the NMR spectrum, or overlapping patterns of toxicity precluded accurate prediction.

POTENTIAL APPLICATIONS

Use of Toxicogenomics in Existing Chemical Screening Programs

Within larger chemical screening initiatives, toxicogenomic technologies could enhance screening-level determinations of basic mode of action. Two such applications involve the High Production Volume (HPV) Chemical Challenge

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Program and the initial evaluation of premanufacturing notices (PMNs) for new chemicals under the TSCA.

Categorizing HPV Chemicals

The EPA HPV Chemical Challenge Program is designed to make basic toxicity data publicly available on approximately 2,800 HPV chemicals or chemicals manufactured in quantities greater than 1 million pounds. The program began in 1998 as a result of a collaborative effort between the EPA Office of Pollution Prevention and Toxics, Environmental Defense, and the American Chemistry Council. To date, chemical manufacturers and importers have committed to providing data for more than 2,200 chemicals to fill in any gaps in the SIDS battery (for further details, see EPA 2006a). More than 400 companies and 100 consortia have sponsored 1,371 chemicals directly in the program. An additional 851 chemicals have been sponsored indirectly in an international counterpart to the HPV Challenge Program, the International Council of Chemical Associations HPV Initiative. Since the inception of the HPV Challenge Program, 365 robust summaries have provided data on approximately 1,328 chemicals.

To obtain information on as many high production chemicals as possible, the HPV program allows chemicals to be grouped into "categories." In theory, chemicals within a category share common modes of action, allowing testing results to be extrapolated from a few representative members of the category for various end points to the other members of the category. At present, there are few objective data to justify category classification. In vivo studies of acute toxicity and 28-day repeat dosing are the most common traditional toxicology tests used to assert the existence of categories in the HPV program, but these tests are done on only a small subset of chemicals in the category. Submitting all members of a category to a short-term study using transcriptome, proteome, or metabonome analysis could allow more rigorous confirmation of categories and provide a mechanism to identify potential outliers. However, assays limited in dose range and time course could not be expected to exclude the possibility of differential toxicity within a category and toxicogenomic data would not be expected to be the sole source of information for category validation. As databases grow and the quality of the interpretative methods improves, gene and protein assays may ultimately support quantitative interpolation within categories as well as provide mechanistic insight to inform further testing protocols.

Enhancing New Chemical Evaluations Under TSCA

Under TSCA, EPA requires chemical companies planning to bring a new product to market to file a PMN. The EPA then has 90 days to determine whether to allow the chemical to proceed to commercialization or to require additional testing or limitations on production based on evidence of toxicity (for additional information, see EPA 2006b). Because TSCA does not require any

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

specific toxicity tests be performed before PMN submission, 85% of PMNs contain no health data and EPA must make a determination based on consideration of structure-activity relationships in most cases.

Categorization based on chemical structures eventually could be combined with transcriptome or proteome profiling. PMNs might then contain toxicogenomic data that show different, nontoxic profiles compared with other chemicals of similar structures. Alternatively, EPA scientists may request rapid toxicogenomic assays of new chemicals and use the profiles generated to make more rigorous determinations of safety. Additional database development and demonstration of predictive accuracy would be required before true toxicogenomic screens would be useful in this setting.

Use of Toxicogenomics to Screen for Hazard in Future Chemical Programs

Current toxicogenomic assays are not sufficiently developed to replace more traditional toxicologic screening tests, but there are important opportunities to build the familiarity and databases needed to inform future toxicogenomic screening regimens by adding toxicogenomic tests to existing chemical programs. For example, The National Toxicology Program uses established methodologies to test relatively large numbers of animals, which is both costly and time-consuming. However, the chronic exposure experiments provide a rare source of biologic material that could be used to advance toxicogenomic applications. Strategically adding appropriate toxicogenomic assays to these testing protocols will generate novel, phenotypically correlated data that would be especially useful in building databases to aid in predictive toxicology. Such data will be suited for helping determine whether short-term in vivo studies using toxicogenomic technologies can substitute for chronic bioassays.

Similar opportunities exist outside the United States. Within OECD screening programs, toxicogenomic assays could be added to the basic SIDS dataset. In the European Union, a new approach to chemical regulation known by the acronym REACH (Registration Evaluation and Authorization of Chemicals) will require hazard data generation as a condition of continued marketing of chemicals, with the amount of data dependent on the volume of the chemical produced. Although the program relies on traditional toxicology assays, the increase in data-generation activities expected under REACH presents another opportunity to expand the datasets and acquire more experience in large-scale applications of toxicogenomic assays.

Future Use in the Pharmaceutical Industry

Whether the pharmaceutical industry will expand the use of toxicogenomic technologies to detect potential toxicity in lead candidates is an open question. For in vivo studies, toxicogenomics may allow ongoing noninvasive sampling (especially for proteomic or metabonomic end points) and provide

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

useful mechanistic information. For certain end points (as demonstrated currently with different types of hepatotoxicity), reliance on toxicogenomic data to predict potential hazard will be possible.

Several companies appear to be working on using in vitro toxicogenomic assays to evaluate multiple cellular effects. Surrogate screens using panels of genes in in vitro systems may be deployed rather cheaply and produce large amounts of data quickly. The critical question for the industry is whether adding toxicogenomic technologies to more standard toxicology testing or other in vitro methods will result in overall cost savings. To date, the economic value of toxicogenomic testing in drug development has not been demonstrated through rigorous studies. In the absence of such data, it is likely that pharmaceutical companies will make individual judgments about incorporating toxicogenomic assays. Regardless of the degree to which pharmaceutical companies use toxicogenomic screens to identify possible hazards, for the foreseeable future full interpretation of risk will be done in the context of additional, required testing information.

CHALLENGES AND LIMITATIONS

For toxicogenomic technologies to be widely applied to screen compounds for toxicity, it will be essential to demonstrate that toxicogenomic-based screens are sufficiently robust and comprehensive for a wide range of compounds and toxic end points. There are several aspects of this generalizability:

- Because the screening of environmental chemicals has different requirements than the screening of candidate drugs, algorithms designed for screening candidate drugs will need to be modified and validated for environmental chemicals.
- The database of toxicogenomic and traditional toxicologic data used for training and validating screening algorithms must be extensive and comprehensive enough to allow accurate predictions for a wide range of chemicals.
- Costs of toxicogenomic assays must be compatible with implementation in screening.
- Throughput of toxicogenomic assays must meet or exceed norms for currently used screens.

Each of these factors is described in more detail below.

Contrasts Between Screening Needs and Methods for Pharmaceutical Candidates and Industrial Chemicals

Pharmaceutical companies and chemical companies face different regulatory and business requirements that lead to different requirements in screening their products. Drug companies need rapid, inexpensive ways to screen candi-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

date compounds for toxicity that avoid expenditures on compounds destined to fail at later stages of development. Chemical companies need rapid, inexpensive ways to screen new chemicals and many existing chemicals for toxicity. In both cases, the ability to gain useful information from short-term tests makes toxicogenomic techniques attractive for screening applications.

Although the screening needs of pharmaceutical and chemical companies may appear similar, the technical requirements of screening in these two settings are quite different. These differences include dose ranges that must be considered, types of toxicity that must be evaluated, and the degree of sensitivity and specificity required. Evaluation of drug toxicity focuses on therapeutic dose ranges. In contrast, exposures to industrial chemicals can range over many orders of magnitude depending on the setting, requiring toxicologic evaluation over a broader dose range.

A second major difference between screening candidate drugs and screening chemicals is that candidate drugs undergo a rigorous set of in vivo tests before they are marketed. In contrast, screening tests are likely to be the only sources of new data generated in the assessment of toxicity that occurs for most industrial chemicals before widespread marketing. The redundancy of animal testing after toxicogenomic prescreening for candidate drugs means that the prescreening test can have relatively low sensitivity and still be valuable. Because the animal tests are believed to be both more sensitive and specific for toxicity, the toxicogenomic screening tests could have a relatively high false-negative rate and still identify enough potentially toxic candidate drugs (thus avoiding more costly animal testing) to be economically useful. Importantly, the animal tests would serve as a backup for identifying those toxic candidate drugs that have been missed by the toxicogenomic screens. In contrast, the absence of any statutory requirement for animal testing for industrial chemicals other than pesticides in the United States means that there is no backstop to the screening tests.

A third major difference is the breadth needed. A set of toxicogenomic screens limited to liver toxicity would be of relatively little value, as liver toxicity is not a primary outcome for exposures to most industrial chemicals. Toxicogenomic screens for industrial chemicals would need to assess a broader array of toxicities, including cancer, developmental and reproductive effects, neurotoxicity, and immunotoxicity.

Whereas any reliable screening information is of great public health value, care must be taken not to assume that toxicogenomic screens for limited types of toxicity are sufficient to conclude that a screened chemical has no significant hazard associated with it. Ideally, for the screening tests to fully protect public health, they should be highly sensitive, detecting the vast majority of chemicals with potential toxicity. Moreover, they should be able to predict not just acute toxicity but also chronic toxicity; they should also be able to predict toxicity at different developmental stages, because environmental exposures occur throughout the life span. This requires new knowledge of early gene, protein, and metabolite markers of chronic effects so that changes indicative of short-term acute toxicity can be distinguished from changes indicative of more latent

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

or chronic effects. Use of more limited toxicogenomic assays must be accompanied by recognition that the screening is only partial. From the industry standpoint, for the screening tests to be economically feasible, they must also be highly specific and avoid falsely identifying benign chemicals as potentially toxic. To be acceptable to public health and industry stakeholders, toxicogenomic screening assays for industrial chemicals need to demonstrate greater accuracy than traditional tests as well as the ability to detect a broader array of outcomes than they currently do.

Database Size and Quality

A second challenge to the development of toxicogenomic approaches to screening is the availability of sufficiently robust and comprehensive data collected into a database. Many of the issues of database quality and comprehensiveness are relevant to predictive toxicology in general and are discussed in detail in Chapter 10. Issues of specific relevance to screening applications include comprehensiveness of end points and modes of action assessed. As discussed elsewhere in this chapter, to be useful for the broad screening of environmental chemicals for many potential effects, toxicogenomic assays must be able to signal the broad array of end points of concern, including cancer, reproductive and developmental toxicity, neurotoxicity, and immunotoxicity. Datasets are used only to classify compounds that act via those modes of action that are represented in their training sets of model compounds. False-negative rates from incomplete datasets may be acceptable in some settings but not in stand-alone screening programs for environmental chemicals.

In addition to being informative about a useful range of end points, the data need to include information useful in predicting responses at relevant doses and exposure timing. This is because responses to chemicals can involve multiple modes of action, depending on dose and time course of exposure. A population of databases with data on dose dependence and timing dependence of well-characterized compounds is required to validate the use of toxicogenomic assays for environmental chemical screening.

Reducing Costs of Toxicogenomics Screening Tests

A basic requirement of an effective screening test is that its cost is low enough to allow widespread application. At present, the cost of toxicogenomic technologies is expected to present a significant barrier to their widespread application in screening programs, especially for environmental chemicals. Toxicogenomic methods are currently being used primarily as adjuncts to standard animal tests. Whereas this may result in selective use of expensive animal assays, the actual costs of screening tests using toxicogenomic techniques in this manner include costs of the short-term animal assays plus the additional costs of toxicogenomic analyses. If toxicogenomic screens with model organisms such

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

as *Caenorhabditis elegans* can replace rodent assays, then cost will become less of a barrier to using toxicogenomic approaches as screening tools.

Achieving High Throughput

The pharmaceutical and biotechnology industries have developed highly automated assays to screen numerous chemicals for relatively simple properties, such as target receptor binding (Walum et al. 2005). Similar assays are being developed to assess toxicity; they range from receptor binding assays linked to toxicity mechanisms to automated cytotoxicity assays (Walum et al. 2005). Such assays are capable of assessing tens or hundreds of thousands of compounds per week. Current toxicogenomic assays are "high throughput" only in the sense that they analyze thousands of genes or proteins simultaneously. To the extent that they can predict outcomes of standard animal tests in a shorter time frame, toxicogenomic technologies allow faster screening of compounds. In vitro screening assays using toxicogenomic techniques, combined with reproducible, validated automation techniques will be needed to truly achieve high-throughput toxicogenomic-based screening.

CONCLUSIONS

Toxicogenomic technologies can facilitate the screening of chemical compounds for their ability to cause toxicity. In vivo and in vitro screens currently are used to detect well-characterized toxicities (hepatotoxicity, nephrotoxicity) in settings where false-negative results can be tolerated, such as drug development. The use of toxicogenomic technologies in the validation of chemical categories for screening programs that also employ traditional toxicologic assays is quite feasible in the near term. However, broad application of toxicogenomic technologies to screen environmental compounds will require demonstration of false-negative rates comparable or superior to those of traditional testing methods. For using toxicogenomic assays to screen for a battery of end points, reliable assays and their associated algorithms for a broader set of end points will need to be developed. These end points include those toxicities of greater interest for environmental chemicals (for example, cancer, neurodevelopmental toxicity) than are currently used in screening candidate drugs (for example, hepatotoxicity, nephrotoxicity). Although it seems unlikely the full range of toxicities assessed in traditional animal screening studies could soon be assessed by toxicogenomic screening, the availability of assays for specific end points (for example, cancer) would help alleviate the dearth of publicly available information on the potential toxicity of many chemical agents.

The pharmaceutical industry has developed the most advanced toxicogenomic screening applications. This reflects incentives to screen out undesirable properties and more efficiently identify drug candidates with the safest and most efficacious profiles. The business case for using toxicogenomic technologies

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

over other testing methods for pharmaceuticals has not yet been clearly demonstrated, but even the touted economic advantage of using toxicogenomic screening to avoid more detailed and expensive testing of pharmaceuticals does not exist for industrial chemicals. This is because regulatory requirements for generating toxicity data for industrial chemicals under the TSCA are far less stringent than those for drugs.

Toxicogenomic technologies for hazard screening have relied primarily on transcriptome profiling assays. New studies have shown the usefulness of proteomics and metabolomics for classifying biologic activity. Because protein and metabolite changes may be more closely linked to pathologic changes, they may ultimately prove to be valuable screening indicators of toxicity. Additional proof of concept research is required to assess their value.

Of the remaining critical challenges to implementing toxicogenomic technologies for screening, most fundamental is the development of comprehensive databases suitable for screening purposes. These databases should include toxicologic end points relevant to environmental chemicals and examples of all known toxic modes of action.

RECOMMENDATIONS

Intermediate

1. Convene an expert panel to provide recommendations for which model compounds, laboratory platforms, and specific data elements are necessary for building toxicogenomic databases for screening applications. Assessment of in vitro approaches for specific toxic end points should be emphasized. All processes, toxicogenomic data, and outcome data must be publicly accessible.

2. Expand the toxicogenomic component in existing testing programs such as the National Toxicology Program. The toxicogenomic data, collected in conjunction with the highest quality traditional toxicology testing, can then be used to help build public databases and increase understanding of the role of toxicogenomics in predicting chronic toxicity.

3. Develop databases and algorithms for using proteomic and metabonomic data in screening.

4. Regulatory agencies (including EPA and the Food and Drug Administration) should continue to develop and refine guidance documents for their staff on interpreting toxicogenomic data. In particular, guidance for environmental chemicals must ensure that screening protocols address the types of end points most relevant for the general population, including sensitive subpopulations. This effort will help define database, basic research, and training needs for the agencies.

5. Develop mechanisms to improve the quantity and quality of data available for deriving screening profiles:

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

a. Establish a dialog with entities holding currently inaccessible toxicogenomic data to evaluate options for increasing the availability of data.
b. Regulatory government agencies (including EPA and the Food and Drug Administration) should consider appropriate ways to address the following disincentives to industry generation and public submission of data that could be used to populate public databases: additional costs of testing; concerns about reporting requirements of the TSCA and the Federal Insecticide, Fungicide, and Rodenticide Act; concerns about competitors' use of toxicity data if they are made public; and concerns about the use of toxicogenomic data in tort liability cases.
c. Integrate data relevant to toxicogenomic screening of environmental chemicals into ongoing biomedical initiatives such as the National Institutes of Health Molecular Libraries initiative. Such data may include physicochemical characteristics, in vitro assay results such as cytotoxicity and receptor binding, and other screening-level types of data.

Long Term

6. Ensure that the regulatory framework provides incentives, or at least removes disincentives, for premarket testing of chemicals.

7. Upon validation and development of adequate databases, integrate toxicogenomic screening methods into relevant current and future chemical regulatory and safety programs.

Copyright © National Academy of Sciences. All rights reserved.

6

Application to Analyzing Variation in Human Susceptibility

As a rule, humans vary in their responses to environmental factors because of variability in their genes and their genes' epigenetic modification. Consequently, the same level of exposure to a chemical compound may give rise to different biologic effects in different individuals. For example, severe life-threatening toxicities can occur in some individuals treated with irinotecan, an anticancer drug. Although multiple genes play a role in irinotecan activity, polymorphisms in the UDP glycuronosyltransferase 1 family, polypeptide A1 (UGT1A1), enzyme have been strongly associated with irinotecan toxicity. Prospective screening of patients before chemotherapy could reduce the frequency of severe toxicities by alerting physicians to consider an alternative therapy (Marsh and McLeod 2004).

With completion of the sequencing effort of the Human Genome Project, new opportunities have arisen to more fully characterize the genetic contributions to variation in human susceptibility to toxic effects of pharmaceuticals and other chemicals. The remarkable advances in our ability to rapidly detect thousands of genetic variations have led to high expectations for the ability to discover and then apply critical new information to understand human susceptibility to disease. More than 6 million single nucleotide polymorphisms (SNPs) have been identified and catalogued in public databases. Research efforts are now under way to identify which SNPs are associated with variation in chemical toxicity as well as drug responsiveness. Animal genome projects also provide opportunities to understand how genetic variation affects toxicity in other animal species.

The new knowledge is likely to have scientific, clinical, and policy effects. Toxicogenomic technologies are expected to revolutionize strategies for predicting disease susceptibility and toxic response to environmental agents. Studies of

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

gene polymorphisms in the paraoxonase I gene (*PON1*) and the resulting differential response to organophosphate pesticides typify the type of genetic markers of toxic response that are likely to surface during the next decade (see Box 6-1).

Another potential impact arising from toxicogenomics is the development of new classifications of disease subgroups. Most adverse reactions to chemical or therapeutic compounds have been classified by biochemical or clinical markers (frequently based on histopathologies). New molecular classifications of disease are likely to arise as researchers better understand the genomic, transcriptomic, proteomic, and metabonomic characteristics of disease.

Finally, new knowledge about genetics and human variability in response is expected to enable greater tailoring of existing pharmaceuticals to patients to reduce toxicities and to better design new pharmaceuticals that produce fewer toxicities.

STATE OF THE ART IN ASSESSMENT OF INDIVIDUAL VARIATION

Toxicogenomic studies relevant to understanding human variability encompass various technologies and study designs. These range from investigations of variability in human gene expression profiles in responseto chemicals to large population-based cohort studies focused on identifying the genetic variations that influence sensitivity to chemicals. Dynamic modification of gene expression patterns without modification of the sequence, known as epigenetic phenomena, are also becoming better understood and characterized. This chapter reviews the state of the art in these areas and assesses future needs and challenges.

Variation in Gene Sequence

Gene-environment interactions refer to effects in which human genetic variability governs differential responses to environmental exposures such as the examples already discussed in this chapter. In this section, this concept is explored through a review of recent studies that identify genetic mutations associated with differential response to cigarette smoke and its association with lung cancer (Box 6-2). This study indicates that smoking is protective in some genotypic subgroups, which raises multiple ethical and policy related issues (see Chapter 11) yet typifies how gene-environment interactions may often appear counterintuitive with respect to our current knowledge base. This type of study also demonstrates the increased information provided by jointly examining the effects of multiple mutations on toxicity-related disease. Studies of polymorphisms in genes involved in Phase II metabolism (*GSTM1*, *GSTT1*, *GSTP1*) have also demonstrated the importance of investigating the combined effects of these variants (Miller et al. 2002).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 6-1 Paraoxonase 1 Gene Polymorphisms and Occupational Exposure to Pesticide

The paraoxonase 1 gene (*PON1*) encodes an enzyme involved in the metabolism of chlorpyrifos, an organophosphate pesticide, widely used in agricultural settings to protect crops from insects. Organophosphate pesticides affect the nervous system of animals and acute toxicity is characterized by nausea, diarrhea, impaired muscle function, and respiratory illness due to the effects on respiratory muscles (Blondell 1999). Studies of agricultural workers indicate that overexposure to organophosphates is relatively common (Ames et al. 1989; Ciesielski et al. 1994). Over the past two decades there has been extensive research on the relationship between pesticide toxicity and variations in the *PON1* gene. Battuello et al. (2002) recently reviewed implications for workplace screening of these *PON1* polymorphisms to assess whether identifying farmers who are more genetically susceptible to the adverse outcomes of pesticide use could reduce the disease burden in this population.

Variations in the *PON1* gene have been associated with both the amount and the type of paraoxonase 1 enzyme produced as part of the body's normal detoxification system. Specifically, a mutation in the amino acid sequence of the gene at position 192 changes a glutamine to arginine (Q192R) and is associated with variable enzyme activity in the population at large—low activity, intermediate activity, and high activity. In addition, at least one mutation in the promoter region of the gene has been associated with a twofold increase in gene expression and, consequently, enzyme production. The combined effect of the Q192R polymorphism and the promoter mutation effect on enzyme production has been referred to as the *PON1* status. Numerous animal studies have documented the relationships between these polymorphisms and pesticide toxicity, and variability in *PON1* status is associated with a greater than 60-fold interindividual difference in the chlorpyrifos detoxification rate in humans (Li et al. 1993).

Moreover, studies of the effects of the Q192R polymorphism in the human *PON1* gene introduced into transgenic mice illustrate the power of using animal models to understand the effects of human genetic variations on differential toxicity to pesticides in genetic subgroups of the human population (Cole et al. 2005). In humans, the R192 alloform has a higher catalytic efficiency of hydrolysis than does the Q192 alloform for chlorpyrifos oxon (CPO), the oxon form of the pesticide chlorpyrifos (CPS). Transgenic mice expressing PON1 Q192 were significantly more sensitive to CPO, and to a lesser extent CPS, than were mice expressing PON1 R192. Dose-response and time course of inhibition studies in these transgenic mice suggest that humans expressing only the PON1 Q192 allele would be more sensitive to the adverse effects of CPO or CPS exposure, especially if they are expressing a low level of plasma PON1 Q192 (Cole et al. 2005).

Although there is ample biochemical evidence of a significant genetic contribution affecting workers' health, there has not been substantial epidemiologic research in populations of exposed workers to quantify genetically influenced risk for this group.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 6-2 Multiple Genetic Factors Influence Response to Cigarette Smoke

Tobacco smoke contains a broad array of chemical carcinogens that may cause DNA damage. Several DNA repair pathways operate to repair this damage and the genes within this pathway are prime biologic candidates for understanding why some smokers develop lung cancers, but others do not. Zhou et al. (2003) examined variations in two genes responsible for DNA repair for their potential association with lung cancer. Briefly, one putatively functional mutation in the *XRCC1* gene and two putatively functional mutations in the *ERCC2* gene were genotyped in 1,091 patients with lung cancer and in 1,240 controls. When the patients and controls were stratified into heavy smokers and nonsmokers, Zhou et al. (2003) found that nonsmokers with the less frequent mutant genotype had a 2.4 times greater risk of lung cancer than those with the more prevalent genotype. In contrast, heavy smokers with the less frequent mutant genotype had a 50% reduction in lung cancer risk compared with their counterparts with the more frequent genotype. When the three mutations from these two genes were examined together, the extreme genotype combination (individuals with five or six mutations present in their genotypes) was associated with a 5.2-fold greater risk of lung cancer in nonsmokers and a 70% reduction of risk in the heavy smokers compared with individuals with no putatively functional mutations. The protective effect of these genetic variations in heavy smokers may be caused by the differential increase in activity of these protective genes stimulated by heavy smoking. Similar types of gene-smoking interactions have also been found for other genes in this pathway, such as *ERCC1*. The findings from this study have not been replicated. It is an illustration of the counterintuitive findings that are likely to emerge in studies of gene-environment interactions.

Gene-gene interactions are another important area of research for understanding human susceptibility to chemical sensitivity. This refers to situations in which one gene modifies the effect of another gene on disease or other adverse effect. In a recent study, McKeown-Eyssen et al. (2004) associated a gene-gene interaction between NAT2 and CYP2D6 enzymes with multiple chemical sensitivity. These results suggest that individuals with the rapid-metabolizing forms of both enzymes were 18 times more likely to have chemical hypersensitivity than individuals with normal metabolizing forms of these enzymes. Gene-gene interactions between CYP2D6 and another P450 enzyme (CYP3A4) have also been found to influence the metabolism of commonly used pharmaceutical agents (Le Corre et al. 2004).

If we are to adequately understand the continuum of genomic susceptibility to toxicologic agents that influences public health, more studies of the combined effects of multiple mutations are needed. The current emphasis on identifying single gene mutations associated with differential response to environmental exposures only delays understanding the distribution of genetic and genomic risks in human populations. Advances in bioinformatics can play a

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

key role in understanding combined effects of multiple mutations. For example, methods to screen SNP databases for mutations in transcriptional regulatory regions can be used for both discovery and functional validation of polymorphic regulatory elements, such as the antioxidant regulatory element found in the promoter regions of many genes encoding antioxidative and Phase II detoxification enzymes (X. Wang et al. 2005). Comparative sequence analysis methods are also becoming increasingly valuable to human genetic studies because they provide a way to rank-order SNPs based on their potential deleterious effects on protein function or gene regulation (Z. Wang et al. 2004). In addition, methods of performing large-scale analysis of nonsynonymous SNPs to predict whether a particular mutation impairs protein function (Clifford et al. 2004) can help in SNP selection for genetic epidemiologic studies and can be used to streamline functional analysis of mutations statistically associated with response to toxicologic agents. The use of bioinformatics in identifying and analyzing the biochemical and physiologic pathways (for example, systems analysis) by which gene-environment interactions occur is another key role toxicogenomics can play in helping genetic epidemiologic studies move beyond simple statistical association.

From a public health point of view, the impact of gene-environment studies on our understanding of the distribution of environmentally induced disease could have major ramifications for public policy. For example, a recent study of drinking water contaminants (commonly associated with trihalomethanes from chlorination) and *CYP2E1* gene mutations found a significant gene-environment interaction that affects fetal growth (Infante-Rivard 2004). Chlorination byproducts in drinking water come from reactions between chlorine and organic material in the source water. Studies of the putative mechanisms underlying such an association are essential to establishing the biologic plausibility of epidemiologic information, with integrated use of transcriptomic, metabonomic, or proteomic technologies to understand environmentally induced disease. These types of studies are likely to play a major role in translating basic genetic epidemiologic science into public health policies and practices.

Epigenetic Variability

Variations in susceptibility are due not only to polymorphisms in DNA sequence. *Epigenetics* refers to the study of reversible heritable changes in gene function that occur without a change in the sequence of nuclear DNA (see Chapter 2). Differences in gene expression due to epigenetic factors are increasingly recognized as an important basis for individual variation in susceptibility and disease (Scarano et al. 2005). The best known mechanism for epigenetic regulation of cell phenotypes is DNA methylation, which turns off a gene or gene region by changing the chemical structure of the DNA (Jaenisch and Bird 2003). For example, as a normal part of human development, genes are turned on and off by methylation processes stimulated by other gene products in the embryo,

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

fetus, newly born infant, adolescent, and aging adult. Environmental factors such as infection, diet, and chemical exposures are known to affect gene methylation (Sutherland and Costa 2003).

Anway et al. (2005) investigated the impact on rats of transient in utero exposures to two endocrine disruptors, vinclozolin (a fungicide commonly used on crops) and methoxychlor (a pesticide used as a replacement to dichlorodiphenyltrichloroethane [DDT]). Mothers were treated at a critical time during gonadal sex determination or a later embryonic period. The adult male offspring developed reduced spermatogenic capacity (decreasing sperm count and spermatogenic cell viability) and decreased fertility in this and two previous studies (Cupp et al. 2003; Uzumcu et al. 2004). In the latest study, although only the original gestating mother for the first generation was treated with vinclozolin, diminished male fertility was transmitted to the subsequent four generations (F_1 to F_4) when offspring males were crossed with offspring females from mothers that were exposed only once. Methoxychlor had similar effects but they extended only to the F_1 and F_2 generations. Further analysis indicated that these were male germ line effects associated with altered DNA methylation patterns. The study thus suggests that environmental factors can induce an epigenetic transgenerational phenotype through an apparent genetic reprogramming of the male germ line. (However, for this to be truly considered to be epigenetic, germline DNA mutations must be ruled out, which would require sequencing the entire genome). Nickel, cadmium, and xenobiotics (such as diethylstilbestrol) have been shown to affect gene methylation (Sutherland and Costa 2003; Bombail et al. 2004).

As this field progresses, it will be important to integrate epigenetic and genetic approaches to better model the risk of disease caused by environmental toxicants. Models of how to merge epigenotype and genotype information are now starting to emerge (Bjornsson et al. 2004) and more theoretical, as well as applied, work is needed in this area of toxicogenomics. Furthermore, work on integrating epigenetic data, both the causes and consequences of epigenetic modification, into dynamic system biology models of the regulation of gene expression, proteomic, and metabonomic profiles is also needed.

Gene Expression Variability

The sections above describe individual variability as assessed by studies that look at variations in gene sequence or epigenetic modification among individuals. Another way to assess human variability is to look downstream of the gene sequence or its epigenetic modification to the amount of mRNA expressed by the genes, examining differences in amount expressed rather than just differences in what is expressed. The variability in gene expression can reflect individual variability due to mutations in the gene, its promoter or other regulatory regions, and other modifications of expression such as epigenetic effects.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Several landmark studies have shown that gene expression may profoundly vary due to gene sequence variation. Lo et al. (2003) investigated allele-specific expression of 602 transcribed SNPs and found that 54% showed preferential expression of one allele over another, frequently greater than a fourfold difference in expression between the two alleles. Similarly, Chueng et al. (2002, 2003) demonstrated that the expression level of genes is highly heritable in humans, with one-third of the genes with heritable expression patterns showing evidence of mutations that directly affected transcription levels. With transcriptomic profiles and genomic data simultaneously provided, new insights into the causes of variability in gene expression are being discovered. This type of research could explain variation in toxic responses to chemical agents that is not due to underlying differences in gene sequence.

A study of SNP variation in human carboxylesterases illustrates how research on both gene expression and genetic sequence together could be used to study human variation in drug responsiveness. Human carboxylesterases 1 and 2 (CES1 and CES2) catalyze the hydrolysis of many exogenous compounds and play an important role in the metabolism of toxic chemicals in the body. Alterations in carboxylesterase sequences could lead to variability in both the activation and inactivation of drugs. Marsh et al. (2004) sequenced the *CES1* and *CES2* genes in individuals in European and African populations, identifying novel SNPs in *CES1* and *CES2*. At least one SNP in the *CES2* gene was associated with reduced CES2 mRNA expression. In summary, functional analysis of novel mutations found to affect gene expression patterns could provide important insight into variation in drug responsiveness.

Using Animal Models to Identify and Evaluate Susceptibility Genes

Animal models offer important experimental research opportunities to understand how genetic factors influence differential response to toxicologic agents. Animal models are advantageous as a first line of research because they are less expensive, less difficult, and less time-consuming than human studies. In addition, animal studies can address questions that are almost insurmountable in human studies, such as questions about sporadic effects or effects that cannot be adequately examined for sex linkage because of sex bias in employment.

Because response is most often quantitative, theoretical models of the cumulative action of mutations in multiple genes and multiple gene-environment interactions have been used to identify which regions of animal genomes are related to response. Genetic analysis of these complex quantitative traits by classic Mendelian methods is not possible. Because of advances in statistical approaches capable of analyzing extensive genetic data, rapid quantitative trait mapping in animal and human genomes has become more feasible (Lander and Botstein 1989; Silver 1995; Manly and Olson 1999). When combined with selective breeding designs in model species, this approach identifies genes in the

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

model species that can then be mapped onto human chromosomes by using comparative genomic and bioinformatic methods.

The mouse offers several advantages in the initial determination of genetic traits that control human conditions. First, inbred and wild-derived inbred mice allow research to focus on the mechanisms of resistance and clear distinctions in susceptibility among inbred strains of mice. Study of inbred mouse strains can also be advantageous because, unlike humans, their polymorphisms often become "fixed" in a population (carried by all the mice) by inbreeding of the strain. Moreover, resistance to one disease may lead to susceptibility to another. Because the mouse genome has been mapped and is largely (>97%) identical to the human genome, studies to identify new genes for susceptibility can be efficiently accomplished in mice, significantly accelerating research on homologous genes in humans.

A number of studies illustrate how these advantages have enabled the mouse to be a powerful model for the dissection of genetic factors contributing to a number of complex diseases, ranging from immune disorders and cancer predisposition (Todd et al. 1991; MacPhee et al. 1995; De Sanctis and Drazen 1997) to coagulation disorders (Mohlke et al. 1996). Genomic animal approaches also increased the ability to uncover genes not previously associated with susceptibility to adverse effects from ozone (Kleeberger 1991; Kleeberger and Hudak 1992; Kleeberger et al. 1993a,b, 1997), lipopolysaccharide-associated lung injury (Arbour et al. 2000; Kiechl et al. 2002; Cook et al. 2004), and acute lung injury (Prows et al. 1997, 1999; Prows and Leikauf 2001; Wesselkamper et al. 2005). These discoveries, although requiring considerable time and effort, yield new information about the biology of the disease process underlying environmental injury and could lead to further detection of human mutations and their functional significance.

Understanding the role of a genetic association in mice can lead to identification of analogous human mutations or analogous alterations in human protein function. Studies of the *SLC11A1* gene for proton-coupled divalent metal ion transporters best illustrates the concepts of using inbred mice and the effects of a single mutation on multiple traits (Liu et al. 1995; Fortier et al. 2005). Nucleotide sequence analyses of the *SLC11A1* cDNA in 27 inbred mouse strains that were either resistant or susceptible to intracellular parasite infection demonstrated that susceptibility was associated with a mutation that caused a glycine-to-aspartic acid amino acid substitution in the corresponding protein product (Malo et al. 1994). The human *SLC11A1* gene encodes a 550-amino acid protein showing 85% identity (92% similarity) with mouse *SLC11A1*. Although the mouse susceptibility polymorphism was not found in the human gene (Blackwell et al. 1995), other human polymorphisms associated with disease resistance were found.

Bellamy et al. (1998) examined the role of SLC11A1 in tuberculosis. In a case-control study in Africa, four SLC11A1 polymorphisms were significantly associated with tuberculosis susceptibility. Searle and Blackwell (1999) found a polymorphism that confers resistance to infection and was also associated with

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

chronic hyperactivation of macrophages. They hypothesized that this polymorphism was functionally associated with susceptibility to autoimmune disease. Analysis of these polymorphisms in patients with rheumatoid arthritis found that increased susceptibility to arthritis was associated with the mutation that conferred resistance to tuberculosis (Shaw et al. 1996; Bellamy et al. 2000). In these examples, understanding the role of a genetic association in mice led to a hand-in-hand assessment of associations in humans. Although the same mutations were not identical across species, knowledge of mutations that can alter protein function (in mice) or gene expression (in humans) were linked by the functional role this gene played in infection and arthritis.

In another study with inbred mouse strains, Arbour and coworkers (2000) compared the susceptibility of 40 strains to bacterial lipopolysaccharide administration. Genetic linkage analysis and transcriptional profiling identified the *TLR4* gene, which encodes the toll-like receptor 4 as the gene primarily responsible for variation in susceptibility. Moreover, the toll receptor showed variation not only among differentially susceptible mouse strains, but variants were also shown to determine differential susceptibility in humans. This is an excellent example of how toxicogenomic investigation of interstrain response variability can be used to study the effects of human variability.

Of particular utility to this approach is the set of recombinant inbred mouse panels that the National Institutes of Health has generated (Churchill et al. 2004). Because each strain represents a random assortment of susceptibility loci, the use of these panels will be particularly helpful in elucidating the effects of quantitative trait loci of susceptibility.

Recently, Churchill et al. (2004) proposed an initiative entitled the Collaborative Cross to promote the development of a genetically diverse set of mouse resources that can be used to understand pervasive human diseases. The goal is to breed current inbred mouse strains to create a more genetically heterogeneous, yet stable, resource for examining polygenic networks and interactions among genes, environment, and other factors. Existing resources optimized to study the actions of isolated genetic loci on a fixed background are less effective for studying the complex interactions among genetic factors that are likely to give rise to a substantial proportion of human susceptibility. The Collaborative Cross will provide a common reference panel specifically designed for the integrative analysis of complex systems and has the potential to change the way animal models can be used to understand human health and disease. New strategies for using animal models of toxicity are likely to yield valuable information for assessing therapeutic strategies and genetic differences that alter susceptibility.

RISK ASSESSMENT

Integrating genetics into the risk assessment process, including protecting sensitive populations, will require more directed research to support estimates of

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

key parameters such as uncertainty factors and on physiologically based pharmacokinetic (PBPK) models associated with genetically influenced human variability.

Risk assessment methodologies currently assume a 10-fold range in sensitivity to toxics in the human population and use a 10-fold uncertainty factor to account for this variability. Developing literature clearly indicates that the range in human sensitivity to toxic exposures has a genetic component for at least some classes of compounds.[1] For example, the Glu-69 polymorphism in the *HLA-DP6* gene has been shown to lead to unusual sensitivity to beryllium, the *PON1* gene appears to be important to the metabolism and detoxification of organophosphate pesticides, and the *NAT2* gene is associated with slow acetylation of arylamine compounds. A review of human data on therapeutic drugs indicates that the metabolism and elimination of most drugs are also subject to wide variation (Renwick and Lazarus 1998). More recently, a review of human variability in different routes of metabolism of environmental chemicals suggests a range greater than 10-fold in individual susceptibility to some chemicals (Dorne et al. 2005). In general, improvements in risk assessment are expected as more research is done to determine the range of allele frequencies and the impact of genetic variability associated with different ethnic groups as well as the elderly, children, and neonates. Currently, the default kinetic uncertainty factor of 3.16 would not be conservative enough to cover the variability observed in all subgroups of the population for compounds handled by monomorphic pathways versus polymorphic pathways (for example, CYP2C19 and CYP3A4 metabolism in Asian populations; CYP2D6, CYP2C19, NAT, and CYP3A4 in the elderly; and CYP2D6 and CYP2C19 in children). Kinetic data available in neonates compared with healthy adults for four pathways (CYP1A2, CYP3A4, glucuronidation, and glycine conjugation) demonstrated that the default value of 3.16 would be adequate for adults, whereas uncertainty factors greater than 12 would be required to cover up to 99% of neonates (Dorne et al. 2005).

In a recent paper, Haber et al. (2002) analyzed the potential contribution of mutations in enzymes influencing the disposition of four different types of compounds—methylene chloride, warfarin, parathion, and dichloroacetic acid—by PBPK modeling. They identified several key uncertainties regarding whether genetic mutations are an important source of variability in human susceptibility to environmental toxicants. The key issues they identified include the following: (1) the relative contribution of multiple enzyme systems, (2) the extent of enzyme induction/inhibition through coexposure, (3) differences in mutation frequencies across ethnic groups, (4) the lack of chemical-specific kinetic data for different genetic forms of the enzymes, (5) the large number of low-frequency mutations with significant effects, and (6) the uncertainty caused by differences

[1]Genetic variations in susceptibility are due not only to polymorphisms in DNA sequence. As discussed above, differences in gene expression due to epigenetic factors are increasingly recognized as an important basis for individual variation in susceptibility and disease (Scarano et al. 2005).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

between in vitro and in vivo kinetic data. There are critical gaps in the data required to assess and integrate genetic information into PBPK modeling to quantitatively assess its impact on population variability.

An example of how genotype-specific PBPK data could be integrated into risk assessment for a population is illustrated by the work of El-Masri et al. (1999), who modeled the effects of *GSTT1* mutations on the risk estimates for dichloromethane toxicity in humans. Dichloromethane is used in many industrial settings, including agriculture and food processing. By modeling the effect of genetic variability in the physiologic and biochemical processes underlying risk estimates, they (El-Marsi et al. 1999) and others (Andersen et al. 1987) concluded that the intrapopulation variability caused by the protective effect of the mutation can significantly increase the variability in the safe dose estimate of dichloromethane in a population. Other work also illustrates how understanding human kinetic variability could influence risk assessment (Dorne et al. 2002; Timchalk et al. 2002; Meek et al. 2003b).

CHALLENGES

There are several significant challenges to using toxicogenomic technologies to understand variation in individual or population susceptibility to chemical and pharmacologic compounds. First, the genetic architecture of human chemical sensitivity is complex. There are likely to be only a few rare instances when single gene mutations convey significant sensitivity to normal levels of exposures regardless of other contexts (e.g., Weber 1997). Much more frequently, there will be many genes with moderate or small effects on susceptibility, which in combination define susceptibility to a toxic agent. Interactions between gene variations, as well as additional gene-environment interactions and epigenetic processes, are likely to play a significant role in determining sensitivity to particular environmental exposures. This etiologic heterogeneity poses substantial challenges from both a methodologic and a risk assessment point of view.

Second, the understanding of the distribution of SNPs in the human gene pool is only beginning, and accurately typing large numbers of SNPs remains a work in progress. Multistaged research strategies (for example, linkage analysis to identify potential genomic regions followed by positional candidate gene studies or genome scans using tag SNPs followed by fine SNP mapping; see Chapter 2) are used to identify the set of genes and their variations that are most significantly associated with differential toxicity to chemical and pharmaceutical compounds. These multistaged research approaches have at their core an assumption that single mutations will have statistically significant, context-independent effects (that is, they will have the same effect in many different populations or contexts). True multigene models of susceptibility have not been practically obtainable to date and need to be a major focus of the next generation of toxicogenomic studies.

Copyright © National Academy of Sciences. All rights reserved.

Third, most large-scale environmental epidemiologic studies have not embraced genomic questions and toxicogenomic technologies as a part of their investigative approach. For example, clinical drug trials do not systematically collect and store blood for toxicogenomic or pharmacogenomic analyses. In some cases, biologic samples are available, but the funds and expertise for conducting the genomic studies in these population resources are limited or difficult to coordinate. Epidemiologic research in toxicogenomics is difficult because it requires multidisciplinary state-of-the-art teams of experts to measure genetic or toxicogenomic-derived markers, to measure environmental exposures, and to conduct clinical assessments, which take coordinated efforts among many different scientific disciplines. Unlike the toxicogenomic studies being carried out in animal models, which often rely on inbred strains, humans have a much higher level of genetic variability. This natural human variability makes large-scale epidemiologic studies imperative to scientific and policy development and it makes the understanding of disease risk incredibly complex.

Fourth, many researchers are finding that results from genetic association studies are not consistent from study to study (Hirschorn et al. 2002). There are several reasons for this lack of replication across studies, ranging from the statistical issues that arise from small studies (stemming from the expense of these technologies) to differences across studies in the distributions of underlying genetic variations, exposure, and accumulated genomic changes that occur at the epigenetic level. Studies of cohorts large enough to offset the small-sample random sources of variation from the important biologic variations will increase the power to identify reliable toxicogenomic predictors of susceptibility.

There are also numerous genetic epidemiologic studies that have developed transformed cell lines from human lymphocytes as a way to create inexhaustible supplies of DNA for genomic studies. These biologic samples could provide extremely valuable experimental material to determine the impact of interindividual variation in genes in response to industrial and pharmacologic compounds through in situ studies. Further research in this area is needed to determine how the results from cell line studies are translatable to human health effects.

A summary of the research issues and potential applications of genetic studies is listed in Box 6-3. In the following section, we outline the recommendations for immediate, intermediate, and long-term actions.

CONCLUSIONS

A key stumbling block to applying toxicogenomic information to risk reduction in humans has been the difficulty in conducting large population studies to understand the distribution of gene-environment interactions in the population at large. Without adequate measures of exposure, studies of gene-environment interactions cannot be carried out effectively.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 6-3 Summary of Research Issues and Potential Applications

Questions to be answered

- How does human genetic variation influence transcriptomic, proteomic, or metabonomic patterns of response to toxic agents?
- Is genetic susceptibility to different classes of toxicologic agents due to a few key genes or does it represent a continuum of multigenic risk?
- How many drugs fail clinical trials because of a toxic response in a genetically susceptible subgroup?
- How do we best use existing environmental cohort studies to identify gene-environment interactions with toxicogenomic approaches?
- How can toxicogenomic research be translated and tested to reduce health risks for the public?

Gaps in knowledge

- The influence of human genetic and epigenetic variation on transcriptomic, proteomic, and metabonomic studies of toxicologic agents is unknown.
- Although animal models and established human cell lines offer some insight into human response to toxic agents, questions remain about how well these studies indicate toxicologic risk in free-living humans.
- Cohort study estimates of the relative risk of disease for most genotype-environment combinations is lacking.
- Multigenic predictive models of toxicologic risk that integrate pleiotropic and polygenic networks of interactions among genes, environments, and other factors have yet to be developed.

How can this technology be applied?

- Identifying genetic variations associated with environmental susceptibility to toxicity can be used to identify at-risk subgroups of the population through genetic testing.
- Developing multigenic models of toxicologic risk can be used to better understand the distribution of risk in a population and can be used in risk communication efforts to reduce exposure and disease.

Animal models provide an important experimental method for identifying and characterizing genetic factors associated with increased susceptibility to toxicity from chemical exposure.

There is substantial evidence that genetic variations in many genes influence individual response to toxic agents. Heterogeneity in the distribution of susceptibility SNPs and environmental exposures, as well as heterogeneity in the relationship to disease of these factors (for example, gene-gene interactions) and how they are affected by other factors (for example, age and sex), needs to be better understood in human populations to identify individuals and subgroups at

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

risk. If we are to adequately understand the continuum of genomic susceptibility to toxicologic agents that influences the public's health, more studies of the joint effects of multiple polymorphisms need to be conducted. Much of the current research emphasizes identifying single gene mutations associated with differential response to environmental exposures. A more holistic approach to the analysis of data, an approach that encompasses gene-gene and gene-environment interactions, is likely to more efficiently advance our understanding of the population distribution of genetic components of risk.

The influence of toxic substances on epigenetic modification of an individual's genome is likely to depend on variation in the type, timing, and duration of exposure as well as the underlying genomic variation.

RECOMMENDATIONS

Immediate Actions

Exposure Assessment

1. Ensure that resources are adequately allocated to exposure monitoring and detection (external and internal to the individual), with approaches that are high speed and high dimensional (that is, can measure multiple chemical compounds simultaneously).
2. Investigate the potential utility of metabonomic technologies to provide quantitative and qualitative assessment of an individual's exposure.

Animal Models

3. Use animal models to identify genes associated with variability in toxicity and to validate causal mechanisms underlying human gene-environment interactions.
4. Use animal models to model the genomic susceptibility (that is, polygenic) that is likely to underlie the continuum of genomic risk found in human populations.
5. Begin developing an animal model resource that mimics the genetic heterogeneity of human populations—a resource that can be used to study the distribution of gene-gene interactions and gene-epigenetic interactions and can serve as a model for understanding population risk.

Intermediate

Population Studies

6. Use genome-wide association studies, ranging from anonymous dense SNP scans to specialized arrays of putative functional SNP approaches, to iden-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

tify the full complement of genes and their variations that influence sensitivity to toxicologic agents.

7. Use existing environmental cohort studies and clinical drug trials to investigate the impact of genetic variations on variation in response to a wide range of chemical exposures and pharmaceutical therapies.

Context-Dependent Genetic Effects

8. In addition to understanding the influence of single SNP variations on susceptibility, focus more attention on investigating context-dependent genetic effects (that is, gene-gene interactions as well as interactions with other biologic contexts such as developmental age, sex, and life course factors) that reflect the state of biologic networks underlying response to toxicologic agents.

Models

9. Develop multigenic and polygenic models of environmental sensitivity to better characterize the continuum of genomic susceptibility to toxicity and to better use genomic information for risk reduction.

Long Term

Epigenetics

10. Conduct research on the influence of exposure variation, genetic variation, and their interaction in determining epigenetic modification of the human genome.

11. Better characterize the influence of epigenetic modifications on disease processes that are associated with exposure to toxicologic agents to use this information for risk characterization and risk reduction.

Copyright © National Academy of Sciences. All rights reserved.

7

Application to the Study of Mechanisms of Action

The study of biologic mechanisms is a priority for basic and clinical researchers interested in elucidating the cellular, biochemical, and molecular basis of chemical and drug toxicity. Mechanistic insight is required for in-depth understanding of the pathobiology that underlies the adverse response to chemical exposures as well as the development of pharmacologic and nonpharmacologic strategies that can control or contain adverse outcomes in chemical toxicity. In addition, toxic chemicals are increasingly being used as tools to unravel complex mechanisms of disease onset and progression and to define the role of environmental injury in acute and chronic pathogenesis. As such, the toxicology research community readily embraces the application of new technologies and approaches to study mechanisms of toxicity.

From a practical perspective, knowledge and insight gained from mechanistic toxicology investigations have proven useful in risk assessment and drug development (Petricoin et al. 2002c). The chemical industry is interested in applying toxicogenomic technologies to monitor the biologic response of humans to their products and to study chemical-biologic interactions in target species. Current risk assessment processes use mechanistic understanding to detect or identify hazards and to make decisions about dose-response relationships. However, despite general acceptance of the concept that mechanistic information is valuable for understanding the basis of the toxic response, data from mechanism-based investigations are not incorporated into risk assessment paradigms as often because either understanding of the mechanism of toxicity is incomplete or competing hypotheses exist. The scarcity of comprehensive assessments of multiple exposures in real time or with adequate quantitative resolution has slowed progress. Recent reports by the National Research Council (NRC 2006b) and the Environmental Protection Agency (EPA 2003) on the health risks from

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

dioxin and related chemicals as well as the report of the International Agency for Cancer Research on the human carcinogenicity of benzo(*a*)pyrene (IARC 1987) exemplify the usefulness of mechanism-based information in risk assessment. In sharp contrast to the chemical industry, the pharmaceutical industry routinely uses mechanisms and molecular level understanding to identify off-target biologic responses to potential new compounds and products already on the market. In fact, these data are used in new drug applications for regulatory approval by the Food and Drug Administration. In both settings, mechanistic studies have greatly facilitated informed decision making and more detailed understanding of critical issues.

Toxicogenomic studies offer the opportunity to evaluate molecular mechanisms of toxic action and the degree of conservation of biologic response pathways across species that are responsible for toxicity. When applied to the study of large classes of chemicals or drugs, toxicogenomic information can be used to globally define modes or mechanisms of toxic action. The application of toxicogenomics to the study of toxicity mechanisms rests on the premise that chemical or physical injury is mediated by, or reflected in, changes at the mRNA, protein, or metabolite level. Abundant evidence supports this concept that toxicity coincides with changes in mRNAs, proteins, and metabolites. As such, these changes under defined conditions of cellular location, time, and biologic context can provide meaningful information about biologic responses to toxic insult. Thus, toxicogenomic studies offer a new dimension in environmental exposure assessment, drug and chemical screening, and understanding of human and animal variability in response to drugs and chemicals.

This chapter provides examples of the application of toxicogenomic analyses for exploring toxicity mechanisms of environmental chemicals and pharmaceuticals. Because transcriptome profiling technologies are technically more mature than methods for proteomics or metabonomics, the discussion focuses primarily on transcriptomics, with complementary insights derived from proteomics and metabonomics. The limitations of the various technologies and a needs assessment are also presented.

STATE OF THE ART IN TRANSCRIPTOMIC ANALYSES

Investigators have exploited transcriptome profiles to understand mechanisms of environmental chemical or drug toxicity. Generally, the experimental approaches used can be categorized in three broad approaches (Box 7-1), which are discussed later in this chapter: the contextual approach, the exploratory approach, and the network building approach.

Transcriptome profiling experiments designed to evaluate toxicity provide a “snapshot” of global gene activity as measured by steady-state levels of mRNA at a precise point in time during the course of the toxic response. Genetic and epigenetic mechanisms that orchestrate the recruitment of RNA polymerase

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 7-1 Types of Experimental Approaches in Toxicogenomic Studies to Evaluate Mechanisms of Toxicity

- **The contextual approach** places newly generated toxicogenomic data in the context of available biologic and toxicologic knowledge by comparing a set of elements (transcriptome profiles) with known response parameters. The reasoning process is deductive.
- **The exploratory approach** moves from a discovery to a hypothesis-driven approach that defines specific mechanisms of action. The information obtained generates novel insights into biologic and toxicologic mechanisms. The reasoning process is, by definition, more inductive.
- **The network building approach** uses patterns of transcriptional coregulation or other means to construct interaction maps for affected genes and regulatory networks. The information provides opportunities to deduce relationships that place individual responses within the broader framework of complex interactions.

and other components of the transcriptional machinery for the synthesis of RNA control the amount of gene activity. Once the DNA template has been copied, the cellular machinery must process the precursor RNA into one (or more) mature forms that act as the template for protein synthesis. Proteins, in turn, are responsible for the biologic processes that govern cellular functions, susceptibility to chemical toxicity, and generation of products of cellular metabolism. Combining transcriptional analysis with parallel or correlational analyses of both protein and metabolite profiles makes it possible to examine the multidimensionality of the toxic response. The study of mechanisms is complicated by the fact that toxicity often involves complex interactions of chemicals with genes, proteins, and cellular metabolic intermediates.

Genetic variability is an important determinant of the transcriptional response elicited by toxic insult. Thus, genetic factors such as polymorphisms that contribute to variable susceptibility to chemical and drug toxicity need to be factored into toxicogenomic analyses. In addition to genetic influences on toxicogenomic response, recent studies have established the importance of epigenetic mechanisms (heritable changes in gene expression profiles that occur without changes in DNA sequence; see Chapter 6) in toxic injury. The epigenome regulates mRNA abundance and is a critical determinant of gene expression. Microarray technology is now capable of assessing epigenetic variability, and impacts of the epigenome on toxicogenomic and transcriptomic responses are being established (Stribinskis and Ramos 2006).

Also relevant to this discussion is the recent introduction of DNA microarray chips to study the role of microRNAs in biology (Monticelli et al. 2005; Shingara et al. 2005; Castoldi et al. 2006). MicroRNAs are a class of naturally occuring, small, noncoding RNAs involved in regulating the translation and

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

processing of mRNAs. MicroRNAs bind to the 3' untranslated region of target mRNAs and target the transcript for degradation. As such, microRNAs are believed to play important roles in regulating cell replication, differentiation, and apoptosis. Aberrant patterns of microRNA expression have now been implicated in human cancer (Iorio et al. 2005), suggesting that toxicity may involve alterations in RNA processing.

Contextual Approach to Study Mechanisms of Action

The contextual approach (Box 7-1) to studying mechanisms of action places newly discovered information within the context of the larger body of biologic and toxicologic knowledge, comparing newly generated toxicogenomic data from a test chemical with data from compounds whose mechanism of action is better understood. The experimenter may have some prior knowledge of the nature of molecular changes that would be consistent with an adverse response. Alternatively, the experimenter may not know of the molecular nature of the adverse response but can use the newly acquired data to determine possible toxicities. The result of this analysis is evidence implicating a set of genes in toxicity and identification of the similarities and differences of the test chemical's toxicogenomic data with known toxicity mechanisms. Thus, applying a contextual approach facilitates the discovery process.

Contextual analysis of gene expression is best exemplified by studies in which the expression profiles of a chemical with unknown effects are compared with the profiles of gene expression of compounds with known effects. This is similar to the class prediction approach described in Chapter 3 and its applications to hazard screening as described in Chapter 5. Although mainly descriptive, the approach is unbiased and does not require a detailed understanding of affected biologic pathways, and categorical classifications do not require annotations or other detailed understanding of affected genes. Compounds with similar toxicities often affect similar genes, as shown in experiments in which "fingerprints" of the genomic response to different chemical classes have been developed (Hammadeh et al. 2002a,b).

Several examples of this approach have been described in the literature and include studies of the responses initiated when 2,3,7,8-tetrachlorodibenzo-*p*-dioxin interacts with the aryl hydrocarbon receptor (AhR) (Puga et al. 2000; Frueh et al. 2001; Kurachi et al. 2002; Boverhof et al. 2005; Fletcher et al. 2005; Thackaberry et al. 2005). Vezina et al. (2004) used DNA microarrays to identify rat hepatic transcriptome profiles associated with subchronic exposure to three AhR ligands and one structurally related non-AhR ligand. The AhR ligands produced similar gene expression profiles that could be readily distinguished from the non-AhR ligand. Thus, the gene expression changes controlled by the AhR could be enumerated. This enumeration led to several genes not previously characterized as AhR targets (carcinoembryonic cell adhesion molecule 4 [C-CAM4] and adenylate cyclase-associated protein 2 [CAP2]) being identified.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

The data provided mechanistic insight by implicating novel genes in the AhR signal transduction pathway and identifying putative targets of toxicity.

In another study of mechanisms of chemical carcinogenesis, Dickinson et al. (2004) compared the effects of the genotoxic drug cisplatin with non-genotoxic osmotic stress by sodium chloride. Transcriptome profiles of cisplatin-treated cells revealed significant increases in transcripts associated with DNA damage (for example, members of the GADD45 family) and adaptive cellular repair (for example, fos and HSP40 homologue). In contrast, at equitoxic concentrations, the gene expression profile of sodium chloride-treated cells did not indicate changes in transcripts associated with DNA damage and repair. This suggests that DNA damage and cellular repair may be important in cisplatin-mediated toxicity. In a similar study, Ellinger-Ziegelbauer et al. (2004) identified transcripts that characterize the DNA damage response, induction of drug metabolism enzymes, and survival of proliferative pathways after a 2-week exposure to four hepatocarcinogens. Observed differences in gene expression profiles may represent the processes preferentially affected by the carcinogens and may be related to the mechanism of carcinogenesis.

Exploratory Approach to Study Mechanisms of Action

The second experimental approach, the exploratory approach to studying mechanisms, moves from generating contextual insights to testing specific hypotheses with lower throughput and more detailed molecular and biochemical analyses. The information derived from exploratory studies provides insight into biologic and toxicologic mechanisms by using a reasoning process that is more inductive. This hypothesis-driven approach is highly complementary to the contextual approach described above, so the two approaches are often used in concert to test specific hypotheses. For example, knowledge of the identity of a specific gene discovered in the contextual approach can be used later in exploratory studies to reveal downstream events involved in a toxic response. More detailed information about mechanisms of toxicity is gained by perturbing the system to study biologic relationships. In this scenario, genes identified by using a contextual design are disrupted pharmacologically, genetically, or molecularly and the genome-wide response is reevaluated to gain more detailed information and additional mechanistic insight. For example, pharmacologic agents or molecular interventions such as posttranscriptional gene silencing, also known as RNA interference, are used to selectively target genes of interest and test specific hypotheses that involve studying gene-gene associations that emerged from transcriptional profiling experiments (Hannon and Rossi 2004). A well-designed series of experiments can provide a more in-depth understanding of how multiple elements within the affected pathways are connected and their roles in toxicity. This approach is biased toward known biologic mechanisms, especially those well represented in the published literature, and success depends on the

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

knowledge and experience of the investigator. Nonetheless, it can provide an effective means to uncover mechanisms of toxicity.

The exploratory approach frequently relies on pathway analyses to determine how discrete elements (transcripts, proteins, and metabolites) contribute to mechanisms of toxic responses. These analyses often use the *Kyoto Encyclopedia of Genes and Genomes* or the Gene Ontology database to define categories of cellular components, biologic process, or molecular function and software tools (for example, Expression Analysis Systematic Explorer, Ingenuity, GeneGo or MicroArray Pathway Profiler) to reveal pathways.

Multiple examples of the exploratory approach can be found in the primary literature. Thomas et al. (2001) identified a set of genes useful for predicting toxicity using transcriptome profiles for 24 hepatotoxic compounds representing five toxicologic categories (peroxisome proliferators, AhR agonists, noncoplanar polychlorinated biphenyls, inflammatory agents, and hypoxia-inducing agents). After surveying 1,200 transcripts, either correlation-based or probabilistic analysis yielded a weak toxicologic classification accuracy (<70%) based on a known mechanism of action. However, with a forward parameter selection scheme, a diagnostic set of 12 transcripts was identified that provided 100% predictive accuracy based on a "leave-one-out cross-validation" approach. A forward parameter selection scheme is an iterative process in which transcripts are examined individually with a naïve Bayesian model and the transcripts with the best internal mechanistic classification rates and highest confidence (representing the sum of all probabilities for correctly classified treatments) are selected. This exemplifies how a classification approach coupled with statistical analysis can be used in exploratory experiments to study mechanisms of toxic action.

Moggs et al. (2004) also used the exploratory approach, in which gene expression changes that drive the response of immature mouse uterus to 17-beta-estradiol were identified. Gene expression changes were sorted into groups of selected cellular processes, such as protein synthesis and cell replication, and expression was analyzed relative to changes in uterine weight. Correlation of phenotypic (uterine weight in this case) and toxicogenomic response, known as "phenotypic anchoring" (Tennant 2002), is useful in linking transcriptional perturbation events to toxicity outcomes. Ulrich and coworkers (Waring et al. 2002) used a classification approach together with discrete gene analysis to examine the biologic pathways affected by an inhibitor of the transcription factor nuclear factor kappa B as well as the genes associated with hepatic hypertrophy and toxicity.

Experiments by Anazawa et al. (2005) exemplify the use of molecular intervention combined with transcriptome profiling study mechanisms of toxicity. This study examined gene expression profiles of prostate cancer cells isolated by laser capture microdissection to exclude signal contamination by other cells. A newly identified gene, prostate collagen triple helix (*PCOTH*), was found to be specific to prostate cancer cells and the corresponding protein was expressed in prostate tumor tissue. Decreasing *PCOTH* expression with small interfering

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

RNA attenuated prostate cancer cell growth, whereas increasing *PCOTH* by DNA transfection enhanced growth. In cells expressing the increased exogenous *PCOTH*, phosphorylation of SET (a nuclear translocation protein expressed in acute undifferentiated leukemia) was detected. A reduction of endogenous SET levels also attenuated the viability of prostate cancer cells, suggesting that PCOTH mediates growth and survival of prostate cancer cells through the SET pathway. This investigation led the investigators to suggest PCOTH as a target for new therapeutic strategies for prostate cancers.

An example of how an iterative process can be useful once key elements have been uncovered is presented in a study by Guo et al. (2006), who used transcriptome profiling to study aflatoxin B1 (AFB1) injury in a genetically engineered yeast model. AFB1 is a potent human hepatotoxin and hepatocarcinogen produced by the mold *Aspergillus flavus* that affects the health of millions of people in developing countries worldwide. The authors engineered a yeast strain to express human cytochrome P4501A2, which metabolizes AFB1 to its ultimate mutagen. They also engineered the genetic background of the yeast to allow for direct correlations between toxicity, gene expression, and mutagenicity. Genes activated by AFB1 treatment included mediators of the DNA damage response. The study also found rapid and coordinated repression of histone and M/G_1 phase-specific transcripts, but these molecular changes were uncoupled from cell cycle arrest, suggesting that histone gene repression by genotoxic stress in yeast involves signaling pathways different from those involved in cell cycle control functions. This study exemplifies the power of transcriptional profiling combined with other molecular approaches to study mechanisms of toxicity.

Finally, animal models that have been engineered with added (knock-in) or deleted (knock-out) genes or that have been engineered to have human genes have been successfully used to evaluate the transcript profiles involved in chemical toxicity. These approaches allow investigators to construct experimental model systems with which to evaluate the interaction of added or deleted genes with the cellular machinery and its micro- or macroenvironment. The combination of genetically modified models with other toxicogenomic approaches provides a powerful tool to examine the molecular basis of the toxic response.

Network Building Approach to Study Mechanisms of Action

The third experimental approach uses patterns of transcriptional coregulation and targeted genetic manipulations to identify biologic interaction networks involved in toxicity. This type of experiment uses relational biologic information to place the toxicologic response within the complex framework that mediates the biologic response. The ultimate goal is to move from traditional reductionist approaches to analyze the behavior of the system as a whole—a “systems biology” approach. The application of systems biology ap-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

proaches will require that toxicogenomic analysis include not only multigene dimensionality but also the dimensions of dose and time. These analyses are highly dependent on databases of gene and protein interaction networks (interactomes), which provide a framework for interpretation. Interactomes for yeast and other model organisms are available and are becoming increasingly useful in this context.

Studies by Samson and colleagues, looking at yeast (*Sacchromyces cerevisiae*) treated with model alkylating agents, were the first to reveal distinct transcriptome profile changes to toxicants (Jelinsky and Samson 1999; Jelinsky et al. 2000). This group elaborated on this approach to generate the first systems biology-level analysis of toxicant action. They produced deletion mutant strains of yeast—strains with different genes deleted—to evaluate roles of individual genes, proteins, and interactome networks (Begley et al. 2002, 2004; Said et al. 2004). These studies revealed new, unanticipated stress signaling networks with novel interactions of genes involved in DNA repair, protein turnover, and metabolic regulation. The power of this yeast model is derived from integration of transcriptome profiling, an extensive protein-protein and protein-gene interaction database (Xenarios et al. 2002) and the availability of a near-comprehensive library of mutants with single genes deleted (Giaever et al. 2002).

Yao et al. (2004), who investigated signal transduction pathways evoked by AhR activation, reported another example of network building. These investigators used an ordered panel of yeast (*S. cerevisiae*) strains that harbored deletions in each of 4,507 genes. Because this collection of deletion mutant strains had essentially every gene in the organism deleted, addition (via transfection) of the human AhR gene in combination with AhR agonists (α- or β-naphthoflavone) could be used to systematically analyze the biologic interaction between specific genes and the AhR activation cascade. A relatively small number of genes (54) exerted a significant influence on AhR signal transduction. They then described the relationship between these modifying genes by using a network map based on the yeast interactome. This revealed that the AhR signaling is defined by five distinct steps regulated by functional modules of interacting modifiers. They classified these modules as mediating receptor folding, nuclear translocation, transcriptional activation, receptor level, and a novel nuclear step related to the Per-Arnt-Sim domain. By coupling computer-assisted and experimental annotations, this study identified the minimum number of genetic loci and signaling events required for AhR signaling; it exemplifies the power of the network analysis approach to study mechanisms of toxicity.

Another example is a study by C. D. Johnson et al. (2003) that evaluated transcriptional reprogramming during the course of environmental atherogenesis in response to benzo(*a*)pyrene, a ligand for AhR and an inducer of oxidative stress. A combined oxidant-antioxidant treatment regimen was used to identify redox-sensitive targets early during the course of the atherogenic response. Supervised and nonsupervised analyses identified transcripts highly regulated by benzo(*a*)pyrene, unaffected by antioxidant, and neutralized by combined chemical treatments. Critical transcripts included lymphocyte antigen 6 complex, his-

Copyright © National Academy of Sciences. All rights reserved.

tocompatibility class I component factors, secreted phosphoprotein, and several interferon-inducible proteins. A predictor algorithm was then applied to define critical gene-gene interactions involved in the atherogenic response. Transcriptional gene networks predictive of the transcriptional activation of AhR were defined in a later study showing that the expression of AhR is most commonly predicted by lymphocyte antigen 6 complex, locus e, a frequent predictor among three-gene combinations that included insulin growth factor binding protein 3 and tumor necrosis factor receptor superfamily member 1b (Johnson et al. 2004). Linkage diagrams of significant predictors were then used to delineate how individual genes integrate into a complex biologic network of genes potentially involved in the toxic response to AhR ligands.

Workman and coworkers (2006) recently used a systems biology approach to map DNA damage response pathways in yeast. In these studies, genome binding sites for 30 DNA damage-related transcription factors were identified after exposure of yeast to methyl methanesulfonate. These binding sites were identified by integrating transcription factor binding profiles with genetic perturbations, mRNA expression, and protein interaction data. The product was a physical map of regulatory interactions and biologic connectivities that can help define the biologic response to DNA-damaging agents.

State of the Art: Proteomic Analyses

The principles and concepts described for transcriptome analyses and its applications to the study of mechanisms of action apply to proteomics analyses. The integration of mass spectrometry (MS) with protein and peptide separation technologies has enabled the characterization of complex proteomes as well as mechanistic study of chemically induced protein modifications (Liebler 2002a; Aebersold and Mann 2003). Because of the extraordinarily wide range of protein expression levels ($>10^6$), no proteomic approach is capable of analyzing all proteins in the cell, tissue, or organism. Global (as opposed to comprehensive) proteome analyses have been applied to derive insight into mechanisms of toxic action. Two-dimensional gel electrophoresis and MS-based protein identification have revealed a number of proteome changes that are characteristic of chemical toxicities (Fountoulakis et al. 2000; MacDonald et al. 2001; Ishimura et al. 2002; Ruepp et al. 2002; Bandara et al. 2003; Xu et al. 2004). In these cases, a contextual or exploratory approach identified proteins and peptides that are correlated with injury and the adaptive biologic responses that follow. Although many of the changes observed may encode mechanistic information about chemically induced injury, distinguishing proteomic changes that represent causes versus effects requires more sophisticated experimental approaches.

As described in Chapter 2, one such study recently described the application of quantitative shotgun proteomics using isotope-coded affinity tag labels to compare proteomes of mice that are susceptible or resistant to the hepatotoxic analgesic acetaminophen both before and after drug treatment (Welch et al.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

2005). These analyses provided a global picture of proteome differences that may govern susceptibility as well as proteome changes that occur with injury.

Modeling protein interaction networks using time-series data is a powerful method to uncover a mechanism in toxicogenomics, as revealed by a study by Allen et al. (2006). These investigators argued that existing analytical approaches of proteomics typically yield a large list of proteins modified under specific conditions, but these data alone may not yield a systems biology understanding of cell signaling or produce hypotheses for future experiments. To improve understanding of proteomic data, these investigators used Monte Carlo approximations to predict behavior and to define relationships in cellular signaling networks. Because signals are often transmitted via covalent modification of protein structure, the investigators examined protein carbonylation upon exposure of yeast to copper as an experimental model to identify protein connections that shut down glycolysis in a reverse, stepwise fashion in response to copper-induced oxidative stress in yeast.

The cellular response to toxic injury often involves direct modification of proteins. The types of modifications seen include carbonylation, phosphorylation, glycosylation, sumoylation, and ubiquitination. Proteomic approaches can go beyond traditional toxicologic approaches in identifying molecular targets of injury and delineating mechanisms of toxicity.

Covalent binding has long been known to contribute to toxicity in many cases, but mechanistic insights have been limited by a lack of knowledge about protein targets of reactive intermediates (Cohen et al. 1997; Liebler and Guengerich 2005). In exploratory studies, the application of MS-based proteome analyses has revealed protein targets. Two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2D SDS-PAGE; see Chapter 2) analyses of proteins from cells and tissues treated with radiolabeled toxicants identified putative protein targets but not the mapping of actual protein adducts (Qiu et al. 1998; Lame et al. 2000). Other work has applied the combination of liquid chromatography and MS (LC-MS-MS) to map the sites of electrophile adduction on proteins. The major problem with the identification of protein adducts is the apparently low stoichiometry of modification, meaning that adducted proteins are relatively difficult to detect in complex protein mixtures. LC-MS-MS analysis of bile from rats treated with the prototypical hepatotoxin 1,1-dichloroethylene (DCE) revealed DCE-derived adducts on biliary proteins (Jones et al. 2003). Future studies with appropriate model compounds and affinity capture of adducts should help address some of these relationships.

State of the Art: Metabonomic Analyses

Metabonomic analyses use high-field nuclear magnetic resonance (NMR), gas chromatography (GC) MS, or LC-MS (see Chapter 2) to analyze complex mixtures of metabolites in biofluids and tissue samples (Nicholson et al. 2002). In this context, “metabolites” include metabolic products of xenobiotics and

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

products of endogenous metabolism, including small molecules generated from bacteria within an organism (for example, gut microflora). Assessment of toxic responses is based on relative intensities of signals corresponding to multiple metabolites, many of which can be identified (for example, by NMR chemical shifts). This allows inferences about metabolites and pathways involved in toxicities. Serial measurements enable metabolite status to be correlated with time-dependent changes in other parameters, and this allows inferences about specific metabolic pathway changes during the course of toxicity and recovery. Most studies to date have applied NMR- and MS-based metabonomic analyses to animal models of tissue-specific toxicity. These studies have described metabolite profiles characteristic of toxicities for different types of chemicals (Nicholls et al. 2001; Waters et al. 2001; Bollard et al. 2002; Coen et al. 2003; Y. Wang et al. 2005b; Williams et al. 2005).

Toxicity mechanisms are commonly envisioned at the molecular-chemical level, yet toxicity is a systems- and organ-level phenomenon. Metabonomic analyses of biofluids, such as urine, can provide mechanistic insights at the level of metabolic pathways and system dysfunction. The advantage over conventional analysis of metabolites is that multiple metabolites in a biologic network can be examined simultaneously and integrated with the responses of related biologic systems. Because urinary metabolite profiles reflect toxicities in multiple organs or tissues, metabolite profiles must be deconvoluted to extract information specific to target tissues. Waters et al. (2005) provided an example of metabolic deconvolution of toxicity signatures in studies of thioacetamide, which is both hepatotoxic and nephrotoxic. They relied on knowledge of both renal and hepatic biochemical physiology to interpret metabolite profile changes accompanying tissue injury. Many signals observed in NMR spectra or MS profiles can be confidently attributed to known compounds, and this linkage provides a systems context to interpret changes. Analysis of time-dependent changes also is critical to distinguishing primary effects from secondary adaptive responses (Keun et al. 2004; Waters et al. 2005).

Metabonomics can elucidate toxicity mechanisms even in the absence of a distinctive toxic phenotype. For example, Mortishire-Smith et al. used metabonomics to characterize mechanisms of a hepatotoxic drug. The compound increased medium-chain fatty acids and intermediates of the tricarboxylic acid cycle, an important metabolic pathway, thus implicating inhibition of fatty acid metabolism as a mechanism of toxicity (Mortishire-Smith et al. 2004). In mechanistic studies, metabonomic analyses of tissues and biofluids are perhaps most powerful when integrated with other data types, including histopathology, clinical chemistry, and transcriptome profiles or proteomics data. A study of the hepatobiliary toxicant alpha-naphthylisothiocyanate combined liver, plasma, and urine metabonomics with histopathology and clinical chemistry measurements to establish a global profile of the development of injury (Waters et al. 2001). Integration of metabonomic and transcriptome profiles from mice treated with a hepatotoxic dose of acetaminophen enabled interpretation of gene expression changes in the context of metabolic status (Coen et al. 2004). Thus, a potentially

Copyright © National Academy of Sciences. All rights reserved.

ambiguous set of gene expression changes was reconciled with a metabolic functional outcome (elevation of glycolysis) as toxicity progressed.

CONCLUSIONS

Toxicogenomic studies are improving our knowledge of the underlying biology and the regulatory networks that integrate the signaling cascades involved in toxicity. Thus, toxicogenomic data may advance the introduction of mechanistic insight into risk assessment and fulfill the promise of more accurate and expedited elucidation of class-related biologic effects or predictive toxicity. One must consider, however, that the data-rich nature of toxicogenomic technologies, coupled to challenges of data interpretation, make the application of toxicogenomics in risk assessment inherently complex and will require the implementation of educational programs for the toxicology and risk assessment communities.

A more immediate need in the field of toxicogenomics is for more accurate identification of orthologous genes or proteins across species. This effort will improve our understanding of conservation of biologic responses to toxic injury and facilitate use of surrogate species that predict the responses in humans. Although there are important differences in the genomes and proteomes, many responses to chemical and physical stressors are evolutionarily conserved and limitations posed by cross-species extrapolation can be mitigated by focusing analyses on processes conserved across species. However, many genes of rats, mice, and humans remain uncharacterized, and divergence can be a major factor in species differences in sensitivity or response.

A key goal of toxicogenomic research is to integrate data from multiple sources to produce a comprehensive understanding of the molecular basis of toxicologic responses. For this reason, there is a pressing need to develop algorithms that combine and interpret data of multiple types (for example, gene expression, proteomic, and metabonomic data). Integration of data from different technologies will lead to synergistic interpretations beyond what can be resolved when data are analyzed in isolation. Examples include the interplay between transcriptional analysis of protein factors and gene expression changes and between levels of metabolizing enzymes and the production or elimination of metabolites. The integration of data from different toxicogenomic technologies, has been explored (Hogstrand et al. 2002; Ruepp et al. 2002; Coen et al. 2004) but has yet to be fully realized.

There is also a need to develop mechanisms to better probe the complexity of toxic responses. Toxicologic responses are typically defined by a linear sequence of events. In contrast, a network and system level of organization reflects nonlinear cellular states that depict the true complexity of biologic systems. The development of a knowledge base to accurately reflect network-level molecular expression and interpretation requires a new paradigm of data management, integration, and computational modeling. As the field of toxicogenomics ad-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

vances, the development of approaches (such as the combined use of model systems and interactome analyses) to unravel the complexity inherent to biologic systems needs to be a research priority.

The application of toxicogenomic approaches to the study of toxicity has advanced our understanding of the biology that underlies the deleterious actions of chemical and pharmaceutical agents on living systems, the regulatory networks that integrate the signaling cascades involved in toxicity, and the pathogenesis of environmental or drug-induced disease. Indeed, mechanistic toxicology investigations have proven useful in risk assessment, drug development, environmental exposure assessment, and understanding of human and animal variability in response to drugs and chemicals. Progress to date has been limited by the scarcity of comprehensive time- and dose-related investigations and the lack of studies using exposure paradigms that reproduce the human condition with fidelity.

RECOMMENDATIONS

Given the status of current mechanistic toxicogenomic investigations, the following recommendations are made:

Immediate Actions

1. Develop richer knowledge bases and models that can integrate knowledge of mechanisms of toxicity and the complex networks information, encouraging the community to use these models to study toxicology as a global response. This requires a new paradigm of data management, integration and computational modeling and will specifically require the development of algorithms that combine and interpret data across multiple platforms in different animal species.
2. Make resources available to advance detailed mechanistic research that is useful for classifying toxic chemicals and to assess the public health relevance of these toxicity classifications.
3. Make resources available to facilitate the identification of orthologous genes or proteins across laboratory species, for identifying suitable surrogate species to predict the response of humans to toxic injury. In the near term, this would include the development of algorithms.

Intermediate Actions

4. Advance proteomic and metabonomic analyses by promoting the integration of peptide and metabolite separation technologies (NMR, GC-MS, LC-MS, MS with protein) into toxicologic investigations and advancing proteomic and metabonomic databases. Advancement of proteomic and metabonomic

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

analysis is necessary to fully elucidate cellular responses to toxic injury, particularly those mediated by direct modification of proteins and/or disruption of metabolic networks involved in tissue homeostasis.

5. Examine the sensitivity of current toxicogenomic-based methodologies and analyses and their ability to distinguish between endogenous and exogenous effects at biologically relevant doses.

6. Implement educational programs to help the toxicology and risk assessment communities incorporate toxicogenomic approaches and data-rich mechanistic assessments into their professional practice.

Long-Term Actions

7. When appropriate, encourage a shift in the focus of mechanistic investigations from single genes to more integrated analyses that embrace the complexity of biologic systems as a whole as well as the multidimensionality of dose- and time-related effects of toxic agents.

Copyright © National Academy of Sciences. All rights reserved.

8

Other Potential Applications of Toxicogenomic Technologies to Risk Assessment

The potential speed, lower cost, and information content of toxicogenomic technologies may offer distinct opportunities for enhancing the design and interpretation of standardized testing protocols and approaches used in risk assessment. A brief overview of risk assessment is provided in Appendix C; Chapters 4, 5, 6, and 7 describe the potential for toxicogenomics to improve exposure assessment, hazard screening, assessment of human variability, and mechanistic insight, respectively. This chapter continues the discussion of toxicogenomic applications, focusing on several topics that are important to risk assessment: understanding dose-response relationships, especially at low doses, and improving the selection of doses used in testing; relating animal model data to human risk; assessing effects of exposure during development; and assessing the relevance of coexposures and the impact of mixtures. The implications that the increased use of toxicogenomics may have on experimental animal use and on federal agency infrastructure needs is also discussed.

POTENTIAL APPLICATIONS TO RISK ASSESSMENT

Dose-Response Relationships

A critical step in assessing risk is determining how responses change depending on the magnitude and nature of exposure to the agent or agents in question. As discussed in Appendix C, the complex nature of the dose-response relationship has traditionally been considerably simplified for risk assessment (1) by using linear slopes at low doses (below the point of departure or below observed data points) for agents with direct mutagenic activity or for which exposures are

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

thought to be near levels associated with key precursor events in the carcinogenic process, and (2) by using nonlinear slopes at low doses when there are enough data to determine the mode of action and conclude the dose-response relationship is not linear or an agent is not mutagenic at low doses.

Because of their ability to detect more subtle changes at the molecular level than those detected by traditional high-dose animal studies, toxicogenomic studies, properly designed and interpreted, can provide greater insight into dose-response relationships with respect to low-dose effects and mode of action (which affects the assumed shape of the dose-response relationship). However, similar to traditional toxicology assays, most toxicogenomic investigations to date have used relatively high doses and conventional end points.

These studies assume that many of the differentially regulated genes are associated with the observed toxic effect. For example, Gant et al. (2003) identified genes induced during chronic liver injury, Amin et al. (2004) investigated renal toxicity, and Hamadeh et al. (2004) identified furan-mediated hepatotoxicity. Moggs et al. (2004) identified the genes and molecular networks associated with the uterotrophic response to estrogens. Such studies are critical for proof of concept, but, ultimately, toxicogenomic technologies will most benefit risk assessments when they provide insight into responses that occur at doses at or near anticipated population exposures—modeling at such low doses is always a challenge. Several illustrations of the potential implications of toxicogenomics for exploring dose-response issues are provided in the following sections.

Low-Dose Responses

When looking at gene expression over a range of doses, some altered genes may reflect homeostatic responses, others may be "early responders" inherent to the ultimate toxic response, and still others may represent perturbations in vital cell pathways resulting from adverse effects. To elucidate quantitative dose-response relationships, the observed changes in gene expression need to predict the toxic response and distinguish it from nontoxic responses. Although this is inherently difficult in complex biologic systems, several concepts have been proposed to help distinguish low-dose effects that predict toxicity from those that do not. They include emphasizing perturbations in critical cellular systems, such as stress responses, apoptosis, and energy production as well as assessing the magnitude of gene expression changes and the number of genes affected as doses increase (Heinloth et al. 2004).

Often a dose increase does not simply increase the magnitude of expression change in the same set of genes but also influences which genes are affected; that is, the dose increase shifts the response profile. For example, Andrew et al. (2003) compared the effects of low noncytotoxic doses with higher cytotoxic doses of arsenic on human bronchial epithelial cells and reported the expression of almost completely nonoverlapping sets of genes. There appeared to be a threshold switch from a "survival-based biologic response at low doses

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

to a death response at high doses." Thukral et al. (2005) exposed animals to two doses of two nephrotoxicants, mercuric chloride and amphotericin. Low doses resulted in damage to and regeneration of tubular epithelium, whereas high doses caused necrosis of tubular epithelium. Gene expression profiles clustered on the basis of similarities in the severity and type of pathology, and necrosis was associated with more changes in gene expression. These studies illustrate the complexity of interpreting toxicogenomic data for the purposes of defining dose-response relationships. Lower doses may cause changes that involve tissue damage and repair and predict some degree of chronic toxicity, even though they do not lead to observable pathologic changes in a given experimental setting. (Note that interpretation of changes in gene expression over the range of the dose-response curve should be supported by appropriately robust statistical methods.)

Experiments conducted at two doses, although helpful, still do not allow for construction of a full dose-response curve useful for risk assessment. To interpret the impact on gene expression, a range of doses must be investigated. Heinloth et al. (2004) measured changes in gene expression due to exposure to acetaminophen at doses ranging from those expected to have no hepatotoxic effect to those known to cause liver toxicity. The goal was to discover changes in low-dose-mediated gene expression that might indicate biologic responses predictive of the toxic effects observed with high doses. The results indicate that subtoxic doses of acetaminophen cause down regulation of genes involved in energy expenditure, coupled to the upregulation of genes involved in ATP production. Cellular systems affected included energy production and energy-dependent pathways as well as cellular stress responses. The results suggest that changes in gene expression induced by exposure to a low dose of a potentially toxic agent may reveal signs of stress or subtle injury that signal potential overt toxicity at higher doses. Thus, in this system, changes in gene expression were more sensitive indicators of potential adverse effects than traditional measures of toxicity.

Sen et al. (2005) tested several concentrations of dimethylarsinic acid on rat urothelium and similarly found a progressive increase in differentially expressed genes associated with apoptosis, cell cycle regulation, adhesion, stress response, and others over much of the dose range. Specifically, the number of genes affected increased as the dose increased from 1 part per million (ppm) to 40 ppm, with increased expression of as many as 74 genes associated with cell cycle regulation and proliferation, apoptosis, and oxidative stress. Interestingly, however, many of these genes were not expressed in animals exposed to a 100-ppm treatment, despite increased toxicity. The authors speculated that toxicity from the high 100-ppm dose likely resulted in the degradation of cellular components, including RNA, which could result in a decrease in the ability to detect altered transcript numbers. The differences may also reflect a different time course of effects after treatment with dimethylarsinic acid at 100 ppm, greater adaptive response at 40 ppm, or a U-shaped dose-response curve (as also ob-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

served in transcriptomic experiments with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) (Ahn et al. 2005).

In another study assessing gene expression over a range of doses, Boverhof et al. (2004) investigated time- and dose-dependent changes in hepatic gene expression from immature ovariectomized mice given several doses of ethynyl estradiol, an orally active estrogen. Thirty-nine of the 79 genes identified as differentially regulated exhibited a dose-dependent response at 24 hours. This study also illustrated how the non-static nature of gene expression complicates the assessment of a relationship between dose and response. That is, this study showed that administering a single dose (100 µg/kg of body weight) resulted in time-dependent changes in the expression of genes associated with growth and proliferation, cytoskeletal and extracellular matrix responses, microtubule-based processes, oxidative metabolism and stress, and lipid metabolism and transport.

A number of papers have reported effects of chemical agents on gene expression profiles at concentrations below those associated with overt toxicity. For example, Boverhof et al. (2004) found that most of the 39 transcripts that exhibited a dose-dependent effect at 24 hours were changed at median effective doses (ED_{50}) comparative to ED_{50} values reported in the literature to cause uterotrophic effects. However, a number of transcripts displayed ED_{50} values at doses that do not elicit a physiologic effect. Naciff et al. (2005a) similarly studied the effects of three estrogenically active compounds over a dose range spanning five orders of magnitude and found that the gene expression profiles provided more sensitive and less variable responses than traditional morphologic end points. In a later study, they demonstrated that all these estrogenically active substances exhibited monotonic dose-response curves, providing critical mode of action insights into the biologic plausibility, or lack thereof, of hypothesized low-dose, nonmonotonic dose-response curves for endocrine-active substances (Naciff et al. 2005a).

No-Effect Threshold

Because conventional animal bioassays generally are conducted at high doses, risk assessors have particular interest in whether toxicogenomic technologies will provide empirical evidence of adverse effects at low doses. This includes not only the shape of the dose-response curve but also whether gene expression indicates a no-effect threshold in a more straightforward manner than conventional means.

A concentration at which no transcriptional effect is observed, referred to as a no-observed-transcriptional-effect level (NOTEL), has been demonstrated for estrogen in cells from a hormonally responsive breast cancer cell line (Lobenhofer et al. 2004). The investigators measured multiple transcriptional changes in response to four concentrations of estrogen and found that only physiologically relevant doses of estrogen induced a transcriptional response, which suggests that it is possible to estimate NOTELs through gene expression

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

microarray experiments. Similarly, Naciff et al. (2005a) found low doses of three estrogenic compounds that did not elicit transcriptional responses. Determining a NOTEL may be of greatest value with well-characterized toxicants because the most sensitive species, sex, organs and cells can be identified and there has been extensive cross-referencing of in vitro and in vivo responses. However, this concept may be less useful in characterizing less-well-studied types of toxicity. In particular, great care must be taken not to assume that a NOTEL derived in one in vitro cellular system is representative of all potential whole-animal responses. It is essential that efforts to incorporate the concept of a NOTEL into regulatory decision making derive a rigorous definition of NOTEL that can be applied to many different mechanisms of toxicity as well as rigorous methods to ensure that minor changes in gene or protein expression associated with toxicity are detected against background variability.

Impact on Cross-Species Extrapolations

A critical challenge for risk assessment is to identify test species that display responses to toxicants similar to those of humans. Animals show species- and strain-specific differences in responses to toxicants and these differences are often inconsistent from compound to compound, thus greatly increasing the complexity of interpreting standard bioassay results. The application of toxicogenomic technologies to standard test species and simpler model organisms offers new opportunities to better understand the range of species' responses to exposures and to identify species-specific mechanisms that affect toxic responses.

For example, lung tumors in mice have long been studied as models for human adenocarcinoma (AC), the most frequently diagnosed human lung cancer. Stearman et al. (2005) recently compared lung tissue from the A/J mouse urethane model and human AC using gene expression analysis to quantify the degree of molecular similarity between the murine model and its human counterpart. Gene expression changes between tumor and adjacent normal lung tissue in human AC were recapitulated in the mouse model with striking concordance. More than 85% of genes had similar expression levels between adjacent tissue and tumor samples in both species, suggesting common pathobiology in AC between humans and mice. The authors highlighted the numerous similarities in orthologous genes associated with well-documented hallmarks of cancer (Hanahan and Weinberg 2000).

Similarly, fish have been used as models in cancer research for almost a century because some fish and human tumors have similar histopathologic features. Still, little has been known about the correspondence of the molecular mechanisms that drive tumorigenesis in these two phylogenetically distant species. In 2006, Lam et al. (2006) evaluated the molecular conservation between human and zebrafish liver tumors. They identified significant similarities between fish and human tumors in the expression profiles of orthologous genes

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

according to histopathologic tumor grade, thus building confidence in the zebrafish as a reliable model system for cancer research. In another study, liver tumors in zebrafish were found to recapitulate the molecular expression patterns of human liver cancers and to possess features that correlated with progressively higher grades of malignancy (Grabher and Look 2006). Similarly, Sweet-Cordero et al. (2005) and Lee et al. (2004) compared gene expression to assess the molecular relationship of mouse models of *KRAS2*-mediated liver cancer with similar human cancers. These studies demonstrate that toxicogenomic technologies could be valuable in identifying animal models that most closely correspond to human toxicity and disease.

Toxicogenomic technologies also offer the opportunity to explore the relevance to humans of toxicity findings that vary dramatically among model species. This scenario is exemplified by the dramatic multispecies and strain toxicity differences observed after treatment with dioxin-like substances (for example, TCDD, polychlorinated biphenyls), ubiquitous environmental contaminants that represent a significant concern for human health risks. Some species and strains are known to be extremely responsive to dioxin exposure, whereas others are relatively resistant. Okey and coworkers (2005) suggested that important events critical to expression of toxicity could be resolved by comparing common and dissimilar gene expression profiles in sensitive and insensitive animals.

TCDD elicits a broad spectrum of aryl hydrocarbon receptor (AhR)-mediated toxic biochemical effects, which are well correlated with its ability to bind to the AhR (Safe 1990). Studies demonstrating that mice with low-affinity AhR alleles are less susceptible to the effects of TCDD (Okey et al. 1989) and that mice lacking the AhR are resistant to prototypical toxicities elicited by TCDD and related substances (Mimura et al. 1997) support this contention. Sun et al. (2004) undertook a comparative computational scanning approach assessing gene expression in livers of human, mouse, and rat (species that differ in dioxin response). They used this approach to identify “dioxin response elements” across the three test species and thereby investigate possible explanations for the species differences in dioxin response. Results suggested that AhR-mediated gene expression may not have been well conserved across species, which could have significant implications in human risk assessment.

Despite important developments in cross-species comparisons, such as those described above, the comparison of gene expression data across species remains an arduous challenge. The DNA probes for analogous genes in two different species may have different hybridization characteristics, may represent different regions of the gene, or may probe different splice variants in each species; these and other factors may confound direct comparisons of expression across species, even for genes expressed at the same level. As a result, the data obtained are not always directly comparable and may, in fact, yield disparate results. Nonetheless, this limitation will be mitigated as genomic sequences are completed and gene annotation is improved and incorporated into microarrays. Investigators have begun to develop computational bioinformatic tools that al-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

low for the identification of common genes across microarray platforms representing different species (Mattingly et al. 2003; Thorgeirsson et al. 2006). Another possible limitation to the generation of informative toxicity data from conventional studies is that test animals and humans may not metabolize or activate the study compound in a similar fashion. This is a major drawback of the Ames bacterial genotoxicity test and necessitates using liver microsomes to represent mammalian metabolism. This issue has also limited the use of other genetically tractable organisms such as yeast. To mitigate this deficit, investigators have engineered human cell lines to constitutively express various combinations of xenobiotic-metabolizing enzymes (Crespi et al. 1997; Crespi and Miller 1999). In addition, several recent genomic studies have used genetically engineered yeast to express genes encoding xenobiotic-metabolizing enzymes such as human CYP1A1 and epoxide hydrolase (Keller-Seitz et al. 2004; Guo et al. 2005, 2006). These yeast strains were then used to study gene expression profiles induced by aflatoxin B1 (AFB1) under a variety of dosing regimens. One of these studies then exploited the power of yeast genetic manipulation to disrupt genes of pathways induced by AFB1 and ascertained DNA repair pathways critical to modulating AFB1-induced mutagenesis (Guo et al. 2005). However, whether the signatures detected in yeast or the pathways affected are also induced by AFB1 in human cells remains to be determined.

Similar approaches can also be used for genes that are not well conserved across species or to investigate the effect of human alleles of specific genes—for example, work with the paraoxonase gene (*PON1*) involved in the metabolism of organophosphate insecticides and nerve gases such as Sarin. To model human genetic variation in susceptibility to these agents, Furlong et al. (2005a,b) generated *PON1* knockout mice and used DNA microarrays to compare gene expression profiles induced in the presence and absence of the gene. They also generated knock-in mice carrying the wild-type human *PON1* or allelic variants that show enhanced susceptibility. This insertion of a variety of human xenobiotic-metabolizing genes in mouse models and subsequent measurement of gene expression in response to chemical toxins should improve the ability to extrapolate from the laboratory to potential human exposures.

Identification of susceptibility genes such as PON1 can also be integrated into PBPK models to further inform risk assessment. The toxicity of the organophosphate pesticide chlorpyrifos is impacted by PON1 activity. Using PBPK analyses, Timchalk and coworkers (2002) demonstrated that PON1 variation in humans would be expected to impact cholinesterase inhibition at high doses (for example, high-dose occupational scenarios) but would be unlikely to impact this toxicity at low environmental concentrations encountered by the general population. Thus, this approach can test whether genetic polymorphisms linked to a chemical's specific mode of action always confer a significant contribution to adverse health outcomes when environmental exposures are very low.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Developmental Effects

In utero exposures have become widely recognized as causes of developmental disorders such as fetal alcohol syndrome, X-ray-induced leukemia, and cerebral palsy-like syndrome due to high levels of methylmercury (Harada 1978; IOM 1996; Doll and Wakeford 1997). In addition, prenatal exposures have long been associated with other chronic diseases such as cancer (Birnbaum and Fenton 2003) and, more recently, with neurodegenerative diseases, asthma, cardiovascular diseases, and immune dysfunction (Holladay and Smialowicz. 2000; Osmond and Barker 2000; Peden 2000; Pinkerton and Joad 2000). Although developmental toxicity testing is required in the U.S. testing batteries for pesticides and drugs and in Europe under the REACH program, relatively little is known about the health impacts of many chemicals in current use. A number of epidemiologic and conventional toxicologic investigations of the impact of early life exposure on adult disease have been reported, but testing primarily has focused on gross morphologic changes of the offspring and reductions in reproductive capability.

One way that toxicogenomics may be useful in assessing risk from exposures that occur during development is that the near completion of mouse and human genome sequencing is advancing molecular level understanding of development by identifying all genes, including regulatory transcripts (such as microRNAs). Hence, even sequences from genes that are expressed only during development and that were historically underrepresented in cDNA libraries can now be identified with bioinformatic tools and included in gene expression microarrays. As a result, comprehensive gene expression profiling can now be performed at any developmental stage, not only providing insight into the normal regulatory networks that control development but also allowing for the analysis of how exposures to toxicants can perturb these networks and cause abnormalities.

Nemeth et al. (2005), in work with early mouse embryos exposed to various teratogens, illustrated that toxicogenomics can advance molecular level understanding of development. They identified a benchmark panel of genes for normal development of the eye, an organ considered a definitive target for the study of teratogens. The authors identified 165 genes differentially expressed during rodent eye morphogenesis, including 58 genes common to both rats and mice. The biologic significance of some of the genes and pathways affected, such as glycolysis genes crucial in maintaining oxygen levels, may provide insights into the mechanisms of teratogenesis.

Toxicogenomic studies conducted to date have focused largely on iconic teratogens, exposing animals on a key gestational day and then attempting to identify gene changes that correlate with the teratogenic action observed. For example, Hard et al (2005) investigated ethanol-induced alterations in gene expression in the fetal mouse brain in an effort to identify a genetic marker for fetal alcohol syndrome and discover the pathways involved in its origin.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Twenty-five genes were associated with ethanol exposure and teratogenic effects. In their transcriptional profiling work with valproic acid (VPA), a potent teratogen that induces neural tube defects, Kultima et al. (2004) defined a subset of VPA-responsive genes for evaluation as potential biomarkers of VPA teratogenicity. Their study also highlighted some potential challenges associated with analyses conducted in embryos, including limited amounts of tissue and complex mixtures of cell types. In addition, alterations in critical gene targets may be difficult to observe because removing cells from the fetus for analysis may ablate the genomic events of interest. Although it is theoretically possible to obtain sufficient starting material (for example, mRNA) from a single cell, it is often not practical and sampling bias can easily arise because the "pooling" of multiple nonresponsive and noncritical target cells will diminish the signal-to-noise ratio (Kultima et al. 2004).

Working with three endocrine disruptors of different potencies, Naciff et al. (2002) sought to identify a gene expression signature to detect estrogenic effects on development. They identified a common set of genes whose expression was significantly and reproducibly modified by each of the chemicals tested. The products of these genes were plausible targets of endocrine disrupters, including, for example, genes encoding steroidogenesis products important to gonadal differentiation.

Toxicogenomics is also an attractive approach to uncover critical molecular events altered by developmental toxicants. This is because the disruption of signaling and gene regulatory networks that control embryonic development is likely to underlie many cases of birth defects. For example, VPA has been subjected to a weight-of-evidence MOA analysis that illustrates how toxicogenomic and other conventional data inputs can be merged to improve confidence in decisions evaluating whether findings of VPA-induced teratogenicity present plausible concern for potential induction of spina bifida in exposed humans (Wiltse 2005). This analysis concluded that VPA-induced alteration in WNT[1]-dependent gene expression in both animal and human cells represented a critical MOA event causing in vivo developmental effects. This conclusion was further supported by integrating supplementary dose-response information; classic enzymatic, biochemical, and pharmacologic studies; and outcomes of cross-species studies.

Understanding Effects of Chemical Mixtures

Standardized animal testing protocols are largely intended to assess the toxicity and potential risks of single chemicals. Data from these types of studies alone provide little evidence of potential adverse interactions that may occur as a result of pharmaceutical or environmental chemical coexposures. Clinical ex-

[1]Wnt proteins are a family of secreted signaling molecules that regulate cell interactions during embryogenesis (http://www.stanford.edu/~rnusse/wntwindow.html).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

perience has clearly demonstrated the potential for adverse drug interactions, which may be mediated via competition for common metabolic activation or detoxification pathways, pharmacologic interactions, unanticipated idiosyncratic responses, or other mechanisms.

Regardless of the nature of the interaction, gene expression data collected during preclinical drug development studies may afford an improved, early method for identifying such potential interactions occurring at intended doses. For example, toxicogenomic technologies been used to explore the idiosyncratic hepatotoxic response associated with ranitidine, a therapeutic histamine-2 receptor antagonist used to treat ulcers and gastroesophageal reflux disease (Luyendyk et al. 2004). Using hierarchal clustering of gene expression data from rat microarrays, these investigators identified key transcriptional responses that appear to account for ranitidine-induced injury associated with a subclinical interaction with bacterial lipopolysaccharide (endotoxin).

However, the identification of potential adverse outcomes surrounding co-exposures to environmental chemicals presents challenges, primarily because typical human exposures are well below doses used in conventional toxicity studies for many compounds (Teuschler et al. 2002). Furthermore, human chemical exposures are by no means limited to chemicals of human origin; natural products also may play a role in multiagent exposures. Thus, future toxicogenomic studies examining the potential health effects of exposures to mixtures must address the confounding overlay of toxicogenomic signals associated with the range of substances present in diets and other environmental exposures. A critical element of this challenge is to determine how an exogenous exposure may truly be differentiated beyond the background pattern and variability of toxicogenomic expression associated with everyday "normal" activities such as dietary patterns, physical activities, disease status, and other lifestyle circumstances.

It is unlikely that toxicogenomic signatures will be able to decipher all interactions among complex mixtures of toxicants, diet, drugs, natural compounds, and other environmental exposures, but it should be possible to use available mechanism-of-action data to design informative toxicogenomic experiments. For example, knowledge of the mechanism of action for a chemical should make it possible to rapidly screen other chemicals for potential points of biologic conversion (overlap) such as shared activation and detoxification pathways, enhancing identification and exploration of potential interactions and moving beyond empirical experiments. Despite the lack of clear MOA understanding of how chemicals may interact under conditions of low-dose environmental exposures, the Environmental Protection Agency, acting under legislative mandates (Superfund Act 1980, FQPA 1996, Safe Drinking Water Amendments 1996), ruled that assessors must consider "how to add" the risk implications of complex environmental mixtures (Wilkinson et al. 2000). The primary methods used in regulatory practice are dose addition and response addition. Dose addition sums the exposures of the individual compounds and then combines the sums into a summed risk. Response addition calculates the risk for each compound in the

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

mixture and then sums the risks to produce a cumulative risk estimate (Wilkinson et al. 2000; Teuschler et al. 2002). These approaches rely on simple addition and do not incorporate synergism and activism—key considerations for interactions (Teuschler et al. 2002).

Valuable knowledge may come from using toxicogenomic technologies to test these approaches and other basic assumptions that underpin current risk assessment models for mixtures. For example, assumed health risks of environmental mixtures of dioxin-like compounds (for example, TCDD, tetrachlorodibenzofurans, polychlorinated biphenyls) are based on the central hypothesis that the individual toxicities of each of these compounds in a mixture can be cumulatively related to each other through a common mechanism of action involving activation of the AhR (EPA 1986; Safe 1990). This approach assumes that each compound in the mixture can be assigned a toxic equivalency factor (TEF) based on its comparative potency to the standard reference chemical TCDD. The total amount of dioxin-like toxic equivalents (TEQs) present in the mixture can then be calculated based on the amounts of individual chemicals present in the mixture. Toxicogenomic approaches afford promising opportunities to test the fundamental biologic assumptions underlying this TEF/TEQ methodology. If dioxin-like chemicals are truly toxicologically equivalent, then gene expression patterns would be expected to have critical commonalities accounting for AhR-driven toxicity responses. These types of informative investigations are in progress (Budinsky 2005).

Toxicogenomic data can also be used to study interactions among compounds that are not structurally related. Studies could be used to determine whether the gene expression, proteomic, or metabolomic patterns induced by individual chemicals overlap, whether they increase in proportion to the sum of compounds added, or whether they suggest new mechanisms of toxicity.

IMPLICATIONS FOR BETTER USE AND POTENTIAL REDUCTION IN THE USE OF EXPERIMENTAL ANIMALS

As societal views on animal welfare are increasingly concerned with ethical and humane treatment of animals—in particular vertebrates—governments have responded with legislation. Legislation affecting research and drug testing seeks to minimize the use of animals in research and eliminate statistically underpowered studies that cannot lead to valid conclusions. There is pressure to develop alternative techniques that replace the use of animals altogether, reduce the number of animals used, and refine study designs to minimize distress to the animals.

Over the past 20 years, companies, universities, and other research institutions have contributed substantial time and resources to developing alternative methods and reducing their reliance on animal studies. Although a large number of alternative in vitro tests and computational approaches are available and can provide useful information, most do not accurately model the complexity of in

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

vivo toxicologic or pharmacologic responses. Since its inception, toxicogenomics has been proposed as a possible way to reduce animal testing (Nuwaysir et al. 1999; Zarbl 2001). It is a challenging task, but if toxicogenomics can be used to predict long-term effects (such as reproductive toxicity, teratogenicity, and carcinogenicity) from short-term animal studies, or even from only in vitro toxicogenomic studies, such approaches might significantly reduce the number of animals needed for conventional tests and possibly reduce the morbidity and mortality to which test animals are subjected. In addition, if a predictive toxicogenomic signature could be detected during early gastrointestinal damage in animals, extreme gastrointestinal ulceration and bleeding could be limited. Indeed, there is a genomics study analyzing fecal material to identify an early biomarker of gastrointestinal damage (Searfoss et al. 2003).

Another potential use of toxicogenomics is the possibility of mechanistically linked surrogate markers. For example, if an exposure known to target an organ elicits a toxicogenomic signature in a surrogate tissue or compartment such as skin or blood, then use of less invasive procedures for assessing toxicity in animals becomes a real possibility. Investigators are exploring the use of blood lymphocytes or proteins present in serum and urine as surrogates for monitoring damage in target tissue organs.

IMPLICATIONS OF TOXICOGENOMICS FOR RISK ASSESSMENT INFRASTRUCTURE NEEDS

A joint report developed by the Society of Toxicology and the Society of Environmental Toxicology and Chemistry pointed out that use of toxicogenomic information in risk assessment decision making will clearly be hindered if the parties ultimately responsible for decisions (regulatory agencies) do not have adequate resources and expertise to confidently analyze and interpret submitted data (Bus et al. 2007). In addition, adequate research and associated expertise capabilities must also be present within agencies so they can better "own" critical elements of the technology being suggested for use in risk analysis. Regulatory agencies historically have dealt with other new technologies, such as physiologically based pharmacokinetic models (see Chapter 4 and Chapter 6), by developing internal capabilities supportive of the technology. However, implementation of toxicogenomic technology within the agencies will likely prove challenging because of the very large datasets and complex bioinformatic needs required for data analysis and interpretation. Practically, until such issues and hurdles are addressed, implementation of toxicogenomic data into risk assessment evaluation may be slowed or even rejected.

As the flow of toxicogenomic data rapidly increases, there will be a need not only for corresponding technical expertise within the agencies but also for enhanced training of risk assessors and risk managers in the fundamentals of the technology. Because of the wide range of stakeholders in agency risk decisions, such individuals must be able to clearly understand and articulate the processes

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

by which toxicogenomic information is incorporated and applied to risk decisions. Additional important challenges include the need to achieve consistency across regulatory agencies as to how they review, interpret, and communicate toxicogenomic data, and to maintain transparency and shared learning experiences with the external research and regulated communities.

CONCLUSIONS

It is not yet appropriate to rely solely on toxicogenomic technologies (such as the applications described in earlier chapters) to support risk decisions. However, it is clear these technologies will address some of the most vexing challenges facing risk assessment. These challenges include (1) establishing the relevance of animal dose-response data to actual human exposures; (2) understanding the relevance to human risk of different responses in different animal models; (3) identifying and establishing the significance of key factors that may confer particular susceptibilities to chemical exposures, including sensitivity to developmental toxicity; (4) understanding risk implications of exposures to complex, low-dose environmental mixtures; and (5) the need to refine, reduce, or replace the use of whole animal studies in toxicology testing.

The collection of dose-response information, appropriately linked to time, will be essential to fully integrate toxicogenomics into risk assessment decision making. To effectively address risk questions associated with human exposures to environmental chemicals, which may be much lower than doses currently used in toxicology studies, special attention must focus on characterizing toxicogenomic responses at low doses. Conducting toxicogenomic studies over a range of low doses may consume considerable laboratory, expertise, and financial resources. These studies may be more valuable when incorporated into traditional toxicity testing programs where toxicogenomic results can be tied to conventional toxicity responses.

RECOMMENDATIONS

1. Develop and expand research programs specifically dedicated to integrating toxicogenomics into challenging risk assessment problems. The development of partnerships among regulatory agencies and private sector stakeholders to incorporate toxicogenomic approaches and data into risk assessments should be encouraged to ensure the most rapid development of useful examples of toxicogenomics in risk assessment. Examples of important research areas include those discussed in the conclusions section of this chapter and in Chapters 4, 5, 6, and 7.

2. Future toxicologic assessment should incorporate dose-response and time course analyses appropriate to risk assessment. An analysis of known toxic compounds that are well characterized would provide an intellectual framework for future studies.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

3. Continue to use toxicogenomics to study differences in toxicant responses between animal models. These results will afford valuable opportunities to extend knowledge of how to effectively translate animal model observations into credible estimates of potential human risk. If toxicogenomics can illustrate how currently known interspecies differences in toxicity can be more rapidly and clearly explained, it will offer the potential to significantly enhance the confidence in animal-to-human toxicity extrapolations that constitute a foundational element of risk evaluations.

4. Use toxicogenomics to investigate how exposure during early development conveys susceptibility to drug and chemical toxicities. Efforts to develop mode-of-action data to clarify the need to apply either specific or universal default uncertainty factors in addressing susceptibility concerns, and to supplant such defaults with data- and principle-driven alternatives, will be of great value.

5. Use toxicogenomic approaches to test the validity of methods for estimating potential risks associated with mixtures of environmental chemicals. Investigations examining responses in the range of relevant exposures and low doses will be particularly valuable.

6. Invest in research and expertise within the infrastructure of regulatory agencies, as well as active collaboration across agencies, to enable toxicogenomic approaches to be effectively and credibly integrated into risk assessment practice. Transparent and participatory mechanisms for educating and engaging the scientific, regulatory, and public communities will also be needed.

7. Support education and training programs at the doctoral and postdoctoral levels to train a new generation of risk assessors fluent in the science of toxicogenomics.

Copyright © National Academy of Sciences. All rights reserved.

9

Validation

The utility of toxicogenomic technologies ultimately depends on how reliable, reproducible, and generalizable the results are from a particular study or individual method of analysis. Moving beyond laboratory assays to more widespread use requires some level of validation, which can be defined as the process of ensuring that a test reliably measures and reports the determined end point(s) and encompasses both technical and platform qualification in addition to biologic qualification. Distinct issues arise from the use of any novel technology in a regulatory context. As discussed in this chapter, validation is an integral part of the more general process of developing and applying toxicogenomic methodology.

LEVELS OF VALIDATION

Validation must be carried out at various levels as described in Box 9-1. First, technology platforms must be shown to provide consistent, reliable results, which includes assessment of device stability and determination of analytical sensitivity and assay limits of detection, interference, and precision (reproducibility and repeatability). Second, the software used to collect and analyze data for an application must provide valid results. Third, the application, consisting of both hardware and software, must be tested and validated in the context of the biologic system to which it will be applied. Fourth, the application, or a related application based on the original, must be shown to be generalizable to a broader population or to be highly specific for a smaller, target population. Finally, one must consider how these technologies and applications based on them can be validated for regulatory use. These five levels of validation are discussed in this chapter.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 9-1 Validation of Toxicogenomic Applications

1. *Platform validation*: Does the particular technology provide reproducible and reliable measurements?
2. *Software/data analysis validation*: Is the software used for analysis of a particular experimental design appropriate and does it provide insight into the biology of the problem under study?
3. *Biologic validation*: Are the results from an "-omics" analysis consistent with the biology or can they be verified by another focused approach such as quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) for microarrays or enzyme-linked immunosorbent assay (ELISA)[1] for proteomics?
4. *Generalizability*: Can the results of a particular analysis be extended from the test samples to the broader population or from animal models to humans?
5. *Regulatory validation*: Is a particular assay or test suitable for use in evaluating the safety and efficacy of new compounds or in diagnostic or prognostic applications?

It is important to recognize that validation is an iterative process, so that, for example, the biologic validation step can refine platform and software validation and help direct efforts to generalize the results.

Platform Validation

Any toxicogenomic study is predicated on the assumption that the technologies provide accurate and relevant measures of the biologic processes underlying what is being assayed. For transcriptomic profiles with microarrays, for which we have the most data, there have been many successful applications, often with high rates of validation using an alternative technology such as Northern analysis or quantitative reverse transcriptase polymerase chain reaction (qRT-PCR); however, it should be noted that each of these techniques has experimental biases. The issue of concordance between different microarray platforms was discussed in Chapter 2. However, recent reports suggest that adherence to good, standard laboratory practices and careful analysis of data can lead to high-quality, reproducible results in which the biology of the system under study drives the gene expression profiles that are observed (Bammler et al. 2005; Dobbin et al. 2005; Irizarry et al. 2005; Larkin et al. 2005). Similar efforts

[1]ELISA is a quantitative in vitro test for an antibody or antigen in which the test material is adsorbed on a surface and exposed to either a complex of an enzyme linked to an antibody specific for the antigen or an enzyme linked to an anti-immunoglobulin specific for the antibody followed by reaction of the enzyme with a substrate to yield a colored product corresponding to the concentration of the test material. (Merriam-Webster's Medical Dictionary, http://dictionary.reference.com/browse/Enzyme-Linked Immunosorbent Assay, [Accessed: April 12, 2007]).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

must accompany the adoption of various genomic, proteomic, and metabolomic technology platforms for toxicogenomics.

This process of technology platform assessment is an essential step in the overall validation process and indicates whether a system provides a reliable and reproducible measure of the biology under study. Two often-confused measures of system performance are repeatability and reproducibility. *Repeatability* describes the agreement of successive measurements when controllable sources of variation are held constant. If one or more of these sources of variation is allowed to have its typical effect, the agreement is called *reproducibility*.

Repeatability can be assessed by consecutive measurements of a single sample at one time under identical conditions. Repeated measurements with the same method but conducted on different days, with different batches of reagents, or with different operators provide a measure of reproducibility. Because the latter scenario best describes the routine application of toxicogenomic technology platforms, reproducibility is the most relevant measure of system performance.

Assays of reproducibility involve analyzing the same biologic sample multiple times to determine whether the platform provides consistent results with a small coefficient of variation. Although this may seem straightforward, toxicogenomic technologies do not measure single quantities but represent hundreds or thousands of measurements—one each for many genes, proteins, or metabolites. Assays optimized for one range of expression level or type of analyte may not perform as well with other samples. For example, a technology that performs well for genes expressed at high levels may not be sensitive to low levels of expression. Careful assessment of the reproducibility of measurements and the relative signal-to-noise ratio in the assay must be evaluated with an emphasis on relevant levels of gene, protein, or metabolite expression for a specific application. This is particularly true of proteomics and metabonomics, in which the range of analyte concentrations may vary by more than a millionfold (Figure 9-1).

Types of Calibration Standards

Another approach that can provide some level of quality assessment and quality control is the use of calibration standards that consist of complex mixtures of analytes spanning the dynamic range normally surveyed in a particular application. In the context of microarray gene expression analysis, the development of "universal" RNA reference samples (Cronin et al. 2004) is under way and the External RNA Control Consortium (ERCC), led by the National Institute of Standards and Technology, is moving toward defining such a standard. The ERCC is composed of representatives from the public, private, and academic sectors working together in a consensus fashion to develop tools for experiment control and performance evaluation for gene expression analysis, including

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

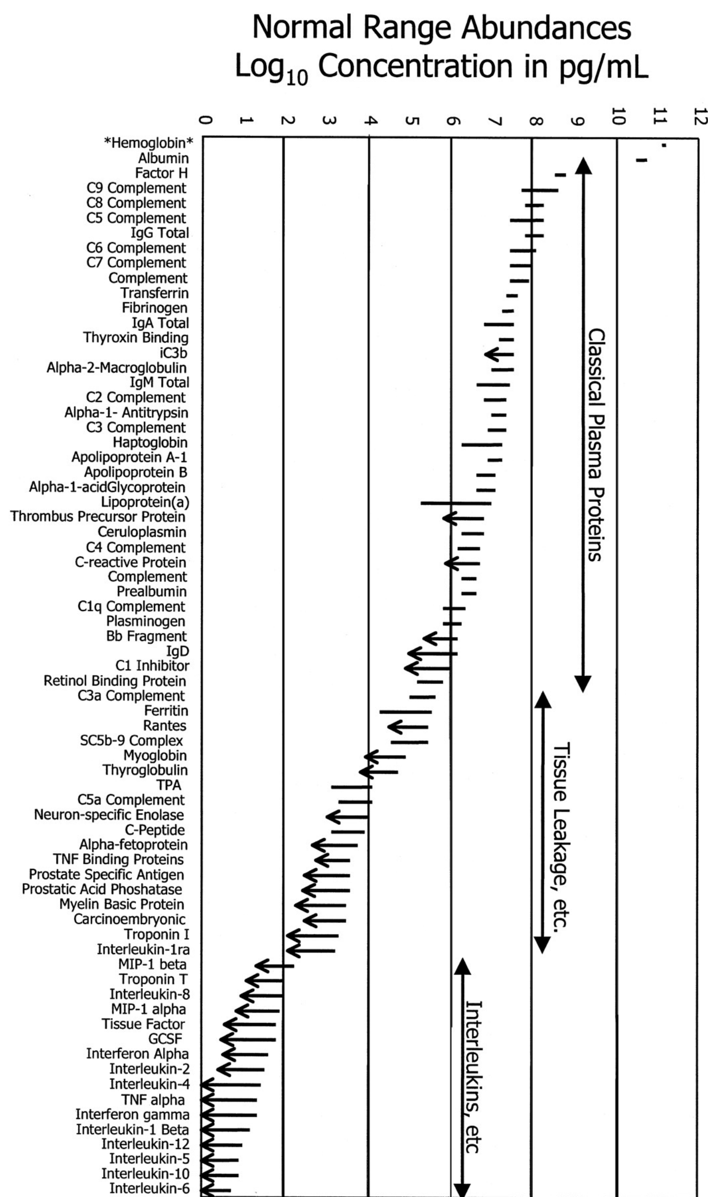

FIGURE 9-1 Human plasma proteome. The large range of protein concentrations in the human proteome represents a significant experimental challenge as technologies must be sensitive across nearly 12 orders of magnitude (a 1 trillionfold range) for comprehensive analysis and the development of biomarkers. Source: Anderson and Anderson 2002. Reprinted with permission; 2002, *Molecular & Cellular Proteomics*.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

"spike-in" controls, protocols, and informatic tools—all intended to be useful for one- and two-color microarray platforms and qRT-PCR.

Ideally, such an RNA "standard" would consist of multiple samples. The first would consist of one or more RNA mixtures that could be used for regular quality control assessment. This approach can be used to monitor the performance of a particular laboratory or platform to document that the results obtained remained consistent both in their ability to detect expression measures for any one sample and their ability to detect differential expression among samples. A second useful control RNA would consist of exogenous spike-in controls (van de Peppel et al. 2003), which correspond to probes on the microarray surface that are not from the species being analyzed. This control measures system performance independent of the quality of the RNA sample being analyzed. Objective measures of RNA quality may provide an additional means of assessing the performance and establishing the credibility of a particular microarray assay, as poor quality RNA samples provide unreliable results. Finally, as the primary measurement in microarray assays is the fluorescence intensity of individual hybridized probes, work is under way to establish quantitative standards to assess the performance of microarray scanning devices.

Efforts at Standards Development

Consortia approaches to standardization and validation have played an important role in working toward platform validation. As the field of toxicogenomics has matured, there has been a realization that groups working together can better understand and define the limitations of any technology and the potential solutions to any problems. Examples include the International Life Sciences Institute's Health and Environmental Sciences Institute (ILSI-HESI 2006), a consortium of industry, government, and academic groups examining applications of toxicogenomics, and the Toxicogenomics Research Consortium sponsored by the National Institute of Environmental Health Sciences (NIEHS) (TRC 2005). The value of these consortium efforts is that they capture the state-of-the-art of multiple groups simultaneously and therefore have the potential to advance an adoptable standard much more quickly than can individual research groups.

In response to the growing need for objective standards to assess the quality of microarray assays, the Microarray Gene Expression Data Society (MGED) hosted a workshop on microarray quality standards in 2005. This workshop and its findings are described in Box 9-2.

While early, informal efforts such as those offered by MGED are important in helping to define the scope of the problem and to identify potential approaches, systematic development of objective standards for quality assessment would greatly facilitate the advancement and establishment of toxicogenomics as a discipline. Ideally, further efforts to establish objective and quantitative

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 9-2 Microarray Gene Expression Data Society Workshop, September 2005, Bergen, Norway

The purpose of this workshop was to examine quality metrics that cut across technologies and laboratories: defined standards that can be used to evaluate various aspects of each experiment, including overall studies, the individual microarrays used, and the reporters that are used to measure expression for each gene. Establishing such standards will allow data in public repositories to be better documented, more effectively mined and analyzed, and will provide an added measure of the confidence of the results in each study. The workshop examined several levels of quality assessment:

1. Quality measures based on external controls, including spike-in controls;
2. Indicators of common artifacts such as background levels, background inhomogeneity, and RNA degradation;
3. Quality metrics based on technical replicates;
4. Model-based quality metrics (using biologic replicates and statistical models of data distributions); and
5. Identification of potential bias in individual measures and evaluation of data.

The consensus was that the diversity in platforms, experimental designs, and applications makes it unlikely that a single universal measure of quality will be possible. However, there was confidence that standards based on universal principles could be developed for each platform—for example, one for Affymetrix GeneChips and a separate, but similar, standard for spotted oligonucleotide microarrays.

Basic principles applicable at different stages of any study were evident. First, external standards as spike-in controls provide a way to assess various steps of the analytical process and facilitate comparisons among laboratories and platforms; they also provide a way to assess the quality of experiments over time. Adding a control RNA from another species without similarity to the genome of the species under study, and for which there are probes on the microarray, can yield data that can be used to assess the overall quality of the starting RNA, the quality of the hybridization, and the general quality of a particular microarray. Second, as most microarrays contained repeated probe sequences for particular genes, they are quite useful for assessing spatial properties of each microarray and for assessing the overall performance of a single microarray as these repeated probes should give consistent measures of gene expression.

Finally, the analysis of replicate samples was identified as one way to assess the quality of a particular study. This requires “technical replicates,” in which the starting material is measured multiple times and which provides an estimate of the overall variability of any assay. However, “biologic replicates,” in which separate experimental subjects from the same treatment group are assayed also provide a very useful and powerful assessment of gene expression, as they include estimates of variability in both the assay and the biologic samples under study but as such are not as useful in assessing quality.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

quality measures for microarray data and other toxicogenomic data will help to advance the field in much the same way that "phred quality scores" characterizing the quality of DNA sequences accelerated genome sequencing.

Software/Data Analysis Validation

The software used in analyzing data from toxicogenomic studies can play as significant a role in determining the final outcome of an experiment as the technology platform. Consequently, considerable attention must be paid to validating the computational approaches. The combination of technology platform data collection and processing algorithms must be appropriately selected and validated for application to the biologic system for each study.

Data Collection and Normalization

Most genomic technology platforms do not perform absolute quantitative measurements. For microarray data collected on an Affymetrix GeneChip, the data from the multiple probe pairs for each gene are combined in various ways to assess an expression level for each gene. However, these analyses do not measure quantities of particular molecules directly; instead, they measure surrogates such as fluorescence, which is subject to unknown sources of variation. These fluorescent signals are used to estimate expression levels. Similar processing of the raw data is an element of all genomic technologies. For gene expression-based assays, these initial measurements are often followed by a "normalization" process that adjusts the individual measurements for each gene in each sample to facilitate intersample comparison. Normalization attempts to remove systematic experimental variation in each measurement and to adjust the data to allow direct comparison of the levels of a single gene, protein, or metabolite across samples. Despite widespread use of image processing and data normalization in microarray analyses, the sources of background signals and how to best estimate their levels are not fully understood; thus, there is no universally accepted standard for this process. Any image processing and normalization approach changes the data and affects the results of further analysis.

The importance of appropriate methods is emphasized here. If the results of multiple studies are to be combined, every effort should be made to apply consistent methods to all data. As the number of microarray experiments grows, there is increasing interest in meta-analyses, which may provide more broadly based information than can be seen in a single experiment. It is important that repositories for toxicogenomic experiments make all "raw" data available, so they can be analyzed by consistent methodologies to extract maximum information.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Data Analysis

Once the data from a particular experiment have been collected and normalized, they are often further analyzed by methods described in Chapter 3. Class discovery experiments typically use approaches such as hierarchical clustering to determine whether relevant subgroups exist in the data.

Class prediction and classification studies, which link toxicogenomic profiles to specific phenotypic outcomes represent a somewhat different validation challenge. Ideally, to validate a classification method, it is most useful to have an initial collection of samples (the training set) that can be analyzed to arrive at a profile and an appropriate classification algorithm as well as an independent group of samples (the test set) that can be used to verify the approach. In practice, most toxicogenomic studies have a limited number of samples and all of them are generally necessary for identifying an appropriate classification algorithm (or classifier). An alternative to using an independent test set, albeit less powerful and less reliable, is to perform *leave k out cross-validation* (LKOCV) (Simon et al. 2003). This approach leaves out some subset k of the initial collection of N samples, develops a classifier using the $(N - k)$ samples that remain, and then applies the classification algorithm to k samples in the test set that were initially left out. This process is then repeated with a new set of k samples to be left out and classified and so on. The simplest variant, which is often used, is *leave one out cross-validation* (LOOCV).

This cross-validation can be extremely useful when an independent test set is not available, but it is often applied inappropriately as a partial rather than a full cross-validation, the distinction being the stage in the process when one leaves k out. Many published studies with microarray data have used the entire dataset to select a set of classification genes and then divided the samples into k and $(N - k)$ test and training sets. The $(N - k)$ training samples are used to train the algorithm, which is tested on the k test samples. The problem is that using all the samples to select a classification set of genes has the potential to bias any classifier because the test and training sets are not independent. Such partial cross-validation should never be performed. The proper approach is to conduct full LKOCV in which the sample data are divided into training and test sets before each round of gene selection, algorithm training, and testing. When iterated over multiple rounds, LKOCV can be used to estimate the accuracy of the classification system by simply averaging the complete set of classifiers. However, even here, optimal validation of any classifier requires a truly independent test set.

The choice of samples for training is an important, but often neglected, element in developing a classifier. It is important to balance representation of sample classes and to ensure that other factors do not confound the analysis. Nearly all algorithms work by a majority consensus rule. For example, if the data represent two classes, A and B, with eight in class A and two in class B, the simplest classifier would just assign everything to class A with 80% accuracy, a result that clearly is not acceptable. Samples should also be selected so that there

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

are no confounding factors. For example, an experiment may be conducted to develop a classifier for hepatotoxic compounds. If all toxicant-treated animals were treated with one vehicle, whereas the control animals were treated with another vehicle, then differences between the treated and control groups may be confounded by the differences in vehicle response. The solution to this problem is a more careful experimental design focused on limiting confounding factors.

Selecting a sample of sufficient size to resolve classes is also an important consideration (Churchill 2002, Simon et al. 2002, Mukherjee et al. 2003). Radich and colleagues recently illustrated one reason why this is so important; they analyzed gene expression levels in peripheral blood and demonstrated significant, but reproducible, interindividual variation in expression for a relatively large number of genes (Radich et al. 2004). Their study suggests that a small sample size *may* lead to biases in the gene, protein, or metabolite selection set because of random effects in assigning samples to classes.

Biologic Validation and Generalizability

Regardless of the goal of a particular experiment, its utility depends on whether its results can be biologically validated. Biologic validation is the process of confirming that a biologic change underlies whatever is detected with the technology platform and assigning a biologic context or explanation to an observed characteristic of a system. The results from any analysis of toxicogenomic data are generally considered a hypothesis that must be validated by more well-established and lower-throughput "standard" laboratory methods. A first step is often to verify the expression of a select set of genes in the original samples by an independent technique. If the results can be shown to be consistent with results from an independent technique, further detailed study is often warranted. For example, upregulation of a gene transcript detected with a microarray may suggest activation of a specific signaling pathway—activation that, in addition to being confirmed by a change in the level of a corresponding protein, can be confirmed by measuring a change in another output regulated by the pathway. In a toxicogenomic experiment, in which thousands of genes, proteins, or metabolites are examined in a single assay, biologic validation is important because there is a significant likelihood that some changes in genes, proteins, or metabolites are associated with a particular outcome by chance.

For mechanistic studies, biologic validation also requires clearly demonstrating a causative role for any proposed mechanism. For class discovery studies in which new subgroups of compounds are identified, biologic validation also typically requires demonstrating some tangible difference, however subtle, among the newly discovered subgroups of compounds. For example, new paths to neurotoxicity may be inferred through transcriptome profiling when neurotoxic compounds separate into groups based on the severity of the phenotype they cause, the time of onset of the phenotype, or the mechanism that produces

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

the phenotype. Finding such differences is important to establish the existence of any new classes.

Generalizability addresses whether an observation from a single study can be extended to a broader population or, in the case of studies with animal models, whether the results are similar across species. For mechanistic studies, generalization requires verifying that the mechanism exists in a broader population. In class discovery, if results are generalizable, newly discovered classes should also be found when the sample studies are extended to include a broader, more heterogeneous population of humans or other species than those used in the initial study. For classification studies, generalization requires a demonstration that, at the least, the set of classification genes and associated algorithms retain their predictive power in a larger, independent population and that, within certain specific populations, the classification approach retains its accuracy (for example, see Box 9-3).

Validation in a Regulatory Setting

Toxicogenomic technologies will likely play key roles in safety evaluation and risk assessment of new compounds. Toxicogenomic technologies for this application must be validated by regulatory agencies, such as the Environmental Protection Agency (EPA), the Food and Drug Administration (FDA), and the Occupational Safety and Health Administration (OSHA), before they can be used in the decision-making process. The procedures by which regulatory validation occurs have traditionally been informal and ad hoc, varying by agency, program, and purpose. This flexible validation process has been guided by general scientific principles and, when applied to a specific test, usually involves a review of the available experience with the test and an official or unofficial interlaboratory collaboration or "round robin" to evaluate the performance and reproducibility of the test (Zeiger 2003).

Deciding whether to accept a particular type of data for regulatory purposes depends on more than such technical validation, however, and is also affected by a regulatory agency's statutory mandate, regulatory precedents and procedures, and the agency's priorities and resources. Because these factors are agency specific, validation is necessarily agency specific as well. The scientific development and regulatory use of toxicogenomic data will be facilitated by harmonization, to the extent possible, of data and method validation both among U.S. regulatory agencies and at the international level. However, harmonization should be a long-term goal and should not prevent individual agencies from exploring their own validation procedures and criteria in the shorter term.

Toxicogenomic data present unique regulatory validation challenges both because such data have not previously been used in a regulatory setting and because of the rapid pace at which toxicogenomic technologies and data are developing. Therefore, regulatory agencies must balance the need to provide criteria

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 9-3 Clinical Validation of Transcriptome Profiling in Breast Cancer

Although toxicogenomic technology applications are still in their infancy, they are being explored in clinical medicine. A notable example that illustrates the path from genomic discovery to biologic and clinical validation is the Netherlands breast cancer study (van't Veer et al. 2002), which sought to distinguish between patients with the same stage of disease but different response to treatment and overall outcome. The investigators were motivated by the observation that the best clinical predictors for metastasis, including lymph node status and histologic grade, did not adequately predict clinical outcome.

As a result, many patients receive chemotherapy or hormonal therapy regardless of whether they need the additional treatment. The goal of the analysis was to identify gene expression signatures that would help determine which patients might benefit from adjuvant chemotherapy. By profiling gene expression in tumors from 117 young patients who had received only surgical treatment and evaluating correlations with clinical outcome, the authors identified 70 genes that compose a "poor prognosis" signature predictive of a short interval to distant metastasis in lymph-node-negative patients.

Initial analysis demonstrated that microarray-based gene expression signatures could outperform any clinically based predictions of outcome in identifying patients who would benefit most from adjuvant therapy. These initial results motivated a more extensive follow-up study (van de Vijver et al. 2002) involving 295 patients (the initial 117 patients along with 178 who had been heterogeneously treated) that confirmed the advantage of the 70-gene classification profile relative to standard clinically and histologically based criteria.

These promising results were independently demonstrated to have utility in a study involving 307 patients from five European centers (Buyse et al. 2006). It has been recognized, however, that a large-scale clinical trial is necessary to fully validate the efficacy of this gene expression signature in predicting outcome. The large collaborative MINDACT (Microarray in Node-Negative Disease May Avoid Chemotherapy Trial) conducted by the Breast International Group and coordinated by the European Organisation for the Research and Treatment of Cancer Breast Cancer Group will recruit 6,000 women with node-negative early-stage breast cancer to investigate the benefit-to-risk ratio of chemotherapy when the risk assessment based on standard clinicopathologic factors differs from that provided by the gene signature (Bogaerts et al. 2006). By extending the analysis to a broader population, the researchers involved will also have the opportunity to determine the extent to which patient heterogeneity, treatment protocols, variations in sample collection and processing, and other factors influence the value of the gene expression signature as a diagnostic tool.

and standardization for the submission of toxicogenomic data with the need to avoid prematurely "locking-in" transitory technologies that may soon be replaced with the next generation of products or methods. Regulatory agencies have been criticized for being too conservative in adopting new toxicologic

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

methods and data (NRC 1994). Consistent with this pattern, such regulatory agencies as EPA and FDA to date have been relatively conservative in using toxicogenomic data, partly due to the lack of validation and standardization (Schechtman 2005).

Although some caution and prudence against premature reliance on unvalidated methods and data is appropriate, agencies can play a critical role and must actively encourage the deployment of toxicogenomic data and methods if toxicogenomic approaches are to be used to their fullest advantage. For example, FDA has issued a white paper describing a "critical path to new medical products" (FDA 2005b) that acknowledges the role of toxicogenomic technologies in providing a more sensitive assessment of new compounds and suggests that new methods be developed to improve the process of evaluation and approval.

EPA and FDA have adopted initial regulatory guidances that seek to encourage toxicogenomic data[2] submissions (see Table 9-1 and Chapter 11). In March 2005, FDA issued a guidance for industry on submission of pharmacogenomic data (FDA 2005a). In that guidance, FDA states that "[b]ecause the field of pharmacogenomics is rapidly evolving, in many circumstances, the experimental results may not be well enough established scientifically to be suitable for regulatory decision making. For example: Laboratory techniques and test procedures may not be well validated" (FDA 2005a, p. 2).

The FDA Guidance describes reporting requirements for "known valid" and "probable valid" biomarkers. A known valid biomarker is defined as "[a] biomarker that is measured in an analytical test system with well-established performance characteristics and for which there is widespread agreement in the medical or scientific community about the physiologic, toxicologic, pharmacologic, or clinical significance of the results." A probable valid biomarker is defined as "[a] biomarker that is measured in an analytical test system with well-established performance characteristics and for which there is a scientific framework or body of evidence that appears to elucidate the physiologic, toxicologic, pharmacologic, or clinical significance of the test results" (FDA 2005a, p. 17). The Guidance provides that "validation of a biomarker is context-specific and the criteria for validation will vary with the intended use of the biomarker. The clinical utility (for example, ability to predict toxicity, effectiveness or dosing) and use of epidemiology/population data (for example, strength of genotype-phenotype associations) are examples of approaches that can be used to determine the specific context and the necessary criteria for validation" (FDA 2005a, p. 17).

[2]"Pharmacogenomic" data and guidances are included in this table and discussion because the term pharmacogenomic is often used to include data about the toxicity and safety of pharmaceutical compounds (referred to in this report as toxicogenomics) because regulatory use and validation of other types of genomic data are relevant to toxicogenomics.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

TABLE 9-1 Worldwide Regulatory Policies and Guidelines Related to Toxicogenomics and Pharmacogenomics

Region/Document Type	Document	Issue Date
United States: Food and Drug Administration	http://www.fda.gov	
Guidance	Multiplex Tests for Heritable DNA Markers, Mutations and Expression Patterns; Draft Guidance for Industry and FDA Reviewers	April 21, 2003
Guidance	Guidance for Industry: Pharmacogenomic Data Submissions	March 2005
Guidance	Guidance for Industry and FDA Staff: Class II Special Controls Guidance Document: Drug Metabolizing Enzyme Genotyping System	March 10, 2005
Concept paper	Drug-Diagnostic Co-Development Concept Paper (Preliminary Draft)	April 2005
Guidance	Guidance for Industry and FDA Staff: Class II Special Controls Guidance Document: RNA Preanalytical Systems	August 25, 2005
Environmental Protection Agency	http://www.epa.gov	
Guidance	Interim Genomics Policy	June 2002
White paper	Potential Implications of Genomics for Regulatory and Risk Assessment Applications at EPA	December 2004
Europe: European Agency for the Evaluation of Medicinal Products	http://www.emea.eu.int	
Position paper	CPMP Position Paper on Terminology in Pharmacogenetics (EMEA/CPMP/3070/01)	November 21, 2002
Guideline	CHMP Guideline on Pharmacogenetics Briefing Meetings (EMEA/CHMP/20227/2004)	March 17, 2005
Supplement	Understanding the Terminology Used in Pharmacogenetics (EMEA/3842/04)	July 29, 2004

(Continued on next page)

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

TABLE 9-1 Continued

Region/Document Type	Document	Issue Date
Concept paper	CHMP Concept Paper on the Development of a Guideline on Biobanks Issues Relevant to Pharmacogenetics (Draft) (EMEA/CHMP/6806/2005)	March 17, 2005
Japan: Ministry of Health, Labour, and Welfare	http://www.mhlw.go.jp/english	
Guidance	Clinical Pharmacokinetic Studies of Pharmaceuticals (Evaluation License Division Notification No. 796)	June 1, 2001
Guidance	Methods of Drug Interaction Studies (Evaluation License Division Notification No. 813)	June 4, 2001
Notification	Submission to government agencies of data and related matters concerning the drafting of guidelines for the application of pharmacogenomics in clinical studies of drugs	March 18, 2005

FDA also lists possible reasons why a probable valid biomarker may not have reached the status of a known valid marker including the following: "(i) the data elucidating its significance may have been generated within a single company and may not be available for public scientific scrutiny; (ii) the data elucidating its significance, although highly suggestive, may not be conclusive; and (iii) independent verification of the results may not have occurred" (FDA 2005a, p. 17, 18). Although FDA outlines clear steps for sponsors to follow with regard to regulatory expectations for each type of biomarker, these classifications are not officially recognized outside FDA.

In addition to this FDA guidance, a number of other regulatory policies and guidelines have been issued worldwide that cover topics related to pharmacogenomics and toxicogenomics (see Table 9-1). Future efforts to harmonize the use and expectations for genomic data will provide value in reducing the current challenge pharmaceutical companies face in addressing guidances for different countries.

EPA issued an Interim Policy on Genomics in 2002 to allow consideration of genomic data in regulatory decision making but stated that these data alone would be "insufficient as a basis for decisions" (EPA 2002, p. 2). The Interim Policy states that EPA "will consider genomics information on a case-by-case basis" and that "[b]efore such information can be accepted and used, agency review will be needed to determine adequacy regarding the quality, representativeness, and reproducibility of the data" (EPA 2002, Pp. 2-3). The EPA is also in the process of standardizing data-reporting elements for new in vitro and in

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

silico test methods including microarrays, using the Minimum Information About a Microarray Experiment criteria as a starting point.

At the interagency level, Congress established a permanent Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) in 2000 to require that new and revised test methods be validated to meet the needs of federal agencies. The NIEHS, EPA, FDA, and OSHA are 4 of the 15 federal regulatory and research agencies participating in ICCVAM. Europe has created a similar validation organization called the European Centre for the Validation of Alternative Methods (ECVAM). At the international level, the Organization for Economic Co-operation and Development has also adopted formal guidelines to validate test methods for use in regulatory decision making (OECD 2001).

The ICCVAM criteria are useful guides for regulatory validation of toxicogenomic data and methods, but toxicogenomic technologies will require unique and more flexible approaches to validation given their rapid pace of change and other unique characteristics (Corvi et al. 2006). To that end, ICCVAM and ECVAM are developing a unique and more flexible approach to validating toxicogenomic test methods for regulatory use, and they have convened a series of workshops on the topic (Corvi et al. 2006). One approach being investigated is a "modular" model, in which different steps in the validation process are independently undertaken, in contrast to the traditional stepwise "linear" model that may unduly delay the validation of rapidly evolving toxicogenomic technologies (Corvi et al. 2006). Agencies such as the EPA are carefully tracking and participating in this initiative and anticipate applying the output of the ICCVAM process in their own regulatory programs (EPA 2004).

CONCLUSIONS

Toxicogenomics has reached the stage where many of the initial technical questions have been resolved, at least for the more mature approaches such as gene expression analysis with microarrays. The community has learned that careful experiments using genomic approaches can provide results that are both comparable among laboratories and reveal insight into the biology of the system under study (Bammler et al. 2005; Irizarry et al. 2005; Larkin et al. 2005). However, the need for standards for assessing the quality of particular experiments remains, and it will affect the utility of datasets that are and will be generated. The work of the ERCC (and other groups) to develop RNA standards (Cronin et al. 2004) is a potentially important component of this effort and should be encouraged and continued, but additional work and development are necessary if truly useful quality assessment standards are to be created. Standard development efforts should not be limited to gene expression microarray analysis, as similar standards will be necessary if other toxicogenomic technologies are to be widely used and trusted to give reliable results.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Beyond quality control of individual experiments, more extensive validation is needed to use toxicogenomic data for the applications discussed in this report. Most toxicogenomic projects have focused on limited numbers of samples with technologies, such as DNA microarrays, that may not be practical for large-scale applications that go beyond the laboratory. Consequently, validation of toxicogenomic signatures should focus not only on the primary toxicogenomic technology (such as DNA microarrays) but also on assays that can be widely deployed (such as qRT-PCR) at relatively low cost. Many issues associated with validation of toxicogenomic signatures will rely on the availability of large, accessible, high-quality datasets to evaluate the specificity and sensitivity of the assays. Those datasets must include not only the primary data from the assays but also the ancillary data about treatments and other factors necessary for analysis. This argues for the creation and population of a public data repository for toxicogenomic data.

A means of regulatory validation of toxicogenomic applications is needed—for example, for toxicogenomic data accompanying new submissions of drug candidates for approval. Specifically, the development of new standards and guidelines that will provide clear, dynamic, and flexible criteria for the approval and use of toxicogenomic technologies is needed at this point. Development of these standards and guidelines requires suitable datasets. For example, the use of toxicogenomics to classify new compounds for their potential to produce a specific deleterious phenotype requires a useful body of high-quality, well annotated data. The existing ICCVAM approaches do not provide guidance for the large-scale toxicogenomic approaches being developed, as evidenced by the different guidelines the FDA and the EPA are developing, and toxicogenomic tools need not be subject to ICCVAM protocols before they are considered replacement technologies. Although multiagency initiatives, such as the one ICCVAM is spearheading, may serve as a basis for establishing the needed standards and criteria in the longer term, the overall ICCVAM approach does not seem well suited for validating new technologies and as such will need to be significantly revised to accommodate the tools of toxicogenomics. Consequently, regulatory agencies such as the EPA and the FDA should move forward expeditiously in continuing to develop and expand their validation criteria to encourage submission and use of toxicogenomic data in regulatory contexts.

In summary, the following are needed to move forward in validation:

- Objective standards for assessing quality and implementing quality control measures for the various toxicogenomic technologies;
- Guidelines for extending technologies from the laboratory to broader applications, including guidance for implementing related but more easily deployable technologies such as qRT-PCR and ELISAs;
- A clear and unified approach to regulatory validation of "-omic" technologies that aligns the potentially diverse standards being developed by various federal agencies, including the EPA, FDA, NIEHS, and OSHA, as well as at-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

tempting to coordinate standards with the relevant European and Asian regulatory agencies.

- A well-annotated, freely accessible, database providing access to high-quality "-omic" data.

RECOMMENDATIONS

The following specific actions are recommended to facilitate technical validation of toxicogenomic technologies:

1. Develop objective standards for assessing sample and data quality from different technology platforms, including standardized materials such as those developed by the ERCC.
2. Develop appropriate criteria for using toxicogenomic technologies for different applications, such as hazard screening and exposure assessment.
3. Regulatory agencies should establish clear, transparent, and flexible criteria for the regulatory validation of toxicogenomic technologies. Whereas the use of toxicogenomic data will be facilitated by harmonization of data and method validation criteria among U.S. regulatory agencies and at the international level, harmonization should be a long-term goal and should not prevent individual agencies from developing their own validation procedures and criteria in the shorter term.

Copyright © National Academy of Sciences. All rights reserved.

10

Sample and Data Collection and Analysis

As discussed in previous chapters, toxicogenomic studies face significant challenges in the areas of validation, data management, and data analysis. Foremost among these challenges is the ability to compile high-quality data in a format that can be freely accessed and reanalyzed. These challenges can best be addressed within the context of consortia that generate high-quality, standardized, and appropriately annotated compendia of data. This chapter discusses issues related to the role of private and public consortia in (1) sample acquisition, annotation, and storage; (2) data generation, annotation, and storage; (3) repositories for data standardization and curation; (4) integration of data from different toxicogenomic technologies with clinical, genetic, and exposure data; and (5) issues associated with data transparency and sharing. Although this chapter touches on how these issues relate to risk assessment and the associated ethical, legal, and social implications, other chapters of this report discuss these aspects of toxicogenomics in greater detail.

LEVERAGING SAMPLES AND INFORMATION IN POPULATION-BASED STUDIES AND CLINICAL TRIALS FOR TOXICOGENOMIC STUDIES

Because the potential of toxicogenomic studies is to improve human risk assessment, it is imperative that susceptible human populations be studied with toxicogenomics. Given the hundreds of millions of dollars already spent on clinical trials, environmental cohort studies, and measurements of human exposures to environmental chemicals, it is imperative that toxicogenomic studies make the best possible use of the human population and clinical resources already in hand.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

This prospect raises many questions. How does the research community ensure that appropriate cohorts and sample repositories are available once the technologies have matured to the point of routine application? What are the limitations of using existing cohorts for toxicogenomic studies? Given the limitations of existing cohorts with respect to the informed consent, sample collection, and data formats, how should future population-based studies be designed? What is the ideal structure of a consortium that would assemble such a cohort, collect the necessary samples and data, and maintain repositories? Consideration of these questions begins with an examination of the current state of affairs.

Current Practices and Studies

Central to all population-based studies are identifying suitable cohorts, recruiting participants, obtaining informed consent, and collecting the appropriate biologic samples and associated clinical and epidemiologic data. Collecting, curating, annotating, processing, and storing biologic samples (for example, blood, DNA, urine, buccal swabs, histologic sections) and epidemiologic data for large population-based studies is a labor-intensive, expensive undertaking that requires a significant investment in future research by the initial study investigators. As a result, study investigators have reasonable concerns about making biologic samples available for general use by others in the scientific community. With the notable exception of immortalized lymphocyte cell lines that provide an inexhaustible resource from the blood of study participants, biologic specimens represent a limited resource and researchers often jealously guard unused samples for future use once the "ideal" assays become available. It is not unusual for institutions with sizeable epidemiology programs to have hundreds of thousands of blood, serum, lymphocyte, and tumor samples in storage that are connected to clinical data as well as information on demographics, lifestyle, diet, behavior, and occupational and environmental exposures.

In addition to cohorts and samples collected as part of investigator-initiated studies and programs, sample repositories have also been accrued by large consortia and cooperative groups sponsored by the public and private sectors. An example of publicly sponsored initiatives in the area of cancer research and national and international sample and data repositories is the National Cancer Institute (NCI)-sponsored Southwest Oncology Group (SWOG), one of the largest adult cancer clinical trial organizations in the world. The SWOG membership and network consists of almost 4,000 physicians at 283 institutions throughout the United States and Canada. Since its inception in 1956, SWOG has enrolled and collected samples from more than 150,000 patients in clinical trials.

Examples in the area of cancer prevention are the NCI-sponsored intervention trials such as the CARET (beta-Carotene and Retinoic Acid) study for prevention of lung cancer, and the Shanghai Breast Self Exam study conducted in factory workers in China. CARET was initiated in 1983 to test the hypothesis

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

that antioxidants, by preventing DNA damage from free radicals present in tobacco smoke, might reduce the risk of lung cancer. CARET involved more than 18,000 current and former male and female smokers, as well as males exposed to asbestos, assigned to placebo and treatment groups with a combination of beta-carotene and vitamin A. Serum samples were collected before and during the intervention, and tumor samples continue to be accrued. The randomized trial to assess the efficacy of breast self-exam in Shanghai accrued 267,040 current and retired female employees associated with 520 factories in the Shanghai Textile Industry Bureau. Women were randomly assigned on the basis of factory to a self-examination instruction group (133,375 women) or a control group that received instruction in preventing back injury (133,665 women) (Thomas et al. 1997). A large number of familial cancer registries have also been established for breast, ovarian, and colon cancer. There are comparable cooperative groups to study noncancer disease end points, such as heart disease, stroke, and depression.

Other examples include several large prospective, population-based studies designed to assess multiple outcomes. Participants provide baseline samples and epidemiologic data and are then followed with periodic resampling and updating of information. The prototype for this type of study is the Nurses' Health Study (NHS), established in 1976 with funding from the National Institutes of Health (NIH), with a second phase in 1989. The primary motivation for the NHS was to investigate the long-term consequences of the use of oral contraceptives. Nurses were selected because their education increased the accuracy of responses to technically worded questionnaires and because they were likely to be motivated to participate in a long-term study. More than 120,000 registered nurses were followed prospectively. Every 2 to 4 years, cohort members receive a follow-up questionnaire requesting information about diseases and health-related topics including smoking, hormone use and menopausal status, diet and nutrition, and quality of life. Biologic samples collected included toenails, blood samples, and urine.

Another example of an ongoing trial to assess multiple disease end points is the NIH-funded Women's Health Initiative. Established in 1991, this study was designed to address the most common causes of death, disability, and poor quality of life in postmenopausal women, including cardiovascular disease, cancer, and osteoporosis. The study included a set of clinical trials and an observational study, which together involved 161,808 generally healthy postmenopausal women. The clinical trials were designed to test the effects of postmenopausal hormone therapy, diet modification, and calcium and vitamin D supplements as well as ovarian hormone therapies. The observational study had several goals, including estimating the extent to which known risk factors predict heart disease, cancers, and fractures; identifying risk factors for these and other diseases; and creating a future resource to identify biologic indicators of disease, especially substances found in blood and urine. The observational study enlisted 93,676 postmenopausal women whose health was tracked over an aver-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

age of 8 years. Women who joined this study filled out periodic health forms and visited the research clinic 3 years after enrollment.

Sample/data repositories are also being assembled in the private sector, where pharmaceutical companies are increasingly eager to form cooperative groups or collaborations with academia and health care providers that provide access to patient cohorts for clinical trials and other population-based studies. In some cases, companies have assembled their own cohorts with samples purchased directly from medical providers. Large prospective cohorts have also been assembled, sampled, and observed by departments of health in countries with socialized medicine (for example, Sweden, Finland). There are multiple examples of established cohort studies with samples that could be used in studies that evaluate the impact of environmental exposures. For example, the Occupational and Environmental Epidemiology Branch of the NCI (NCI 2006a) conducts studies in the United States and abroad to identify and evaluate environmental and workplace exposures that may be associated with cancer risk. Another example is the Agricultural Health Study (AHS 2006), a prospective study of nearly 90,000 farmers and their spouses in Iowa and North Carolina, carried out in collaboration with the National Institute of Environmental Health Sciences and the Environmental Protection Agency (EPA). These cohort studies involve sophisticated exposure assessments and mechanistic evaluations and include intensive collaborations among epidemiologists, industrial hygienists, and molecular biologists. These studies often involve collecting biologic samples to assess biologic effects from exposure to agricultural (for example, pesticides and herbicides), industrial, and occupational chemicals.

Another valuable resource is the National Health and Nutrition Examination Survey (NHANES) conducted by the Environmental Health Laboratory of the Centers for Disease Control and Prevention (CDC) at the National Center for Environmental Health. The goal of NHANES is to identify environmental chemicals in participants, quantify their levels, and determine how these amounts relate to health outcomes. The survey design calls for collecting blood, urine, and extensive epidemiologic data (demographic, occupational, lifestyle, dietary, and medical information) from people of all ages and from all areas of the country. As such, NHANES provides representative exposure data for the entire U.S. population. Rather than estimating exposures, NHANES measures the amounts of hundreds of chemicals or their metabolites in participants' blood and urine, using the most sensitive, state-of-the-art analytical techniques. The number of people sampled varies among compounds but is typically several hundred to thousands, which is sufficient for determining the range of exposures in the population; determining racial, age, and sex differences in exposure; detecting trends in exposures over time; and analyzing the efficacy of interventions designed to reduce exposures. In addition to the measurements of exposures, blood and urine are also used to measure biomarkers of nutrition and indicators of general health status. A report of the findings is published every 2 years.

In summary, academia, numerous government agencies, health care providers, and companies in the private sector have already invested tremendous

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

effort and resources into accruing cohorts for population-based studies to design better drugs to predict responses to drug therapy, to assess the efficacy of drugs, to improve disease classification, to find genetic markers of disease susceptibility, to understand gene-environment interactions, and to assess the health effects of human exposures to environmental chemicals. These resources present an opportunity for developing partnerships to apply toxicogenomic technologies to ongoing and future collaborative projects.

Limitations and Barriers

The most expensive and arduous components of any population-based study are the collection of epidemiologic data and samples in a form that can be used to answer the largest number of questions in a meaningful way. Ideally, the same cohorts and samples should be available for multiple studies and reanalyses. However, for a wide variety of reasons, ranging from the types of samples collected to issues of informed consent, this is rarely possible given the structure of existing cohorts.

Structure of Cohorts

Given the cost and logistics of assembling, obtaining consent, and following large cohorts, many studies are designed to address only a few specific questions. Typically, case-control or association studies are used to test a new hypothesis before testing is expanded and validated with robust and expensive population-based cross-sectional designs or prospective studies. As a result, many studies lack the statistical power to address additional hypotheses. For example, a case control study designed to investigate the contribution of single nucleotide polymorphisms (SNPs) in a gene encoding a specific xenobiotic metabolizing enzyme in cancer will typically be underpowered to investigate the interactions of these SNPs with SNPs in other genes. Using larger cohorts that allow for a broader application would clearly be advantageous to future toxicogenomic studies.

Heterogeneity of Disease or Response Classification

Another important aspect of any population-based study is accurate genotyping and precise phenotyping of diseases and responses. Whether a study is looking at genetic linkage, disease association, susceptibility, or responsiveness to environmental agents, any phenotypic or genotypic misclassification will reduce or even obscure the truth by weakening the association between the correct genotype and the phenotype. Some diseases, such as clinical depression, migraine headaches, and schizophrenia are inherently difficult to diagnose and accurately classify at the clinical level. Likewise, an increasing number of molecular analyses of tumors indicate that significant heterogeneity exists even

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

among tumors with a similar histopathology and that these molecular differences can influence the clinical course of the disease and the response to therapy or chemoprevention (Potter 2004). The resulting inaccuracies in genotypic and phenotypic stratification of disease in cases can limit the utility of cohorts and their associated samples. However, increased stratification of disease based on genotype and molecular phenotype can also have an adverse effect by reducing the statistical power of a cohort to detect disease association or linkage. As the capacity to define homogeneous genotypes and phenotypes increases, the size of cohorts will need to be increased well above those used in present studies.

Sample Collection

A major impediment to the use of existing cohorts is that many studies have not collected appropriate types of specimens, or the available specimens are in a format that is not amenable to toxicogenomic studies. For example, the NCI funded the Cancer Genetic Network to identify and collect data on families at elevated risk for different forms of cancer. No provisions were made for collecting biologic specimens within the funded studies, and the samples that are collected are often inappropriate for genomic analysis. Traditionally, cancer cohort studies have collected formalin-fixed, paraffin-embedded samples of tumors and adjacent normal tissue. Although DNA for SNP-based genotyping can be extracted from such fixed specimens, albeit with some difficulty (e.g., Schubert et al. 2002), these samples usually do not yield representative mRNA appropriate for gene expression profiling.

Clinical trials and epidemiologic studies have not usually collected or analyzed DNA samples (Forand et al. 2001). On the other hand, most molecular epidemiology studies diligently collected blood or serum samples for biomarker analyses, but multiple challenges remain. The way samples are collected, handled, shipped, and preserved varies greatly among studies. Many samples are flash frozen and hence are adequate for DNA extraction and genotyping, but such samples are usually limited in size and this may not allow for comprehensive genotyping. To deal with this limitation, some studies have developed transformed cell lines from human lymphocytes as a way to create inexhaustible supplies of DNA for genomic studies (e.g., Shukla and Dolan 2005). These cell lines could provide valuable experimental material to determine the impact of interindividual genetic variation on the response to compounds in in vitro studies. However, whether the results from studies with cell lines can be translated to human health effects remains to be determined.

In the case of the primary blood lymphocytes from study participants, unless steps were taken to preserve mRNA at the time of collection, the samples may have limited utility in gene expression profiling, although serum samples obviously could be used for proteomic and metabonomic studies. The NCI is currently funding a large initiative in the application of serum proteomics for early detection of disease (NCI 2006b). However, whether the methods of pres-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

ervation currently used will allow for accurate analyses after many years of storage and to what extent different methods of sample preparation and storage affect the applicability of samples to proteomic and metabonomic analyses remain to be determined. In summary, there are numerous impediments to using existing samples available through cohort studies. There are both a need and an opportunity to standardize methodologies in ongoing studies and to design future studies to accommodate the requirements of new toxicogenomic platforms.

Data Uniformity

Another impediment to using existing cohorts for toxicogenomic applications is the lack of uniformity in data collection standards among population-based studies. Although there will always be a need to collect data specific to a particular study or application, much can be done to standardize questionnaires that collect demographic, dietary, occupational, and lifestyle data common to all studies. Moreover, efforts should be launched to develop a standardized vocabulary for all data types, including clinical data, which can be recorded in a digitized electronic format that can be searched with text-mining algorithms. An example of such a standardized approach is the attempt to standardize the vocabulary for histopathology within the Mouse Models of Cancer Consortium (NCI 2006c). The NCI launched a related effort, the Cancer Bioinformatics Grid, to develop standardized tools, ontologies, and representations of data (NCI 2006d).

Sharing and Distributing Data

The biologic samples and data collected by clinical trials, epidemiologic studies, and human exposure studies funded by agencies such as NIH, the CDC, the EPA, and other agencies represent a significant public scientific resource. The full value of these studies for basic research and validation studies cannot be realized until appropriate data-sharing and data-distribution policies are formulated and supported by government, academia, and the private sector. Several NIH institutes such as the NCI and the National Heart, Lung, and Blood Institute (NHLBI) have drafted initial policies to address this need.

These policies have several key features that promote optimal use of the data resources, while emphasizing the protection of data on human subjects. For example, under the NHLBI policy, study coordinators retain information such as identifiers linked to personal information (date of birth, study sites, physical exam dates). However "limited access data," in which personal data have been removed or transformed into more generic forms (for example, age, instead of date of birth), can be distributed to the broader scientific community. Importantly, study participants who did not consent to their data being shared beyond the initial study investigators are not included in the dataset. In some cases, consent forms provide the option for participants to specify whether their data can

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

be used for commercial purposes. In these cases, it is important to discriminate between commercial purpose datasets and non-commercial purpose datasets to protect the rights of the human subjects.

Because it may be possible to combine limited access data with other publicly available data to identify particular study participants, those who want to obtain these datasets must agree in advance to adhere to established data-distribution agreements through the Institute, with the understanding that violation of the agreement may lead to legal action by study participants, their families, or the U.S. government. These investigators must also have approval of the institutional review board before they distribute the data.

Under these policies, it is the responsibility of the initial study investigators to prepare the datasets in such a way as to enable new investigators who are not familiar with the data to fully use them. In addition, documentation of the study must include data collection forms, study procedures and protocols, data dictionaries (codebooks), descriptions of all variable recoding, and a list of major study publications. Currently, the format of the data is requested to be in Statistical Analysis Software (SAS) and a generation program code for installing a SAS file from the SAS export data file is requested for NHLBI datasets. The NCI cancer bioinformatics information grid (CaBIG) initiative is exploring similar requirements.

Timing the release of limited access data is another major issue. Currently, policies differ across institutes and across study types. For large epidemiologic studies that require years of data to accumulate before analysis can begin, a period of 2 to 3 years after the completion of the final patient visit is typically allowed before release of the data. The interval is intended to strike a balance between the rights of the original study investigators and the wider scientific community.

Commercialization of repository data is another issue that could impede or enhance future studies. In Iceland, DNA samples from the cohort composed of all residents who agree to participate are offered as a commercial product for use in genetic studies. Based in Reykjavik, deCODE's product is genetic information linked, anonymously, to medical records for the country's 275,000 inhabitants. Iceland's populations are geographically isolated and many share the same ancestors. In addition, family and medical records have been thoroughly recorded since the inception of the National Health Service in 1915. Icelanders also provide a relatively simple genetic background to investigate the genetics of disease. The population resource has proved its worth in studies of conditions caused by single defective genes, such as rare hereditary conditions including forms of dwarfism, epilepsy, and eye disorders. deCODE has also initiated projects in 25 common diseases including multiple sclerosis, psoriasis, preeclampsia, inflammatory bowel disease, aortic aneurism, alcoholism, and obesity. Clearly, future toxicogenomic studies will have to take into account possible use of such valuable, albeit commercial, resources.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Informed Consent

Perhaps the greatest barrier to using existing sample repositories and data collections for toxicogenomic studies is the issue of informed consent and its ethical, legal, and social implications (see Chapter 11). Government regulations increasingly preclude the use of patient data and samples in studies for which the patient did not specifically provide consent. Consequently, many potentially valuable resources cannot be used for applications in emerging fields such as toxicogenomics. Legislation to protect the public, such as the Health Insurance Portability and Accountability Act (HIPAA), which was enacted to protect patients against possible denial of health insurance as a result of intentional or unintentional distribution of patient health or health risk data, has unwittingly impaired population-based studies (DHHS 2006). Researchers now face tightening restrictions on the use of any data that can be connected with patient identity. Despite the significant barriers imposed by patient confidentiality and issues of informed consent, mechanisms exist for assembling large cohorts and collecting samples that can be used for multiple studies and applications to benefit public health. For example, the NHANES program under the aegis of the CDC collects blood samples for measurement of chemicals, nutritional biomarkers, and genotypes from large numbers of individuals using a consent mechanism that expressly authorizes multiple uses of biologic samples. However, because NHANES functions under the aegis of the CDC, it is exempt from many of the requirements of HIPAA. Nonetheless, solutions can be found that meet research needs while protecting the rights of individuals. An example of a successful approach is the BioBank in the United Kingdom (UK Biobank Limited 2007), a long-term project aimed at building a comprehensive resource for medical researchers by gathering information on the health and lifestyle of 500,000 volunteers between 40 and 69 years old. After providing consent, each participant donates a blood and urine sample, has standard clinical diagnostic measurements (such as blood pressure), and completes a confidential lifestyle questionnaire. Fully approved researchers can then use these resources to study the progression of illnesses such as cancer, heart disease, diabetes, and Alzheimer's disease in these patients over the next 20 to 30 years to develop new and better ways to prevent, diagnose, and treat such problems. The BioBank ensures that data and samples are used for ethically and scientifically approved research. Issues such as informed consent, confidentiality, and security of the data are guided by an Ethics and Governance Framework overseen by an independent council.

In summary, the current state of sample repositories and their associated data are less than ideal and there are numerous limitations and barriers to their immediate use in the emerging field of toxicogenomics. Future studies require that issues in the design of cohort studies, phenotype and genotype classification, standardization of sample and data collection, database structure and sharing, and informed patient consent be addressed and mitigated. However, even when all these issues have been addressed, toxicogenomics faces additional and

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

significant challenges related to the complexity of data collected by the various toxicogenomic technologies.

EXISTING TOXICOGENOMIC DATA REPOSITORIES: STANDARDIZATION, AVAILABILITY, AND TRANSPARENCY

The following section provides an overview of the current standards for toxicogenomic data and a brief overview of existing databases and repositories. The remaining needs that must be met to move forward are described in the Conclusions section of this chapter.

Standards for Toxicogenomic Data

Although each individual study conducted using toxicogenomic approaches may provide insight into particular responses produced by compounds, the real value of the data goes beyond the individual assays. The real value of these data will be realized only when patterns of gene, protein, or metabolite expression are assembled and linked to a variety of other data resources. Consequently, there is a need for well-designed and well-maintained repositories for toxicogenomic data that facilitate toxicogenomic data use and allow further independent analysis. This is a daunting task. Genome sequencing projects have generated large quantities of data, but the ancillary information associated with those data are much less complex than the information necessary to interpret toxicogenomic data. Genome sequencing does not vary significantly with the tissue analyzed, the age of the organism, its diet, the time of day, hormonal status, or exposure to a particular agent. Analyses of expression patterns, however, are dramatically affected by these and a host of other variables that are essential for proper interpretation of the data.

In 1999, a group representing the major DNA sequence databases, large-scale practitioners of microarray analysis, and a number of companies developing microarray hardware, reagents, and software tools began discussing these issues. These discussions resulted in creation of the Microarray Gene Expression Data Society (MGED). MGED took on the task of developing standards for describing microarray experiments with one simple question in mind: What is the minimum information necessary for an independent scientist to perform an independent analysis of the data? Based on feedback received through numerous discussions and a series of public meetings and workshops, a new set of guidelines called the MIAME standard (Minimal Information About a Microarray Experiment) was developed (MIAME 2002). The publication of this new standard was met with general enthusiasm from the scientific community and the standards have evolved with continued input from scientists actively conducting microarray studies. To facilitate usage of this new standard, brief guidelines and a “MIAME checklist” were developed and provided to scientific journals for

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

their use when considering publication of manuscripts presenting microarray data (Brazma et al. 2001, MGED 2005).

Whereas MIAME can be thought of as a set of guidelines for describing an experiment, it is clear that these guidelines must be translated into protocols enabling the electronic exchange of data in a standard format. To meet that challenge, a collaborative effort by members of MGED and a group at Rosetta Inpharmatics led to the development of the microarray gene expression (MAGE) object model as well as its implementation as an XML-based extensible markup language (Spellman et al. 2002). The adoption of MAGE by the Object Management Group has promoted MAGE to the status of an "official" industry standard. MAGE is now being built into a wide range of freely available microarray software, including BASE (Saal et al. 2002), BioConductor (Dudoit et al. 2003), and TM4 (Saeed et al. 2003). An increasing number of companies are also adopting MIAME and MAGE as essential components of their products. The guidelines in MIAME have continued to evolve, and both MIAME and MAGE are being prepared for a second release.

Efforts are also under way to develop an extended version of the MIAME data standard that allows for integration of other data specific to toxicologic experiments (for example, dose, exposure time, species, and toxicity end points). Deemed MIAME/Tox, this standard would form the basis for entry of standardized data into toxicogenomic databases (MIAME 2003). Similar efforts have produced microarray standards for a variety of related applications, including a MIAME/Env for environmental exposures (Sansone et al. 2004) and MIAME/Nut for nutrigenomics (Garosi et al. 2005). Work is also ongoing to develop standards for proteomics (the Minimal Information About a Proteomics Experiment) (Orchard et al. 2004) and for metabolomic profiling (the Standard Metabolic Reporting Structure) (Lindon et al. 2005a).

Public Data Repositories

The ArrayExpress database of the European Bioinformatic Institute (EBI) (Brazma et al., 2003), as well as the National Center for Biotechnology Information's Gene Expression Omnibus (Edgar et al. 2002) and the Center for Information Biology gene Expression database at the DNA Data Bank of Japan have adopted and supported the MIAME standard. Following the same model used for sequence data, data exchange protocols are being developed to link expression data found in each of these three major repositories. Other large public databases such as the Stanford Microarray Database (SMD 2006) and CaBIG (NCI 2006d) have been developed in accordance with the MIAME guidelines. However, these are essentially passive databases that apply standards to data structure but provide little or no curation for data quality, quality assurance, or annotation. Other small-scale, publicly available tools and resources that have been developed for sharing, storing, and mining toxicogenomic data include db Zach

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

(MSU 2007) and EDGE[2], the Environment, Drugs and Gene Expression database (UW-Madison 2006).

Private and Public Toxicogenomic Consortia

Databases developed within private or public consortia are often much more proactive in curating data on the basis of quality and annotation. The pharmaceutical industry, for example, has generated very large compendia of toxicogenomic data on both proprietary and public compounds. These data are high quality but in general are not accessible to the public. Datasets and data repositories developed for toxicogenomics within the public sector, or in partnership with the private sector, are also actively curated and annotated but are made available for public access after publication of findings. Examples of such cooperative ventures are described below.

The Environmental Genome Project

The Environmental Genome Project (EGP) is an initiative sponsored by the National Institute of Environmental Health Sciences (NIEHS) that was launched in 1998. Inextricably linked to the Human Genome Project, the underlying premise for the EGP was that a subset of genes exists that have a greater than average influence on human susceptibility to environmental agents. Therefore, it was reasoned that the identification of these environmentally responsive genes, and characterization of their SNPs, would lead to enhanced understanding of human susceptibility to diseases with an environmental etiology. The EGP provided funding for extramural research projects in multiple areas, including bioinformatics/statistics; DNA sequencing; functional analysis; population-based epidemiology; ethical, legal, and social issues; technology development; and mouse models of disease. A variety of mechanisms including centers such as the Comparative Mouse Genomics Centers Consortium, the Toxicogenomics Research Consortium, the SNPs Program, and the Chemical Effects in Biological Systems database, which are described below, provided research support.

The NIEHS scientific community selected the list of environmentally responsive genes to be analyzed, which currently numbers 554 genes, although this list is not inclusive and is subject to ongoing modification. Among the environmentally responsive genes are those involved in eight categories or ontogenies including cell cycle, DNA repair, cell division, cell signaling, cell structure, gene expression, apoptosis, and metabolism. The goal of the EGP included resequencing these genes in 100 to 200 individuals representing the range of human diversity. Polymorphisms are then assessed for their impact on gene function. However, the small size of the program precluded disease-association studies. The goal of the program was to provide a database of SNPs that could then be used in larger epidemiologic studies.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

The Pharmacogenetics Research Network

The NIH-sponsored Pharmacogenetics Research Network is a nationwide collaborative research network that seeks to enhance understanding of how genetic variation among individuals contributes to differences in responses to drugs. Part of this network is the Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB), a publicly available Internet research tool that curates genetic, genomic, and cellular phenotypes as well as clinical information obtained from participants in pharmacogenomic studies. The database includes, but is not limited to, information obtained from clinical, pharmacokinetic, and pharmacogenomic research on drugs targeting the cardiovascular system, the pulmonary systems, and cancer as well as studies on biochemical pathways, metabolism, and transporter domains. PharmGKB encourages submission of primary data from any study on how genes and genetic variation affect drug responses and disease phenotypes.

The NIEHS Toxicogenomics Research Consortium

As a complement to its ongoing participation with the International Life Sciences Institute's Health and Environmental Sciences Institute (ILSI-HESI) Consortium, the NIEHS, in conjunction with the National Toxicogenomics Program (NTP) established the Toxicogenomics Research Consortium (TRC). The goals were to enhance research in the area of environmental stress responses using transcriptome profiling by (1) developing standards and "best practices" by evaluating and defining sources of variation across labs and platforms, and (2) contributing to the development of a robust relational database that combines toxicologic end points with changes in the expression patterns of genes, proteins, and metabolites.

The TRC included both consortium-wide collaborative studies and independent research projects at each consortium member site. On the whole, the TRC was successful and achieved many of its goals. The TRC published a landmark paper that was the first to not only define the various sources of error and variability in microarray experiments but also to quantify the relative contributions of each source (Bammler et al. 2005). The study confirmed the finding of the ILSI-HESI Consortium (see page 165), indicating that, despite the lack of concordance across platforms in the expression of individual genes, concordance was high when considering affected biochemical pathways. The TRC results were also concordant with work by other groups who arrived at similar conclusions (Bammler et al. 2005, Dobbin et al. 2005, Irizarry et al. 2005, Larkin et al. 2005). Although the amount of funding limited the scope of independent research projects funded within each center, this component was highly successful and has led to numerous outstanding publications in the field (e.g., TRC 2003).

Copyright © National Academy of Sciences. All rights reserved.

The Chemical Effects in Biologic Systems Knowledge Base

The Chemical Effects in Biological Systems (CEBS) (Waters et al. 2003b) knowledge base developed at the National Center for Toxicogenomics at the NIEHS represents a first attempt to create a database integrating various aspects of toxicogenomics with classic information on toxicology. The goal was to create a database containing fully documented experimental protocols searchable by compound, structure, toxicity end point, pathology end point, gene, gene group, SNP, pathway, and network as a function of dose, time, and the phenotype of the target tissue. These goals have not been realized.

ILSI-HESI Consortium

The ILSI-HESI Consortium on the Application of Genomics to Mechanism Based Risk Assessment is an example of a highly successful, international collaboration between the private and public sectors, involving 30 laboratories from industry, academia, and national laboratories. The ILSI-HESI Toxicogenomics Committee developed the MIAME implementation for the toxicogenomic community in collaboration with the European Bioinformatics Institute (a draft MIAME/Tox standard is included in Appendix B). In addition to helping develop data standards, the consortium was also the first large-scale experimental program to offer practical insights into genomic data exchange issues.

Using well-studied compounds with well-characterized organ-specific toxicities (for example, acetaminophen), the consortium conducted more than 1,000 microarray hybridizations on Affymetrix and cDNA platforms to determine whether known mechanisms of toxicity can be associated with characteristic gene expression profiles. These studies validated the utility of gene expression profiling in toxicology, led to the adoption of data standards, evaluated methodologies and standard operating procedures, performed cross-platform comparisons, and investigated dose and temporal responses on a large scale. The studies also demonstrated that, despite differences in gene-specific data among platforms, expression profiling on any platform revealed common pathways affected by a given exposure. Multiple landmark papers published the research findings of the consortium (Amin et al. 2004; Baker et al. 2004; Chu et al. 2004; Kramer et al. 2004; Mattes 2004; Mattes et al. 2004; Pennie et al. 2004; Thompson et al. 2004; Ulrich et al. 2004; Waring et al. 2004).

The Consortium for Metabonomic Toxicology

Another example of a highly successful consortium, the Consortium for Metabonomic Toxicology (COMET) was a collaborative venture between University College London, six major pharmaceutical companies, and a nuclear magnetic resonance (NMR) instrumentation company. The main objectives of the consortium were similar to those of the ILSI-HESI Consortium but with pro-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

ton NMR-based metabonomic profiling (Lindon et al. 2003). The group developed a database of ^{1}H NMR spectra from rats and mice dosed with model toxins and an associated database of meta-data on the studies and samples. The group also developed powerful informatic tools to detect toxic effects based on NMR spectral profiles.

Since its inception, COMET defined sources of variation in data, established baseline levels and ranges of normal variation, and completed studies on approximately 150 model toxins (Lindon et al. 2005b). In addition, the analytical tools developed have been validated by demonstrating a high degree of accuracy in predicting the toxicity of toxins that were not included in the training dataset.

Commercial Databases: Iconix Pharmaceuticals and Gene Logic

Databases designed for use in predictive toxicology have also been developed in the private sector for commercial applications. An example is the DrugMatrix database developed at Iconix in 2001. Working with Incyte, MDS Pharma Services, and Amersham Biosciences, Iconix selected more than 1,500 compounds, all of which were used to generate in vitro expression profiles of primary rat hepatocytes and in vivo gene expression profiles of heart, kidney, and liver tissue from exposed rats. They then used these data to populate a database that included publicly available data on biologic pathways, toxicity associated with each compound, and pathology and in vitro pharmacology data. The database was then overlaid with software tools for rapid classification of new compounds by comparisons with the expression patterns of compounds in the database.

A similar database for predictive, mechanistic, and investigational toxicology, deemed Tox-Express, was developed at Gene Logic. These databases are large by current standards and hence these companies occupy a market niche in predictive toxicology. The information in these databases far exceeds that available through publicly available sources.

CONCLUSIONS

Leveraging Existing Studies

Given the potential of toxicogenomic studies to improve human risk assessment, it is imperative to conduct toxicogenomic studies of human populations. Academic institutions, government agencies, health care providers, and the private sector have already invested tremendous effort and resources into accruing cohorts for population-based studies to predict responses to drug therapy, to improve disease classification, to find genetic markers of disease susceptibility, to understand gene-environment interactions, and to assess effects of human exposures to environmental chemicals. Given the hundreds of millions of

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

dollars already spent on clinical trials, environmental cohort studies, and measurements of human exposures to environmental chemicals, it is imperative that toxicogenomic studies make the best possible use of the resources at hand. Ensuring access to the samples and data produced by existing studies will require consortia and other cooperative interactions involving health care providers, national laboratories, academia, and the private sector. Sharing of data and samples also raises important ethical, legal, and social issues, such as informed consent and confidentiality, that must be addressed.

Even though existing cohorts and ongoing studies should be leveraged when possible, the current state of sample repositories and their associated data is less than ideal, and there are numerous limitations and barriers to their immediate use in toxicogenomics. For example, there is little uniformity in data collection standards among population-based studies, and these studies for the most part were not designed with genomic research in mind. This lack of uniformity and the fact that samples and data are seldom collected in a format that can be assessed by genomic technologies present a major impediment to using existing cohorts for toxicogenomic applications. Although there will always be a need to collect data specific to a particular study or application, much can be done to standardize collection of samples and of demographic, dietary, occupational, and lifestyle data common to all studies.

Databases and Building a Toxicogenomic Database

Toxicogenomic technologies generate enormous amounts of data—on a scale even larger than sequencing efforts like the Human Genome Project. Current public databases are inadequate to manage the types or volumes of data expected to be generated by large-scale applications of toxicogenomic technologies or to facilitate the mining and interpretation of the data that are just as important as its storage. In addition, because the predictive power of databases increases iteratively with the addition of new compounds, significant benefit could be realized from incorporating greater volumes of data. A large, publicly accessible database of quality data would strengthen the utility of toxicogenomic technologies, enabling more accurate prediction of health risks associated with existing and newly developed compounds, providing context to toxicogenomic data generated by drug and chemical manufacturers, informing risk assessments, and generally improving the understanding of toxicologic effects.

The type of database needed is similar in concept to the vision for the CEBS database (Waters et al. 2003b), which was envisioned to be more than a data repository and, more importantly, to provide tools for analyzing and using data. The vision also included elements of integration with other databases and multidimensionality in that the database would accommodate data from various toxicogenomic technologies beyond gene expression data. While CEBS (CEBS 2007) is not well populated with data, adding data will not solve its shortcomings; the original goals for the database were not implemented.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

The lack of a sufficient database in the public sector represents a serious obstacle to progress for the academic research community. Development of the needed database could be approached by either critically analyzing and improving the structure of CEBS or starting a new database. Creating an effective database will require the close collaboration of diverse toxicogenomic communities (for example, toxicologists and bioinformatic scientists) who can work together to create "use cases" that help specify how the data will be collected and used, which will dictate the structure of the database that will ultimately be created. These are the first and second of the three steps required to create a database.

1. Create the database, including not only the measurements that will be made in a particular toxicogenomic assay but also how to describe the treatments and other information that will assist in data interpretation.
2. Create software tools to facilitate data entry, retrieval and presentation, and analysis of relationships.
3. Develop a strategy to populate the database with data, including minimum quality standards for data.

Where will the data for such a database come from? Ideally, this should be organized as an international project involving partnership among government, academia, and industrial organizations to generate the appropriate data; an example of such coordinated efforts is the SNP Consortium. However, developing such a database is important enough that it needs to be actively pursued and not delayed for an extensive time.

One potential source of toxicogenomic data is repositories maintained by companies, both commercial toxicogenomic database companies and drug or chemical companies that have developed their own datasets. Although industry groups have been leaders in publishing "demonstration" toxicogenomic studies, the data (e.g., Daston 2004; Moggs et al. 2004; Ulrich et al. 2004; Naciff et al. 2005a,b) published to date are thought to represent a small fraction of the data companies maintain in internal databases that often focused on proprietary compounds. It is unlikely that much of these data will be available in the future without appropriate incentives and resolution of complex issues involving the economic and legal ramifications of releasing safety data on compounds studied in the private sector. Furthermore, extensive data collections even more comprehensive and systematically collected than these maintained by companies are necessary if the field is to advance.

The NTP conducts analyses of exposure, particularly chronic exposures, that would be difficult to replicate elsewhere and could serve as a ready source of biologic material for analysis. One possibility is to build the database on the existing infrastructure of the NTP at the NIEHS; at the least, the NTP should play a role in populating any database.

Although creating and populating a relevant database is a long-term project, work could begin immediately on constructing and populating an initial

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

database to serve as the basis of future development. A first-generation dataset could be organized following the outline in Box 10-1.

This dataset could be analyzed for various applications, including classifying new compounds, identifying new classes of compounds, and searching for mechanistic insight. The preliminary database would also drive second-generation experiments, including a further study in populations larger than those used in the initial studies to assess the generalizability of any results obtained and second-generation experiments examining different doses, different exposure durations, different species and strains, and relevant human exposure, if possible.

A practical challenge of a toxicogenomic data project would be that much of the database development and data mining work are not hypothesis-driven science and generally are not compatible with promotion and career advancement for those in the academic sector. This is despite the fact that creating useful databases, software tools, and proven analysis approaches is essential to the success of any project. If the ability to analyze the data is limited to a small community of computational biologists able to manipulate abstract data formats or run command line programs, then the overall impact of such a project will be minimized. Thus, ways must be found to stimulate the participation of the academic sector in this important work.

Finally, although this may seem like an ambitious project, it should be emphasized that it would only be a start. Moving toxicogenomics from the research laboratory to more widespread use requires establishing useful data and standards that can be used both for validating the approaches and as a reference for future studies.

RECOMMENDATIONS

Short Term

1. Develop a public database to collect and facilitate the analysis of data from different toxicogenomic technologies and associated toxicity data.

2. Identify existing cohorts and ongoing population-based studies that are already collecting samples and data that could be used for toxicogenomic studies (for example, the CDC NHANES effort), and work with the relevant organizations to gain access to samples and data and address ethical and legal issues.[1]

3. Develop standard operating procedures and formats for collecting samples and epidemiologic data that will be used for toxicogenomic studies. Specifically, there should be a standard approach to collecting and storing demographic, lifestyle, occupation, dietary, and clinical (including disease classification) data that can be included and queried with a public database.

---

[1]Ethical and legal issues regarding human subject research are described in Chapter 11.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 10-1 Possible Steps for Developing an Initial Database

1. Select two or more classes of compounds targeting different organ systems. These compounds should be well-characterized compounds with well-understood modes of action.
2. Define appropriate sample populations (for example, a particular mouse strain) in which the effects of the compounds will be measured.
3. Design experiments, including an appropriate range of doses, time course, and level of biologic replication, to ensure that useful data are captured.
4. Select technology (for example, gene expression microarray or proteomics) to be studied. Consider using more than one platform for a particular technology (for example, Affymetrix GeneChips, and Agilent Arrays) or more than one technology (for example, gene expression microarray, and proteomics). Using more than one technology would enable development of methods for comparing and integrating different types of data (for example, gene expression changes and protein data).
5. Develop and implement standards for data collection and reporting, working with communities that specialize in standard development and using private sector, national laboratory, academic, and government agency participation. The raw data, resulting phenotypes, and ancillary information about samples and experimental design should be described with a precise controlled vocabulary.

   While descriptions of information should be along the lines of what has been outlined in MIAME microarray standards, compliance with MIAME microarray standards is insufficient to describe toxicology experiments.

   Adequate description of ancillary information is essential but often overlooked, but it is critical because toxicogenomic technologies are sensitive to a wide range of perturbations, and analysis requires paying careful attention to the complete information about the samples and experiments.
6. Develop a database and its software systems that can capture the relevant information and make it freely available to the scientific community. This requires a significant effort because the quality and ease of any subsequent analysis will largely depend on the design of the database and the information it captures. It is expected that new software tools will need to be developed, and they should be a freely available and open source to facilitate their distribution, use, and continued assessment and development.
7. Develop tools to analyze the results. Although there are many tools for data analysis, at present none fully serves the needs of the toxicogenomic community. Such development must be carried out as a close partnership between laboratory and computational scientists to ensure that the tools developed are both useful and accessible.

Intermediate

4. Develop approaches to populate a database with high-quality toxicogenomic data.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

a. Incorporate existing datasets from animal and human toxicogenomic studies, from both the private and public sectors, into the public data repository, provided that they meet the established criteria.
b. Communicate with the private sector, government, and academia to formulate mutually acceptable and beneficial approaches and policies for appropriate data sharing and data distribution that encourage including data from the private and public sectors.

5. Develop additional analytical tools for integrating epidemiologic and clinical data with toxicogenomic data.

6. NIEHS, in conjunction with other institutions, should cooperate to create a centralized national biorepository for human clinical and epidemiologic samples, building on existing efforts.

Long Term

7. Work with public and private agencies to develop large, well-designed epidemiologic toxicogenomic studies of potentially informative populations—studies that include collecting samples over time and carefully assessing environmental exposures.

Copyright © National Academy of Sciences. All rights reserved.

11

Ethical, Legal, and Social Issues

Toxicogenomic research and its applications will raise many ethical, legal, and social issues. Because toxicogenomics involves the collection and analysis of personal genetic and phenotypic information from large numbers of individuals, it raises more significant ethical, legal and social issues than does, for example, release of reference genome sequences. Although the issues often overlap, they are discussed below in the following five categories: research issues, ethical and social issues, regulatory implications, litigation, and communication and education. The early parts of the chapter cover general concepts and the later parts apply the concepts to toxicogenomics. The following discussion includes examples relating to the efficacy of pharmaceuticals, but similar issues may also apply to genomic-related investigations of the toxicity of pharmaceutical and environmental chemicals.

RESEARCH ISSUES

Research on or using toxicogenomics raises three categories of ethical issues. First, toxicogenomic research not involving human subjects may raise issues common to all biomedical research, including research integrity, conflicts of interest, and commercial relations and disclosures. Second, toxicogenomic research involving human subjects raises generally applicable human subject issues, such as the Common Rule and the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule, equitable selection of subjects, informed consent, privacy and confidentiality, and special protections for vulnerable populations. Third, toxicogenomic research with human subjects involves distinct issues related to the collection and use of biorepositories in research, special issues of informed consent in genetics research, community

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

consultation when doing genetics research in discrete subpopulations, and the duty to notify sample donors and their relatives of research findings.

Research Priorities

One myth about scientific research is that science is value neutral—that new scientific understandings await discovery, that these discoveries have no independent moral significance, and that they take on moral significance only when individuals and groups of individuals assign weight to scientific findings. One flaw in this argument is that there are seemingly limitless areas for scientific inquiry, yet there are finite numbers of scientists and limited resources to pursue research. Therefore, scientists and society must set priorities for research, and those priorities are a function of societal values. Even though scientists often make adventitious discoveries, they generally discover what they are looking for, and what they look for are the things that science and society value discovering.

Pharmacogenomics provides a good illustration. Pharmaceutical researchers may discover numerous polymorphisms in the gene for a particular drug target. Before spending tens or hundreds of millions of dollars on studying a particular polymorphism, any revenue-conscious biotech or pharmaceutical company will try to ascertain the population frequency of these drug targets. Consequently, initial drug development does not focus on genetic variations found most often among individuals in developing countries, where most people do not have the money to pay for basic medicines, let alone expensive new products that treat persons with particular genotypes. Even in developed countries, without some external support from government or private sources, developing drugs directed at rare genotypes is not cost-effective.

Another way to set research priorities is to focus on the nature of the harm to be prevented. It has been asserted that a disproportionate share of health resources (research and treatment) are directed at specific diseases simply because of the effectiveness of advocacy groups working on behalf of affected individuals (Greenberg 2001). Similarly, environmental policy may be influenced by considering certain environmental risks to be of greater societal concern than others. For example, it has been argued that U.S. environmental policy is flawed because we spend millions of dollars on a relatively few Superfund sites in need of remediation but insufficient sums on air and water pollution, which is a more widespread threat to public health (Breyer 1995). Undoubtedly, the targets of toxicogenomic research will be influenced by a myriad of social factors, and priority setting will be influenced by economic and political concerns.

The Federal Rule for the Protection of Human Subjects

The federal Rule for Protection of Human Subjects (45 C.F.R. Part 46, Subpart A) is usually referred to as the Common Rule, because most federal

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

agency sponsors of research in addition to the Department of Health and Human Services (DHHS) have adopted it (DHHS 1997). The Common Rule describes basic procedures and principles for conducting federally sponsored research on human subjects. It applies to researchers who receive federal financial support from a signatory federal agency, research conducted in contemplation of a submission to the Food and Drug Administration (FDA) for approval, and research conducted by an institution that has signed a "multiple project assurance" with DHHS promising to comply with the Common Rule in all research, regardless of the funding source. As a result, the Common Rule applies to much, but not all, research involving human subjects in the United States.

The Common Rule is designed to safeguard the welfare of human subjects of research, and therefore an important preliminary matter is to define "research." The Common Rule defines it as "a systematic investigation, including research development, testing, and evaluation designed to develop or contribute to generalizable knowledge" (45 C.F.R. § 102[d]). Research also includes the development of repositories for future research (DHHS 1997).

The Common Rule does not apply to research conducted on specimens or health records that are not individually identifiable. According to a guidance document issued by the Office for Human Research Protections (OHRP), private information and specimens are not "individually identifiable when they cannot be linked to specific individuals by the investigator(s) either directly or through coding systems" (DHHS 2004). The guidance further provides that research involving only coded private information or specimens does not involve human subjects if the investigator cannot "readily ascertain" the identity of the individual because the key was destroyed before the research began, the key holder has agreed not to release the key to investigators under any circumstances, there are Institutional Review Board (IRB)-approved written policies prohibiting release of the key until the individuals are deceased, or other legal requirements prohibit releasing the key to the investigators until the individuals are deceased. Neither the OHRP guidance nor other authoritative policies address all the major issues surrounding specimen collection and use, including identifiability (McGuire and Gibbs 2006) and the potential for group harm (Foster and Sharp 2002). It is also possible that definitions of "readily identifiable" will need to change with the incorporation of large amounts of polymorphism and similar data in databases.

Health Insurance Portability and Accountability Act

HIPAA (42 U.S.C. §§ 300gg-300gg-2) contains a section called "Administrative Simplification" (Title II, subtitle F, §§ 261-264), which commits the United States to more efficient health claim processing through standard electronic transactions. To protect the confidentiality of protected health information, DHHS was directed to promulgate Standards for Privacy of Individually Identifiable Health Information (Privacy Rule) (45 C.F.R. Parts 160, 164). The Privacy Rule applies to three classes of covered entities: health care providers,

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

health plans, and health clearinghouses. The covered entities under the Privacy Rule and the Common Rule overlap to a great extent (for example, both cover large academic medical center institutions), but they are not of equal scope. The Privacy Rule and the Common Rule also differ on other key issues, including reviews preparatory to research, research involving health records of deceased individuals, and revocation of consents and authorizations (Rothstein 2005a).

Biorepositories

Biorepositories are repositories of human biologic materials collected for research. According to the most recent and widely cited estimate, more than 300 million specimens are stored in the United States, and the number grows by 20 million per year (Eiseman and Haga 1999). Because the ability to access large numbers of well-characterized and annotated samples is an essential part of new genomic research strategies (see Chapter 10), biorepositories are increasingly important (Andrews 2005; Clayton 2005; Deschênes and Sallée 2005; Janger 2005; Knoppers 2005; Majumder 2005; Malinowski 2005; Palmer 2005; Rothstein 2005b). Although biorepositories raise numerous issues involving research ethics and social policy, three issues are especially important to toxicogenomics, pharmacogenomics, and related research.

First, biorepositories collect samples and data for unspecified, unknown, future research and thereby differ from traditional biomedical research, which generally contemplates a single, discrete project for which a single process of informed consent is used. Thus, the question that arises is whether it is permissible under the Common Rule for researchers to obtain informed consent for future uses of specimens collected by a biorepository. Under the Common Rule, a research subject may consent to future, unspecified research. Nevertheless, IRBs generally do not approve "blanket consent" for all possible research on the grounds that consent about unknown future uses by definition is not "informed" (Strouse 2005). Using "tiered" or "layered" consent, research subjects are presented with a menu of possible future research categories, such as cancer research, AIDS research, and genetic research. They can then indicate for which types of research it is permissible to use their sample (NBAC 1999). Because of the growth of biorepositories and new methods of informed consent, researchers need to monitor the understanding of research subjects and the effectiveness of the informed consent process to determine whether changes or improvements in the informed consent process are needed.

Under the HIPAA Privacy Rule, however, each new research project requires a new authorization. Thus, tiered consent does not satisfy the Privacy Rule. Although researchers can obtain a waiver of the authorization requirement (meaning that the researchers do not need to recontact each sample donor for each new research project) (45 C.F.R. §164.512[l][2][ii]), a separate waiver from an IRB or Privacy Board is needed for each proposed new use. Therefore, the inconsistency between the Common Rule and the Privacy Rule complicates

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

compliance for researchers, especially when toxicogenomic research may involve large repositories of individually identifiable specimens that may be used for multiple research projects (Rothstein 2005a).

A second key ethical issue about biorepositories concerns if or when researchers have a duty to recontact sample donors to notify them that research using their samples has identified health risk information of clinical or social significance to them or their relatives (Rothstein 2005b; Bookman et al. 2006; Renegar et al. 2006). If the samples are deidentified, it will be impossible to notify any individual of sample-specific research findings, but there still may be a duty to notify all sample donors of the overall conclusions of the study and the potential desirability of consulting a physician. The greatest challenge, however, involves individually identifiable samples. Because it is too late to develop a recontact policy after a discovery has been made, the most prudent course is for researchers to ask sample donors at the time of sample collection if they want to be recontacted in certain circumstances (for example, when there is a predictive test or therapy) and, if so, whether the contact should be through their physician or personally.

A third set of issues on which there has been considerable debate in the bioethics literature, is who owns the property rights to tissue samples and whether researchers have a duty to undertake any type of "benefit sharing" with individual donors or groups of donors (Andrews 2005; Marchant 2005). For example, over and above informed consent, who ultimately controls the use, distribution, and disposal of donated tissue samples? In one of the first judicial decisions to address this issue, a federal district court recently held that the research institute, not the individual researcher or tissue donor, "owns" donated tissue samples (Washington University v. Catalona, 2006 U.S. Dist. LEXIS 22969 [E.D. Mo. 2006]). Some tissue donors, especially those who are members of identifiable groups such as Native Americans, are now insisting that they retain the property rights in their donated tissues as a precondition for providing samples for research (Marchant 2005).

A related question concerns those instances in which research on biorepository samples has led to commercially valuable discoveries: are sample donors entitled to some share of the proceeds (see Greenberg v. Miami Children's Hospital Research Institute, Inc., 264 F. Supp. 2d 1064 [S.D. Fla. 2003]; Moore v. Regents of Univ. of Cal., 793 p.2d 479 (Cal. 1990)? There have been a few documented cases in which one individual's sample has been extremely valuable in developing cell lines or other research products (Moore v. Regents of Univ. of Cal., 793 p.2d 479 [Cal. 1990]; Washington 1994). It is far more likely, however, for analyses of numerous samples to be needed for development of any commercially viable finding. Whether it is ethical for individual research subjects to waive all their interests in the products of discovery remains unclear, although it is common for informed consent documents to so provide. Regardless of the language in the informed consent document, the principle of benefit sharing would be satisfied if researchers declare in advance that they will set aside a small percentage of revenues derived from commercialization for dona-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

tion to a charitable organization that, for example, provides health care to indigent individuals with the condition that is the target of the research (HUGO 2000; Knoppers 2005).

Intellectual Property

The development, exploitation, and protection of intellectual property are important elements of contemporary research strategy. Many of the contentious intellectual property issues, such as publication, timing of filing, ownership, and licensing, arise "downstream," after the discovery of a patentable invention. These issues are not unique to toxicogenomic and pharmacogenomic research. Nevertheless, because many toxicogenomic applications involve assays that produce large amounts of data, often involving large numbers of genes or proteins, obtaining intellectual property rights to all the materials used in an assay could be particularly burdensome and problematic for researchers and platform manufacturers in the field of toxicogenomics (NRC 2005b).

An important, "upstream" issue, and one that needs to be addressed specifically in the context of toxicogenomic and pharmacogenomic research, is whether to have an "open access" policy to raw samples and data and, if so, how open it should be and how it would work. Traditionally, both toxicologic and pharmaceutical research have been undertaken with closed access, with private companies maintaining proprietary control over their research until a patent has been filed or a company has decided to disclose information for some other reason. Increasingly, however, there is pressure to make preliminary data from government-funded (and even some privately funded) research more widely available to researchers in general. For example, the Human Genome Project (public), the Pharmacogenetics Research Network (public), and the SNP Consortium (private) have adopted the policy of making research findings available online promptly for other researchers to use.

The Committee on Emerging Issues and Data on Environmental Contaminants of the National Research Council held a Workshop on Intellectual Property in June 2006 at the National Academies. Among other issues, the workshop discussed the position of the National Institutes of Health (NIH) in support of immediate (or prompt) release of and free access to human genome sequence data. These policies, embodied in the so-called Bermuda Principles, have been endorsed by the International Human Genome Organization and other organizations (HUGO 1997). They also have been applied more broadly, and they form the basis of the National Human Genome Research Institute's policies of prompt or immediate disclosure of genomic information generated by NIH-supported research or collected in NIH-supported repositories.

No standard policy for data release has been developed for toxicogenomic and pharmacogenomic research. In the absence of such a policy, the default position is nondisclosure. Consequently, affirmative steps should be undertaken to

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

ensure that toxicogenomic and pharmacogenomic data are promptly and freely made available to all researchers whenever possible.

Conflicts of Interest

Environmental, pharmaceutical, and occupational health research often is sponsored or undertaken by entities with an economic interest in the outcome of the study. Especially for sensitive research, such as genetic research involving vulnerable populations, concerns about the investigators' actual or perceived conflicts of interest will be magnified. Transparency, informed consent, and external oversight are essential (Rothstein 2000a). Even with such precautions, conflicts may arise with regard to what actions to take on the basis of the study.

ETHICAL AND SOCIAL ISSUES

A number of ethical and social issues may apply to toxicogenomics. These issues include privacy and confidentiality, issues related to socially vulnerable populations, health insurance discrimination, employment discrimination, individual responsibility, issues related to race and ethnicity, and implementation.

Privacy and Confidentiality

Privacy has many dimensions. In the information sense, privacy is the right of an individual to prevent the disclosure of certain information to another individual or entity (Allen 1997). Increasingly, the bioethics literature also has recognized a negative right of informational privacy—that is, the right "not to know" certain information about oneself (Chadwick 2004). Toxicogenomics affects both of these aspects of privacy. Individuals may want to restrict the disclosure of their toxicogenomic or pharmacogenomic information that was obtained in a research, clinical, workplace, or other setting. They also may not want to know of certain risks, especially when nothing can be done to prevent the risk of harm (for example, where there was past occupational or environmental exposure).

Confidentiality refers to a situation in which information obtained or disclosed within a confidential relationship (for example, physician-patient) generally will not be redisclosed without the permission of the individual (Rothstein 1997). With regard to toxicogenomics, the most important relationship is the physician-patient relationship. Restrictions on redisclosure are an important ethical precept of health care professionals, and maintaining the confidentiality of health information is important in preventing intrinsic and consequential harm to individuals (Orentlicher 1997). Although toxicogenomic and pharmacogenomic information would be covered under the "genetic privacy" laws several states have enacted (NCSL 2005a), these laws afford limited and inconsistent protection.

Copyright © National Academy of Sciences. All rights reserved.

It should also be noted that DHHS is taking the lead in promoting widespread adoption of electronic health records and creation of a national infrastructure to support a system of interoperable, longitudinal, comprehensive health records (NCVHS 2006). In such an environment, privacy and confidentiality protections are even more important because detailed health information can be disclosed to numerous sources instantaneously (Rothstein and Talbott 2006).

Although it is important to safeguard the security of health records to prevent *unauthorized* disclosure of information, breaches of confidentiality through *authorized* disclosures may be an even greater problem. For example, in health care settings, not all employees (for example, billing clerks, meal service employees, maintenance employees) need the same degree of access to patient health records. Thus, the degree of access needs to comport with the need to know about protected health information. "Role-based access restrictions" help to limit the scope of disclosure.

In non-health care settings, individuals often are required to execute an authorization for disclosure of their medical records as a condition of employment or insurance. Development and application of "contextual access criteria" will help to limit these disclosures to the health information relevant to the purpose of the disclosure (for example, ability to perform a job) (Rothstein and Talbott 2006).

Socially Vulnerable Populations

Many genetic loci of actual or potential significance to toxicogenomics and pharmacogenomics have a differential allelic frequency in discrete subpopulations. Due to ancestral patterns of endogamy (marrying within a social group), migration, geographic isolation, and other elements of ancestral origins, some genetic traits have a higher frequency among subpopulations socially defined by race or ethnicity. Thus, there is great potential for stigma when an increased risk of a particular undesirable health condition is associated with a particular population group, especially when the group is a racial or ethnic minority in a society. These population groups are then said to be "vulnerable" with respect to a particular health condition.

A socially vulnerable population also can be based on shared somatic mutations. For example, workers with an acquired mutation or biomarker based on a particular occupational exposure or residents of a certain area with toxic exposures who demonstrated the subclinical effects of a particular exposure may also be said to be socially vulnerable populations. These polymorphisms may not be phenotypically obvious and often do not correlate with race, ethnicity, and other traditional social categories.

The notion of vulnerable populations based on toxicogenomics and pharmacogenomics raises numerous ethical concerns, including the ethical principles of justice, respect for persons, and beneficence. Because of the social risks to vulnerable populations from research, special efforts are needed to obtain infor-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

mation about community interests and concerns. Community engagement and consultation are essential elements of ethical research (Foster and Sharp 2002). In addition, if environmental or pharmaceutical exposures have already adversely affected a subpopulation or a subpopulation has been identified to be at greater risk, ethical considerations might require that vulnerable populations have fair access to health care for diagnostic and treatment measures; that steps be taken to prevent unfair discrimination in employment, insurance, and other opportunities; that there is adequate compensation for harms; that feasible remediation be undertaken; and that effective protections are in place for protecting privacy and confidentiality (Weijer and Miller 2004).

Health Insurance Discrimination

The leading concern among individuals at a genetically increased risk of illness is that they will be unable to obtain or retain their health insurance (Rothstein and Hornung 2004). In the United States, individuals obtain health care coverage mostly through government-funded programs (for example, Medicare, Medicaid) or employer-sponsored group health insurance. A relatively small percentage (about 10%) of those with health coverage have individual health insurance policies. Although this group is the only one in which individual medical underwriting takes place, individuals who currently have group coverage might be legitimately concerned that if they lost their job and their group health insurance, they would need to obtain individual health insurance.

To address the concern about discrimination in health insurance, nearly every state has enacted a law prohibiting genetic discrimination in health insurance (NCSL 2005b). These laws have a fatal flaw: they prohibit discrimination only against individuals who are asymptomatic. If an individual later becomes affected by a condition, he or she is no longer protected by the law and is at risk of rate increases or even cancellation in accordance with general provisions of state insurance law (Rothstein 1998).

Health insurance discrimination is very difficult to resolve. As long as a key element of our system of health care finance is individually underwritten, optional, health insurance, then it will contain less favorable access for individuals at an increased risk of future illness. Meaningful reform would require major changes in health care financing.

Employment Discrimination

Employers have two concerns about employing individuals who are at a genetic risk of future illness. First, employee illness and disability of any etiology represents significant costs in terms of lost productivity, lost work time, high turnover, and increased health care costs. Genetic testing makes it possible to predict future risks and would make it more attractive for employers to exclude at-risk individuals from their workforces. To combat such discrimination,

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

about two-thirds of the states have enacted laws prohibiting genetic discrimination in employment (NCSL 2005c). These laws are seriously deficient because, although they make discrimination based on genetic information illegal, they do not prohibit employers from lawfully obtaining genetic information contained within the comprehensive health records that employers may require individuals to disclose as a condition of employment. Consequently, at-risk individuals, who might benefit from genetic testing, are often reluctant to do so because a future employer can lawfully obtain the results (Rothstein and Hornung 2004).

Second, employers may be reluctant to hire or assign individuals to work where they are at an increased risk of occupational illness because of their genetics. In addition to the employers' illness concerns listed above, occupational illness could result in additional costs based on workers' compensation, compliance with the Occupational Safety and Health Act of 1970, personal injury litigation, and reduced employee morale. With regard to toxicogenomics in the workplace, is it lawful or ethical for employers to request or require workers to undergo genetic testing? See Box 11-1 for additional discussion on toxicogenomic research in the workplace.

There are two ethical cross-currents in the laws regulating the workplace: paternalism and autonomy. Much workplace health and safety regulation may be characterized as paternalistic, including child labor laws, minimum wage laws, and the Occupational Safety and Health Act. A greater appreciation for worker autonomy in the workplace is reflected in the Supreme Court's leading decision in International Union, UAW v. Johnson Controls, Inc. (1991). The Court held that the employer discriminated on the basis of sex, in violation of Title VII of the Civil Rights Act of 1964 (42 U.S.C. § 2000e), by excluding all fertile women from jobs with exposure to inorganic lead because of concerns that a woman could become pregnant and give birth to a child with deformities caused by maternal workplace exposures. According to the Court: "Congress made clear that the decision to become pregnant or to work while being either pregnant or capable of becoming pregnant was reserved for each individual woman to make for herself" (499 U.S. at 206).

A middle position between permitting employers to mandate genetic testing and prohibiting all employee genetic testing is that a worker who is currently capable of performing the job should have the option of learning whether he or she is at increased risk of occupational disease based on genetic factors. Optional genetic testing would be conducted by a physician or laboratory of the individual's choice; the testing would be paid for by the employer; and the results would be available only to the individual. The individual would then have the option of deciding whether to assume any increased risk of exposure (Rothstein 2000b). Only if employment of the individual created a direct and immediate threat of harm to the individual or the public would the employer be justified in excluding the individual based on future risk (Chevron U.S.A. Inc v. Echazabal 2002). See Box 11-2 for additional discussion on genetic screening in the workplace.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 11-1 Toxicogenomic Research on Workers

Some employers that use toxic substances in the workplace have considered or have engaged in research on their workers to determine whether exposures are causing physiologic changes. Toxicogenomics increases the potential scope of research to include the possible effects of individual genetic variability. The Federal Regulations for the Protection of Human Subjects, generally called the "Common Rule," specifically provide: "When some or all of the subjects are likely to be vulnerable to coercion or undue influence, such as children, prisoners, pregnant women, mentally disabled persons, or economically or educationally disadvantaged persons, additional safeguards [should be] included in the study to protect the rights and welfare of these subjects" (45 C.F.R. § 46.111[b]).

Although research conducted by private employers is unlikely to be subject to federal regulations, the underlying principles of ethical conduct of research suggest that special steps are undertaken to prevent coercion and to safeguard the rights and welfare of workers. It has been suggested that the following guidelines, intended to supplement the Common Rule, should apply to medical research involving workers. These guidelines appear to have equal force when applied in the context of toxicogenomic research.

1. If possible, the research should be performed by a party other than the employer, such as a university, medical institution, nongovernmental organization, or government agency.
2. Employers and employees (including union representatives) should be involved from the beginning in developing all aspects of the study, including study design, recruitment practices, criteria for inclusion and exclusion, informed consent process, confidentiality rules, and dissemination of findings.
3. The sponsor of the research should be indicated to potential participants, and investigators should disclose any financial conflicts of interest in the research.
4. Individuals with supervisory authority over potential research participants should not be involved in the recruitment process, and lists of research participants should not be shared with supervisors.
5. No inducements should be offered for participation in the research.
6. The informed consent process should make it clear to potential participants that there will be no adverse employment consequences for declining to participate or withdrawing from the research; potential participants also should be informed whether treatment or compensation for injuries will be provided.
7. Research should be conducted, and results should be disclosed, in the least identifiable form consistent with sound scientific methodology.
8. If the investigators believe the findings will be of sufficient scientific validity and clinical utility to warrant offering participants the opportunity to obtain their individual results, the participants should be advised of all the potential risks of disclosure, including any psychologic, social, or economic risks.
9. Reasonable steps should be taken to ensure the confidentiality of participant-specific information generated by the research, including storing the information in secure locations separate from other employment records, destroying biologic samples that are no longer needed, destroying linking codes when they are no

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

longer needed, and applying for a certificate of confidentiality from the Department of Health and Human Services when appropriate.

10. Investigators should take special precautions at all stages of the study if the research has the potential to adversely affect groups of individuals on the basis of race, ethnicity, gender, age, or similar factors.

Source: Rothstein 2000a.

Individual Responsibility

To what extent should individuals have a responsibility to learn if they are at increased risk from certain exposures, to avoid those exposures, and to deal with the consequences of illness caused by those exposures? To what extent does society have a duty to respect the autonomy of individuals to decline to learn of possible increased risks or to accept those risks? To what extent is society obligated to provide special protections for such individuals and health care if they become sick?

Several of these policy choices have been settled in the workplace setting for some time. Under workers' compensation law, an employer "takes the employee as is," and therefore it is no defense that the individual was at an increased risk of illness (Perry v. Workers' Comp. App. Bd. 1997). Under the Occupational Safety and Health Act, health standards, at least in theory, are designed to protect all workers (29 U.S.C. § 655 [c][5]). Under HIPAA, employer-sponsored group health plans may not charge employees different rates or have different levels of coverage based on their health status, including genetic predisposition (42 U.S.C. § 300gg-1[a] and [b]). Nevertheless, even with regard to employment, many issues remain to be resolved, including the role of individual variation in setting permissible exposure levels.

It may be even more difficult to develop rules of general applicability beyond the workplace setting. Both scientific evidence (for example, absolute risk, relative risk, severity, latency, treatability) and context-specific social conceptions of reasonable conduct in the face of increased risk are likely to affect the development of public policy. For example, if members of a low-income family with a genetic predisposition to illness and with few options of places to live "voluntarily" agree to accept the increased risk of living in the vicinity of toxic emissions, it is doubtful that society would shift the moral blame from the polluter to the individuals harmed by the pollutants. On the other hand, if toxicogenomic testing indicated that a particular individual would be much more likely to suffer serious harm from exposure to tobacco smoke and the individual started smoking or made no effort to stop, it is likely that at least some of the moral blame would shift to the cigarette smoker (Christiani et al. 2001).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 11-2 Genetic Screening for Susceptibility to Chronic Beryllium Disease

Beryllium is a lightweight, stable, and strong metal that is ideal for use in the aerospace, nuclear, electronics, and other industries. Workplace exposure to beryllium can result in chronic beryllium disease (CBD), a granulomatous lung disease occurring in 2% to 16% of exposed workers (Fontenot and Maier 2005). Researchers have identified an increased risk of CBD among workers with the allelic substitution of glutamate for lysine at position 69 in the *HLA-DPB1* gene on chromosome 6. This marker, known as $Glu^{69\ P}$, has a population prevalence of 40% (McCanlies et al. 2003).

The Brush Wellman Company, the leading commercial beryllium-processing company, adopted a pilot program of genetic screening for the $Glu^{69\ P}$ marker in job applicants and employees before exposure to beryllium. The program featured the following elements: testing was voluntary, informed consent was obtained before testing, testing was performed at the employer's expense by an independent laboratory under contract with the company, results were provided directly to the individuals, management received no individual test results, and counseling was available to workers upon request. The decision to accept employment rested with the worker.

Although the testing program for beryllium susceptibility incorporated the essential elements of voluntariness, confidentiality, and autonomy, the program had limitations that may be instructive in the development of similar screening programs (Holtzman 1996). First, the test's positive predictive value was only about 11%. Second, the presence of the Glu^{69} marker is not the only risk factor for development of CBD and the absence of the marker may give workers a false sense of protection. Third, in any screening program, care must be taken to ensure that the individual's decision to undergo testing is truly informed and voluntary and that confidentiality of participation and results is maintained. Fourth, screening out workers from exposure, even indirectly, is no substitute for industrial hygiene controls to ensure the health of all workers.

Race and Ethnicity

Concerns about race and ethnicity are subsumed under the heading of "socially vulnerable populations," discussed above, but they are raised in distinctive ways by pharmacogenomics. All humans are genetically 99.9% alike. There is more genetic diversity within certain socially defined groups (for example, race, ethnicity) than between groups (Ossorio and Duster 2005). The individual and group differences, however, are measurable and may be significant. Among the traits for which subpopulation variance has been observed are those dealing with pharmacokinetics and pharmacodynamics (Rothstein and Epps 2001a).

Although pharmacogenomics holds great promise for increasing drug efficacy and decreasing adverse drug events, there is a risk of "racializing" medicine. It is facile to rely on race as a surrogate for genotype, but the social cost of

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

doing so may be high. "By heedlessly equating race with genetic variation and genetic variation with genotype-based medications, we risk developing an oversimplified view of race-specific medications and a misleading view of the scientific significance of race" (Rothstein 2003, p. 330). The controversy surrounding the FDA's approval of the drug BiDil in 2005 for self-identified African Americans demonstrates the currency of the issue (Saul 2005). See Box 11-3.

Implementation Issues

Traditionally, pharmaceutical companies have searched for new blockbuster drugs to provide effective treatments for common conditions, such as hypertension, hypercholesterolemia, and musculoskeletal pain. A single product is designed to be safe and effective for substantially all members of the adult population. With pharmacogenomics, however, the potential market is segmented by genotype and many different companies may develop products for each affected population. Although virtually every biotech and pharmaceutical company is using genomic technologies in researching new products, whether "personalized medicine" will prove to be a viable business model remains to be seen (Reeder and Dickson 2003; Royal Society 2005; Ginsburg and Angrist 2006).

Once medications that target people with particular genotypes are developed, it is not clear how widely they will be adopted. Targeted medications may eliminate the costly trial and error of current prescribing and dosing, but the substantial research and development costs almost certainly will result in more expensive medications. Whether private and public payers will include these therapies on their formularies will depend on several factors, including the condition being treated, the relative improvement in safety and efficacy over standard therapies, and cost (La Caze 2005). Patient demand produced by direct-to-consumer advertising (Ausness 2002) and provider concerns about potential liability (Palmer 2003) undoubtedly will increase adoption of the new products.

New technologies will also expand the responsibilities of physicians, nurses, pharmacists, and other health care providers to use genetic information in prescribing, dispensing, and administering medications (Rothstein and Epps 2001b). Professionals not only will need to integrate genetic information into their practices, they also will need to become competent in genetic counseling (Clayton 2003). In addition, even though genetic testing for medication purposes is more limited in scope than genetic tests for assessment or prediction of health status, the test results will still contain sensitive genetic information. For example, identifying someone as being more difficult or expensive to treat could lead to discrimination in employment or insurance. Consequently, pharmacogenomics could increase the demand for new protections for the confidentiality of health information.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 11-3 BiDil and Race-Linked Pharmaceuticals

BiDil is a combination of two drugs, hydralazine and isosorbide (nitroglycerin), that have been available in generic form for decades. Researchers postulated that this combination of drugs would benefit individuals with end-stage cardiovascular disease. In 1997, the FDA rejected the drug (based on a lack of efficacy) after a trial in a mixed-race group, although a subgroup of African Americans appeared to show benefit.

According to Nitromed, Inc. (which now has rights to BiDil), African Americans are more likely than white Americans to have decreased amounts of nitric oxide, and BiDil is a nitric oxide enhancer. This was the biologic hypothesis for an improved outcome in this subpopulation. Nitromed joined with the Association of Black Cardiologists to sponsor the African American Heart Failure Trial, which involved 1,050 patients at 170 sites. On July 19, 2004, the trial was halted because of the significant success of patients enrolled in the treatment arm of the trial. Subjects treated with BiDil demonstrated a 43% improvement in survival and a 33% reduction in first hospitalization. A patent was issued for BiDil on August 31, 2004.

In June 2005, based on the assumption that BiDil has a differential efficacy, the FDA approved Nitromed's new drug application for use only by self-identified African Americans. Even if there is a polymorphism with some role in this process, it is scientifically implausible either that the genotype would be found exclusively in this population group or that members of the group would have only this polymorphism. The BiDil approval is very controversial because race is being used as a crude surrogate for an assumed but unknown genotype. Genotyping efforts in different populations (for example, African, European, and Asian) have demonstrated that most human genetic variation is shared across human populations, with different allele frequencies between populations. Moreover, most functional human genetic variation is not population specific (Hinds et al. 2005 and references therein). Thus, there is no scientific basis for using race as an indicator of genetic variability or as a surrogate for a specific genotype. According to one expert, "Race as a proxy for genotype should never count as a justification acceptable to regulatory authorities" (Ossorio 2004, p. 424). In contemplating more widespread application of race in pharmacogenomics, substantial scientific, ethical, social, and legal issues need to be resolved (Kahn 2005).

REGULATORY IMPLICATIONS

Toxicogenomic data have numerous potential applications in environmental, pharmaceutical, and occupational health regulation (Rothstein 2000b; Baram 2001; Marchant 2003a,b; Salerno and Lesko 2004; Grodsky 2005; Gallagher et al. 2006). Regulatory agencies such as the Environmental Protection Agency (EPA) and the FDA have already produced guidance documents on the incorporation of genomic information into their regulatory programs (see Box

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

11-4). Some examples of potential regulatory applications of toxicogenomics and their implications are discussed below.

Defining New Adverse Effects

Many regulatory requirements are based on a finding of "adverse effect." For example, national ambient air-quality standards set under the Clean Air Act are set at a level that protects against adverse effects in susceptible populations. Some subclinical effects (for example, changes in erythrocyte protoporphyrin concentration in blood) have been found to be adverse effects under this statutory provision and have required more stringent standards to prevent such effects from occurring in exposed populations (Lead Industry Association v. EPA 1980). Furthermore, many EPA regulations for noncarcinogenic substances are based on the reference concentration (RfC) or reference dose (RfD). Some gene expression or other changes measured with toxicogenomic technologies may constitute new adverse effects under these programs and thus lower the RfDs and RfCs, resulting in more stringent regulations. Manufacturers of pesticides and toxic chemicals are required to notify the EPA of new scientific findings about adverse effects associated with their products. In pharmaceutical regulation, the detection of a biomarker that suggests an adverse effect may likewise trigger additional regulatory scrutiny or restrictions.

Therefore, a critical issue in the regulatory application of toxicogenomics will be determining whether and when a change constitutes an adverse effect. Many changes in gene expression, protein levels, and metabolite profiles will be adaptive responses to a stimulus that are not representative or predictive of a toxic response. Other toxicogenomic changes may be strongly indicative of a toxic response. Therefore, it will be important to distinguish true biomarkers of toxicity from reversible or adaptive responses and to do so in a way that is transparent, predictable, and consistent for the affected entities (Gallagher et al. 2006). At least initially, phenotypic anchoring of toxicogenomic changes to well-established toxicologic end points will likely be necessary to identify toxicologically significant markers.

Regulatory Decisions Based on Screening Assays

The availability of relatively inexpensive and quick toxicogenomic assays that can be used for hazard characterization of otherwise untested materials offers a number of potential regulatory opportunities (Gallagher et al. 2006). Full toxicologic characterization of the approximately 80,000 chemicals in commerce using a battery of traditional toxicology tests (for example, chronic rodent bioassay) is not economically or technologically feasible in the foreseeable future. Screening with a toxicogenomic assay, such as a gene expression analysis that classifies agents based on transcript profiles, might be useful to quickly

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BOX 11-4 Regulatory Agency Guidance Documents

EPA: The EPA adopted its "Interim Guidance Policy" in June 2002 (EPA 2002). This Interim Policy states that the EPA expects to receive genomic data for various risk assessment and regulatory purposes and expresses the agency's belief "that genomics will have an enormous impact on our ability to assess the risk from exposure to stressors and ultimately to improve our risk assessments." Specifically, the EPA identifies potential applications of genomic data to include better understanding of the relationship between exposure, mechanisms of action, and adverse effects in risk assessment as well as in setting agency priorities, ranking chemicals for testing programs, and supporting possible regulatory actions. The EPA expressly states that it will accept and consider genomic data in making regulatory decisions "on a case-by-case basis." However, although "genomics data may be considered in decision-making at this time, these data alone are insufficient as a basis for decisions." In other words, whereas the EPA will consider available genomic data as part of a weight-of-evidence risk assessment, it will not base regulatory decisions solely on genomic data at this time.

Two years later, the EPA elaborated on the potential applications of genomics for risk assessment and regulation purposes in a white paper issued in December 2004 entitled "Potential Implications of Genomics for Regulatory and Risk Assessment Applications at EPA" (EPA 2004b). This 50+-page document identified four major areas where the EPA could use genomic data: (1) prioritizing contaminants and contaminated sites, (2) monitoring, (3) reporting provisions, and (4) assessing risk. The document describes potential, and in some cases actual, current applications of genomic data in these four areas.

FDA: The FDA issued a "Guidance for Industry: Pharmacogenomic Data Submissions" in March 2005 (FDA 2005a). This guidance seeks to encourage the collection and use of pharmacogenomic data in drug development by providing the agency's position on when and how pharmacogenomic data should be submitted to the FDA as part of the approval process for drugs or biologics, and how such pharmacogenomic data may be used in FDA regulatory decisions. The guidance sets forth "algorithms" describing submission requirements for investigational applications, unapproved marketing applications, and approved marketing applications. Data on "known valid biomarkers" must be submitted to the FDA for each stage of approval, whereas probable biomarker data generally need to be submitted only at the final approval stage and not at the investigational approval stage for a drug or biologic. The guidance document recognizes that most pharmacogenomic data will be "exploratory" at this time and will not meet the definition of a valid known biomarker. The FDA guidance encourages the voluntary submission of such exploratory pharmacogenomic data and provides assurances that the agency will use such data only for learning purposes and not for regulatory purposes. The FDA has issued other draft guidance documents to support the development of pharmacogenomic data and products, such as its Drug-Diagnostic Co-Development Concept Paper (FDA 2005b) to encourage a drug and a related pharmacogenomic test to be developed and obtain regulatory approval together.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

screen chemicals for prioritizing substances for further investigation and possible regulation (see Chapter 5). Another possibility is to require the manufacturer of a new chemical substance, or an existing substance that will be used in a new application, to include the results of a standardized toxicogenomic assay as part of a premanufacturing notice required under section 5 of the Toxic Substances Control Act (Marchant 2003a).

Protection of Genetically Susceptible Individuals

Many regulatory programs specifically require protection of susceptible individuals. For example, the Clean Air Act requires that national ambient air-quality standards be set at a level that protects the most susceptible subgroups within the population. Under this program, the EPA focuses its standard-setting activities on susceptible subgroups such as children with asthma. Recent studies indicate a significant genetic role in susceptibility to air pollution (Kleeberger 2003), which may lead to air-quality standards being based on the risks to genetically susceptible individuals (Marchant 2003b; Grodsky 2005). Regulations under other environmental statutes, such as pesticide regulations under the Food Quality Protection Act and drinking water standards under the Safe Drinking Water Act, may likewise focus on genetically susceptible individuals in the future, as might occupational exposure standards promulgated by the Occupational Safety and Health Administration. Likewise, pharmaceutical approvals may require considering and protecting individuals with genetic susceptibilities to a particular drug.

The identification of genetic susceptibilities to chemicals, consumer products, pharmaceuticals, and other materials raises a number of regulatory issues. One issue is the question of the feasibility of protecting genetically susceptible individuals. On the one hand, protecting the most susceptible individuals in society may be extremely costly, and perhaps even infeasible without major, formidable changes in our industrial society. On the other hand, the concept of government regulators leaving the health of some individuals unprotected, who through no fault of their own are born with a susceptibility to a particular product or chemical, also seems politically and ethically infeasible (Cranor 1997). As more information on individual genetic susceptibility becomes available, regulators and society generally will confront difficult challenges in deciding whether and how to protect the most genetically vulnerable citizens in our midst (Grodsky 2005).

A related set of issues will involve the extent to which we rely on societal regulation versus individual self-help in protecting genetically susceptible individuals. All regulatory programs confront choices between societal regulation and targeting high-risk individuals (Rose 1994), but these choices will be heightened by the identification of genetic susceptibilities to specific products or substances. For example, should products be labeled to warn people with particular genotypes to avoid using the product, just as diet sodas are labeled to

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

warn people with the genetic disease phenylketonuria that the product contains phenylalanine? Should regulators approve for sale pharmaceuticals that are effective only for individuals with compatible genotypes and are ineffective or even hazardous for other people? Under what conditions and with what types of warnings? Will workers with genetic susceptibilities to specific workplace hazards be precluded from working in those jobs, or will they be required to accept personal responsibility for any harms that result? In all these scenarios relying on self-help, how will susceptible individuals discover they carry a particular genetic susceptibility? How will the privacy and confidentiality issues associated with such genetic knowledge be addressed? How well will people understand and adapt their behavior appropriately to information on their genetic susceptibilities? Policymakers, scientists, regulators, and other interested parties will need to address these and other questions raised by new knowledge about genetic susceptibilities.

Generic Regulatory Challenges

Several generic issues will confront all regulatory programs that use toxicogenomic data. One issue will be ensuring that toxicogenomic data are adequately validated (see Chapter 9). Each agency has its own procedures and criteria for ensuring data quality. In deciding whether and when toxicogenomic data are ready for "prime time" application in formulating federal regulations, agencies must balance the risk of premature use of inadequately validated data versus the harm from unduly delaying the use of relevant data by overly cautious policies. Agencies such as the EPA have been criticized for being too conservative in accepting new types of toxicologic data (NRC 1994).

Another trade-off that agencies must face involves whether and when to standardize toxicogenomic platforms and assays (Gallagher et al. 2006). On the one hand, standardization is important in that it facilitates comparison between datasets (see Chapter 3) on different substances and ensures consistent treatment of different substances. On the other hand, standardization runs the risks of prematurely freezing technology before it has fully matured.

A related issue is how regulatory agencies can encourage or require private parties to generate and submit toxicogenomic data. Many product manufacturers are likely to be concerned that toxicogenomic markers may detect biologic effects from their products or emissions at lower doses that may lead to increased regulatory scrutiny. To address such disincentives, agencies have adopted different approaches to encourage data submission. The FDA has adopted an approach under which the voluntary submission of certain types of toxicogenomic data will be for informational purposes only to help regulators and regulated parties better understand toxicogenomic responses, with an assurance that the data will not be used for regulatory purposes (Salerno and Lesko 2004). The EPA has adopted an interim policy that it will not base regulatory

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

decisions solely on genomic data, alleviating concerns that a new regulatory requirement may be imposed based solely on a toxicogenomic finding when no other indication of toxicity is present (EPA 2002).

LITIGATION

Toxic tort litigation involves lawsuits in which one or more individuals (the plaintiffs) who have been allegedly injured by exposure to a toxic agent sue the entity responsible for that exposure (the defendant) for compensation. Both plaintiffs and defendants are likely to seek to use toxicogenomic data for various purposes in future toxic tort litigation (Marchant 2002, 2006).

Proving Exposure

Many toxic tort claims involving exposures to environmental pollutants fail because the plaintiffs are unable to adequately demonstrate and quantify their exposure to a toxic agent. Individuals exposed to contaminated drinking water, hazardous chemicals in the workplace, or toxic releases from an environmental accident often lack access to objective environmental monitoring data that can be used to quantify their exposures. In such cases, courts often dismiss claims because the plaintiffs are unable to prove sufficient exposure with objective data (Wright v. Willamette Indus 1996). Some courts have endorsed, in principle, the possibility of using genetic biomarkers (for example, chromosomal aberrations) rather than environmental monitoring data to demonstrate and even quantify exposure (in re TMI litigation, 193 F. 3d 613, 622-623 [3d Cir. 1999] cert. denied, 530 U.S. 1225 [2000]).

Toxicogenomic data may be able to help prove or disprove exposure in appropriate cases. If a particular toxic agent creates a chemical-specific fingerprint of DNA transcripts, protein changes, or metabolic alterations, those biomarkers potentially could be used to demonstrate and perhaps even quantify exposure. Alternatively, a defendant could argue that the lack of a chemical-specific toxicogenomic marker in an individual proves that there was not sufficient exposure. However, a number of technical obstacles and data requirements need to be addressed before toxicogenomic data can be used reliably in this manner (see Chapter 4). The chemical-specific attribution of the toxicogenomic change would need to be validated, as would the platform used to quantify the changes. The potential variability in expression between cell types and individuals would also need to be addressed. Perhaps most significantly, the time course of toxicogenomic changes during and after exposure would need to be understood to correctly extrapolate an individual's exposure history from after-the-fact measurements of toxicogenomic changes (Marchant 2002).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Causation

The other major evidentiary challenge that plaintiffs in toxic exposure cases must overcome is to prove causation. Plaintiffs bear the burden of proof to demonstrate both general causation and specific causation. General causation involves whether the hazardous agent to which the plaintiff was exposed is capable of causing the adverse health effect the plaintiff has incurred. Specific causation concerns whether the exposure did in fact cause the health effect in that particular plaintiff. Courts generally consider these questions separately, and both inquiries frequently suffer from a lack of direct evidence, resulting in outcomes that are highly uncertain, speculative, and often unfair. Toxicogenomic data have potential applications for both general causation and specific causation.

General causation determinations are impaired by "toxic ignorance" (NRC 1984; EDF 1997), in which valid scientific studies do not exist for many combinations of toxic substances and specific health end points. The lack of a valid published study evaluating whether the agent to which the plaintiff was exposed can cause the plaintiff's health condition will generally bar tort recovery for failure to demonstrate general causation. Toxicogenomic assays that can reliably be used for hazard identification may provide a relatively inexpensive and quick test result that could be used to fill the gaps in general causation. For example, a gene expression assay that can identify a particular type of carcinogen might be used to classify a chemical as a carcinogen (and hence establish general causation) in the absence of a traditional chronic rodent bioassay for carcinogenicity.

Toxicogenomic data can also play a role in proving or disproving specific causation. Current toxicologic methods are generally incapable of determining whether a toxic agent caused an adverse health effect in a specific individual for all but so-called "signature diseases" that usually have a single cause (for example, mesothelioma and asbestos; diethylstilbestrol and adenocarcinoma) (Farber 1987). Lacking any direct evidence of specific causation, courts generally adjudicate specific causation based on a differential diagnosis by an expert or based on epidemiologic evidence showing a relative risk greater than 2.0, which suggests that any individual plaintiff's disease was more likely than not caused by the exposure under study (Carruth and Goldstein 2001). Such methods are very imprecise and prone to under- and overcompensation depending on the facts of the case. A toxicogenomic marker could provide direct evidence of specific causation if an individual who develops a disease is shown to have the specific molecular signature of toxicity attributable to a specific agent.

Duty of Care

Toxic tort litigation involves judgments about the duty of care by an actor who creates risks to those who may be injured by those risks. Toxicogenomic data could affect or shift those judgments about duty of care in a number of con-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

texts. For example, the finding that some members of the population may have a genetic susceptibility that makes them particularly sensitive to a product may impose new duties on the product manufacturer with regard to testing, labeling, and selling that product. Some individuals who alleged they had been harmed by the Lyme disease vaccine Lymerix brought lawsuits claiming that the vaccine manufacturer had a legal duty to warn vaccine users to obtain a genetic test for a polymorphism that allegedly affected the user's propensity to develop serious side effects from the vaccine (Noble 2000). The litigation settled before a judgment was issued but likely contributed at least indirectly to the vaccine being removed from the market and was the first of what will likely be a new trend of plaintiffs claiming that a product manufacturer has additional duties to protect individuals genetically susceptible to their products. Alternatively, under the "idiosyncratic defense doctrine," defendants may be able to argue in some cases that they have no duty to protect individuals with a rare genetic susceptibility to a product. Under this line of cases, courts have held that a product manufacturer can be reasonably expected to ensure the safety of "normal" members of the population and not individuals with an unusual susceptibility to the agent in question (Cavallo v. Star Enterprise 1996; Marchant 2000).

The detection of toxicogenomic changes in exposed individuals using toxicogenomic technologies may also result in lawsuits seeking damages for an increased risk of disease or for medical monitoring. Historically, courts have been reluctant to award damages for an increased risk of disease that has not yet manifested in clinical symptoms, but courts in some states have recognized such a claim if the at-risk individual can sufficiently quantify a substantial increased risk of disease (Klein 1999). It is possible that toxicogenomic data could support such a claim (Marchant 2000). A related type of claim is for medical monitoring, now recognized by more than 20 states, which requires a defendant responsible for a risk-creating activity to pay for periodic medical testing of exposed, at-risk individuals (Garner et al. 2000). Toxicogenomic assays could be used to identify at-risk individuals who would be entitled to ongoing medical evaluations, or the assays could serve as a periodic medical test for people who incurred a hazardous exposure.

Damages

Toxicogenomic data may also be relevant in the damages stage of toxic tort litigation. A defendant who is liable for a plaintiff's injuries may seek to undertake genetic testing of the plaintiff to identify potential predispositions to disease that might have contributed to his or her disease or that might otherwise reduce the plaintiff's life expectancy (Rothstein 1996). Either of these findings could reduce the damages the defendant would be ordered to pay to the plaintiff. Although the plaintiff's genetic predispositions to disease may sometimes be pertinent (and may be used to benefit the plaintiff in some situations), there is a danger that defendants will undertake "fishing expeditions" in the plaintiff's

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

genome that may reveal sensitive and private personal information and potentially violate the plaintiff's right not to know (Chadwick 2004).

Legal, Policy, and Ethical Aspects of Toxic Tort Applications

The many potential applications of toxicogenomic data in toxic tort litigation raise a number of scientific, legal, policy, and ethical issues. One central concern is the potential for premature use of toxicogenomic data (Marchant 2000). Unlike regulatory agencies, which generally consider new types of data cautiously and deliberately, toxic tort litigants are unlikely to show similar restraint. A toxic tort case is a one-time event often involving large stakes that, once filed, tends to move forward expeditiously toward a decision under a court-ordered schedule. Therefore, toxic tort litigants have every incentive to use any available data that may help them prevail, regardless of how well those data have been considered and validated by the scientific community. Premature use of toxicogenomic data should obviously be discouraged, but it is important to note that current scientific evidence on issues in litigation such as causation are often inadequate, and toxicogenomic data have enormous potential to make the resolution of toxic tort litigation more scientifically informed, consistent, and fair.

The reliability of toxicogenomic data introduced in a court proceeding will be evaluated under the standards for admissibility of scientific evidence. In federal courts and many state courts, the evidence will be evaluated under the criteria for scientific evidence announced by the U.S. Supreme Court in the 1993 Daubert decision (Daubert v. Merrell Dow 1993). Under that standard, the trial judge is to serve as a gatekeeper for scientific evidence to ensure that any such evidence presented to a jury is both relevant and reliable. The Supreme Court identified the following four nonexclusive criteria that courts can use to evaluate the reliability of scientific evidence: (1) whether the evidence can and has been empirically tested, (2) whether it has a known rate of error, (3) whether it has been peer reviewed and published, and (4) whether it is generally accepted within the relevant scientific field. These criteria are consistent with general criteria used to validate toxicologic tests and, if applied rigorously, should help ensure against the premature or inappropriate use of toxicogenomic data in toxic tort litigation.

Specifically, a court might want to consider the following factors in deciding whether to admit toxicogenomic data in a toxic tort lawsuit.

- Has the toxicogenomic response been shown to be associated with or predictive of a traditional toxicologic end point (for example, cancer, toxicity)?
- Are there data showing that the observed toxicogenomic response is characteristic of exposure to the specific toxic agent at issue, with a similar time course of exposure as experienced by the plaintiff?

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

- Have the data used or relied on for making the above determinations been published in peer-reviewed scientific journals?
- Have one or more other laboratories replicated the same or similar results under similar conditions?
- Has the toxicogenomic platform used been shown to provide consistent results to the platform used in any other studies relied on?

Ideally, all these questions would be answered affirmatively before toxicogenomic data were introduced into evidence. Given that it is not reasonable to impose greater barriers to the introduction of toxicogenomic data than other types of toxicologic evidence used in toxic tort litigation, data satisfying most of the above criteria would likely be sufficiently reliable to be admitted.

The use of toxicogenomic data in toxic tort litigation raises a series of other issues. For instance, the capability of lay jurors to adequately understand and apply toxicogenomic data in their decision making will present challenges. The potential privacy and confidentiality issues raised by genetic testing of plaintiffs create another set of concerns. Protective orders that protect sensitive information used at trials from being publicly disclosed will be needed to help protect the confidentiality of personal genetic information.

Product manufacturers may have concerns about future liabilities associated with toxicogenomic biomarkers. For example, a manufacturer may sponsor a study that detects a biomarker of unknown toxicologic significance today but that years later may become established as a validated biomarker of toxicity. The use of such data to impose liability retroactively could raise fairness concerns and may deter manufacturers from conducting toxicogenomic studies. At the same time, shielding manufacturers from such liability could create the wrong incentives by deterring them from investigating the risks of their products and taking appropriate mitigation measures and may deprive injured product users from being compensated for their injuries. These countervailing factors demonstrate the complex and sensitive role that potential liability can exert on scientific research and applications of toxicogenomics.

COMMUNICATION AND EDUCATION

The increasing scientific capability of making individualized predictions of risk from toxic substances raises the important question of how this information will be communicated to individuals and the public. Such predictions, tailored for the individual, may be years away. But it is still important to communicate with the public in advance about relevant scientific advancements to maintain public support for the science and to help educate members of the public for the day when many of them may have to make personal decisions based on their capacity to understand the customized predictions of the risk they face from one or more toxic hazards (NRC 2005c). A great deal of research is needed in this field. Among other things, we do not know how much information people

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

want, in what form, their likely comprehension, their likely response, and the degree of variability in different socioeconomic and cultural groups. Nevertheless, some basic risk-communication principles are instructive.

Risk communication has been defined as "an interactive process of exchange of information and opinion among individuals, groups, and institutions" (NRC 1989). Thus, the prescription for agencies, media, and others involved in risk communication is to abandon the traditional top-down, sender-based, "public education" model of risk communication (for example, launching a flurry of messages in an attempt to get the public to see things the way experts do). Instead, they should favor an approach that emphasizes greater understanding of the emotional reactions, concerns, and motivations of a segmented public who can seek (or even avoid), process, and evaluate critically the risk information they encounter and who have varying desires and capacities to do so (Griffin et al. 1999). Indeed, "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" has become the prime definition of "health literacy" and has been proposed as an essential component of health and risk communication programs (IOM 2004, p. 32).

Thus, there are at least four key issues in risk communication: (1) sufficiency of information; (2) capacity of the individual or society to access, assess, and understand information; (3) emotional responses to risks; and (4) trust in scientific and mass media organizations that oversee communication channels (NRC 2005d).

It is important to understand that effective risk communication requires multiple messages tailored to a particular audience. People interpret health risks in light of their everyday events and experiences, and the messages must be framed within a familiar context. Health disparities and environmental justice issues are of heightened concern in some communities. In all communities, a person's emotional response to risk, such as worry, fear, anger, and hope, can determine risk perception as well as response to risk (e.g., Witte 1994; Griffin et al. 1999). The emotional response to risk is also a key factor in risk acceptability. Better understood and familiar risks may be more acceptable than dreaded, poorly understood risks of a lower magnitude.

Complicating any discussion of risk communication is the public's low level of scientific comprehension, low level of understanding probabilistic information, and low level of understanding basic numerical concepts. Thus, risk communication, especially about predictions of individualized risk from toxic substances, must include upfront efforts designed to help the public improve their health and risk literacy.

There are many dimensions to the translation of toxicogenomic and pharmacogenetic knowledge into health benefits. Research needs to focus on both individual-level and population-level educational and risk communication aspects of translation into medical and public health practice (Kardia and Wang 2005). For example, critical research issues on an individualized level are exemplified by research on genetic risk communication (Hopwood et al. 2003), in-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

formed consent processes (Geller et al. 1997), the decision-making process (Shiloh 1996), and provider knowledge and awareness of genetics (Suchard et al. 1999). These research arenas have direct practice implications for genetic education and genetic counseling (C. Wang et al. 2004), interventions incorporating genetic information (Lerman et al. 1997), patient adherence to screening recommendations (Hadley et al. 2004), effectiveness of decision supports and aids (Green et al. 2004), and health care provider training (Burke et al. 2002).

Researchers and practitioners in the field of health behavior and health education can play a pivotal role in integrating toxicogenomics into practice to improve the public's health (e.g., Sorenson and Cheuvront 1993; C. Wang et al. 2005). Priority areas that are ripe for further exploration, understanding, and application include the following: (1) public and provider education about genetic information, (2) risk communication and interventions for behavior change, (3) sociologic sequelae of genetic testing, and (4) public health assurance and advocacy. An ecologic perspective should be considered when addressing the educational and communication issues involved in applying toxicogenomics to decrease health risks. Many different types of stakeholders and practitioners should be considered, and, consequently, many levels of intervention and analysis should be pursued.

CONCLUSIONS AND RECOMMENDATIONS

Because this chapter presents a number of complex issues, the conclusion text precedes the corresponding, numbered recommendations.

It is critical to ensure adequate protections on the privacy, confidentiality, and security of toxicogenomic information in health records. Safeguarding this information will further important individual and societal interests. It will also prevent individuals from being dissuaded from participating in research or undergoing the genetic testing that is the first step in individualized risk assessment and risk reduction. The potential consequences of disclosure of toxicogenomic information are greater with the growth of electronic health records.

There is a lack of comprehensive legislation protecting the privacy, confidentiality, and security of health information, including genetic information. These protections are needed at all entities that generate, compile, store, transmit, or use health information, not just those affected by HIPAA.

1. Address the privacy, confidentiality, and security issues that affect the use of toxicogenomic data and the collection of data and samples needed for toxicogenomic research.
2. Role-based access restrictions should be used for the disclosure of health information in health care settings. Contextual access criteria should be developed and used for the disclosure of health information, pursuant to an authorization, beyond health care settings.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Toxicogenomic research often uses large biorepositories and databases in anonymous, deidentified, linked, or identifiable forms as well as phenotypic data in health records. Inconsistencies between the Common Rule informed consent requirement and the HIPAA Privacy Rule authorization requirement burden and interfere with toxicogenomic research.

3. Consider approaches to harmonize standards for deidentification and informed consent and authorization under the Federal Rule for the Protection of Research Subjects (Common Rule) and the HIPAA Privacy Rule, to minimize unnecessary barriers to research while continuing to protect the privacy and welfare of human subjects.

4. DHHS should explore new approaches to facilitate large-scale biorepository and database research while protecting the welfare and privacy of human subjects.

Subpopulation groups considered socially vulnerable based on race or ethnicity, income, age, or other factors are at increased risk for discrimination, stigma, and other adverse treatment as a result of individualized toxicogenomic information.

5. In toxicogenomic research, especially involving or affecting socially vulnerable populations, special efforts should be made at community engagement and consultation about the nature, methods, and consequences of the research.

6. To minimize the risk of adverse impacts on socially vulnerable populations from toxicogenomic research and implementation, access to adequate health care for diagnostic and treatment purposes will be critical and should be a priority for funding agencies and legislators.

The decision to use toxicogenomic testing to learn about one's individual risk should rest with the individual, including risk posed by the workplace setting. Employers have the primary responsibility, under the Occupational Safety and Health Act, to provide a safe and healthful workplace and, under the Americans with Disabilities Act, to provide nondiscriminatory employment opportunities and reasonable accommodations for individuals with disabilities.

7. When toxicogenomic tests to provide individualized risk information have been validated, individuals should be able to learn of their particular risk from workplace exposures without the information being disclosed to their current or potential employer. Upon learning of their individualized risk information they should be able to decide for themselves whether to accept the risk of such employment.

8. Only in extraordinary circumstances, when employing an individual with a risk would create a direct and immediate threat to the individual, cowork-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

ers, or the public, would the employer be justified in excluding an individual from a particular employment based on increased risk.

Toxicogenomic data have many promising applications in regulation and litigation. These data have the potential to help fill many of the scientific uncertainties and gaps regarding exposure, causation, dose response, and extrapolation that currently limit the toxicologic knowledge critical for making sound regulatory and liability decisions.

9. Although caution, scrutiny, and validation are required to protect against premature, inappropriate, and unethical use of toxicogenomic data in regulatory and litigation contexts, care should also be taken to ensure that a higher standard of proof is not imposed for toxicogenomic data relative to other types of toxicologic data used in regulation and litigation.

10. A regulatory agency or court should give appropriate weight to the following factors in deciding whether to rely on toxicogenomic data:[1]

a. Has the toxicogenomic response been shown to be associated with or predictive of an adverse health outcome (for example, cancer, toxicity)?
b. Has the specificity and sensitivity of the test been established to be within reasonable bounds?
c. Are there data showing that the observed toxicogenomic response is characteristic of exposure to the specific toxic agent at issue, with a similar time course and level of exposure as experienced by the plaintiff?
d. Have the data used or relied on for making the above determinations been published in peer-reviewed scientific journals?
e. Have one or more other laboratories replicated the same or similar results under similar conditions?
f. Has the toxicogenomic platform used been shown to provide results consistent with the platforms used in any other studies that were relied on?

Risk communication is an essential component of translating toxicogenomic information into reduced health risks for the public. Currently, the general public, as well as health care practitioners, are ill-equipped to understand and use toxicogenomic information to alter adverse health outcomes.

11. Research is needed on how to communicate toxicogenomic risk information to the public using culturally and psychologically appropriate methods.

---

[1]These factors could be operationalized through agency risk assessment guidelines in the regulatory context and through judicial precedent relying on authoritative sources, such as this report in the litigation context.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

12. Educational initiatives are needed for vulnerable subgroups and the general public to raise awareness about toxicogenomic findings that can affect their health.
13. Educational initiatives need to be developed and implemented to prepare the medical and public health workforce to use toxicogenomic information.

Finally, several areas of future research would address some of the issues raised in this chapter. These include the following:

14. Federal agencies should increase their support for research on the issues of ethical, legal, and social implications in applying toxicogenomic technologies, including public attitudes toward individualized risk, social effects of personalized information on increased risks, and regulatory criteria for toxicogenomics.
15. The National Institute of Environmental Health Sciences (or other federal agencies) should develop "points-to-consider" documents that identify and discuss the issues of ethical, legal, and social implications relevant to individual researchers, institutional review boards, research institutes, companies, and funding agencies participating in toxicogenomic research and applications.

Copyright © National Academy of Sciences. All rights reserved.

12

Conclusions and Recommendations

Toxicogenomic technologies provide new means to evaluate complex biologic systems and the impact of chemicals on living systems. Specifically, toxicogenomic technologies may be applied to improve cross-species extrapolation in the analysis of chemical hazard, identify susceptible subpopulations, assess effects of early life exposures to chemicals, analyze compounds' modes of action, screen for potential toxic responses, refine exposure assessment, and analyze biologic effects of combined exposures or mixtures. Applying toxicogenomic technologies to these important problems in toxicology can improve understanding and minimize adverse effects of environmental exposures and drugs and contribute to a knowledge base of toxicity end points.

To date, applications of toxicogenomic technologies in risk assessment and the regulatory decision-making process have been exploratory. Although they clearly have great potential to affect decision making, toxicogenomic technologies are not ready to replace existing required testing regimes in risk assessment and regulatory toxicology. However, toxicogenomic technologies are assuming an increasing role as adjuncts to, and extensions of, existing technologies for predictive toxicology. Toxicogenomics can provide molecular level information and tests that add to the "weight of the evidence" for or against the safety of specific environmental toxicants and drugs. Ultimately, toxicogenomic technologies are envisioned to be more sensitive and more informative than existing technologies and may supplant some approaches currently in use, or at least be a component of batteries that will replace certain tests.

This chapter summarizes the committee's conclusions and recommendations. The summary includes several areas of important overarching recommendations that are not discussed elsewhere in the report (recommendations for a new initiative, for general risk assessment, and to address educational needs) as well as recommendations discussed in detail at the end of individual chapters.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Recommendation 1: Regulatory agencies should enhance efforts to incorporate toxicogenomic data into risk assessment.

The following actions are needed to move toward this objective:

a. Substantially enhance agency capability to integrate toxicogenomic approaches into risk assessment practice, focusing on exposure assessment, hazard screening, identification and understanding of variation in human susceptibility, mechanistic insight, assessment of dose-response relationships, cross-species extrapolation, and assessment of mixtures.

b. Invest in research and expertise within the infrastructure of regulatory agencies as well as active collaboration across agencies.

c. Develop and expand research programs dedicated to integrating toxicogenomics into challenging risk assessment problems, including the development of public and private sector partnerships.

A NEW HUMAN TOXICOGENOMICS INITIATIVE

Fully integrating toxicogenomic technologies into predictive toxicology will require a coordinated effort analogous in concept to the Human Genome Project. Such an effort would, of necessity, be multidisciplinary and multi-institutional and would require broad standardization of technologies and sharing of information. To achieve its goals, such an effort would require funding and resources significantly greater than what are currently allocated to existing research programs and would require partnerships between the public and private sectors.

A national, publicly available database is an essential element of this effort. This is because toxicogenomic technologies generate enormous amounts of data—on a scale even larger than sequencing efforts like the Human Genome Project. Current public databases are inadequate both to manage the types and volumes of data expected to be generated by large-scale applications of toxicogenomic technologies and to facilitate mining and interpretation of the data, which are just as important as their generation and storage. Finally, the ethical, legal, and social implications of collecting, sharing, and using toxicogenomic information are likely to be substantial.

The project as envisioned would require a large-scale, coordinated effort, involving partnerships of government agencies, academic institutions, and commercial organizations. One essential element is to leverage, where possible, large publicly funded studies so they can be used to generate toxicogenomic data. Another essential element is to take steps to facilitate the production and sharing of data between the public and private sectors. Because many of the most extensive applications of toxicogenomic technologies have occurred in the private sector, public-private partnerships will be a vital component of such a large coordinated effort.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

This resource would strengthen the utility of toxicogenomic technologies in toxicity assessment and enable more accurate prediction of health risks associated with existing and newly developed compounds and formulations. For example, the Human Toxicogenomics Initiative (HTGI) data resource could provide context to toxicogenomic data generated by drug and chemical manufacturers. Regulatory agencies could use the data resource in risk assessment to inform exposure limits, and the data could improve understanding of the effects of a broad range of exposures ranging from pollution to natural disasters to terrorist attacks.

Recommendation 2: NIEHS should cooperate with other stakeholders in exploring the feasibility and objectives of implementing a Human Toxicogenomics Initiative dedicated to advancing toxicogenomics. The HTGI is intended to coordinate efforts by the government, academia, and industry to advance the evaluation and, as appropriate, implementation of toxicogenomic technologies.

Elements of the HTGI need to include the following:

a. Creation and management of a large, public database to store and integrate the results of toxicogenomic analyses with conventional toxicity-testing data (see Chapter 10 for more specifics on database needs and recommendations).

b. Assembly of toxicogenomic and conventional toxicologic data on a large number (hundreds) of compounds into the single database. This includes generating new toxicogenomic data from humans and animals for a number of compounds on which other types of data already exist and consolidating existing data. Every effort should be made to leverage existing research studies and infrastructure (such as those of the National Toxicology Program) to collect samples and data that can be used for toxicogenomic analyses.

c. Creation of a centralized, national biorepository for human clinical and epidemiologic samples, building on existing efforts.

d. Further development of bioinformatic tools (for example, software, analysis, and statistical tools).

e. Explicit consideration of the ethical, legal, and social implications of collecting and using toxicogenomic data and samples.

f. Specific coordinated subinitiatives to evaluate the application of toxicogenomic technologies to the assessment of risks associated with chemical exposures.

DATA ANALYSIS TOOLS

There is a need for bioinformatic, statistical, and computational approaches and software to analyze data. The HTGI effort described above in-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

cludes development of specialized bioinformatic and computational tools, including development of appropriate statistical methodologies, to analyze toxicogenomic data.

Recommendation 3: Develop specialized bioinformatic, statistical, and computational tools and approaches to analyze toxicogenomic data.

a. Develop algorithms that facilitate accurate identification of orthologous genes and proteins in species used in toxicologic research.
b. Develop tools to integrate data across multiple analytical platforms (for example, gene sequences, transcriptomics, proteomics, and metabolomics).
c. Develop computational models to enable the study of network responses and systems-level analyses of toxic responses.

EXPOSURE ASSESSMENT

The application of toxicogenomics to define biomarkers of exposure requires consensus on what constitutes an exposure biomarker and standardization of toxicogenomic platforms that are appropriate for identifying signatures of environmental or drug exposures in target and surrogate tissues and fluids. Additional challenges include the individual variation in response to an environmental exposure and the persistence of a toxicogenomic signature after exposure.

Recommendation 4: Toxicogenomic technologies should be adapted and applied for the study of exposure assessment by developing signatures of exposure to individual chemicals and perhaps to chemical mixtures.

The following immediate steps could be taken toward this goal:

a. Using transcriptomic, proteomic, and metabolomic technologies to identify signatures of environmental exposures in target and surrogate tissues and fluids, primarily with animal models.
b. Beginning to test complex mixtures for possible identification of distinct exposure signatures.
c. Examining the time course of persistence of chemical versus toxicogenomic signatures after initial chemical exposures.

The following intermediate steps could be taken toward this goal:

d. Including toxicogenomic (transcriptomic, metabolomic, and proteomic) analysis of samples in large human population studies and in studies designed to

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

assess exposures at toxicant levels commonly encountered in the workplace and in certain communities.[1]

e. Including toxicogenomic analysis of samples in relevant case-control, cohort, and panel studies that involve repeated measurements over time as well as in clinical trials when possible and appropriate.

f. Using the information collected from studies to help develop and populate a database that will support further development of toxicogenomic exposure assessment.

HAZARD SCREENING

Toxicogenomic technologies provide new and potentially useful specific end points for use in toxicity screening. In contrast to applications in evaluating new drug candidates, screening approaches for environmental chemicals will need to address a broader range of exposures and a more comprehensive set of end points. A database of signatures that are informative of the appropriate range of phenotypic end points and doses is essential. Much of the toxicogenomic data collected reside in publicly inaccessible databases. The process of deriving useful toxicogenomic signatures for screening would be accelerated if these data were publicly available.

Recommendation 5: Upon validation and development of adequate databases, integrate toxicogenomic screening methods into relevant current and future chemical regulatory and safety programs.

The following steps could be taken toward this goal:

a. Having regulatory agencies (including the Environmental Protection Agency and the Food and Drug Administration) continue to develop and refine guidance documents for their staff on interpreting toxicogenomic data. In particular, guidance for environmental chemicals must ensure that screening protocols address the types of end points most relevant for the general population, including sensitive subpopulations.

b. Developing mechanisms to improve the quantity and quality of data available for deriving screening profiles and developing a database to organize this information:

- i. Establishing a dialogue with entities holding currently inaccessible toxicogenomic data to evaluate options for increasing data availability.
- ii. Integrating toxicogenomic assays into ongoing initiatives such as the National Institutes of Health Molecular Libraries Initia-

[1]See issues raised about the protection of humans in Chapter 11.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

tive, the National Toxicology Program, and other large chemical screening programs.

iii. Establishing a dialogue among regulators, regulated industries, and other relevant stakeholders to address current disincentives to generating and submitting toxicogenomic data in regulatory settings.

iv. Convening an expert panel to provide recommendations for which model compounds, laboratory platforms, and specific data elements are necessary for building toxicogenomic databases that are useful for screening applications. Assessment of in vitro approaches for specific toxic end points should be emphasized. All processes, toxicogenomic data, and outcome data must be publicly accessible.

c. Developing databases and algorithms for using proteomic and metabonomic data in screening.

d. Ensuring that the regulatory framework provides incentives, or at least removes disincentives, for premarket testing of chemicals.

VARIABILITY IN SUSCEPTIBLITY

People vary in their susceptibility to the toxic effects of chemical exposures, and information about genetic variability is generally used retrospectively (that is, after safety problems are discovered). Toxicogenomic technologies provide the opportunity to use genetic information in a prospective fashion to identify susceptible subpopulations and to assess the distribution of differences in susceptibility in larger populations. Toxicogenomic technologies could reduce the uncertainty about assumptions used in regulatory processes to address population variability.

Recommendation 6: Use toxicogenomic information to prospectively identify, understand the mechanisms of, and characterize the extent of genetic and epigenetic influences on variations in human susceptibility to the toxic effects of chemicals, with the goal of improving the certainty about assumptions used in the regulatory processes to address population variability.

The following immediate step could be taken toward this goal:

a. Using animal models to identify and study genes associated with human variation in toxicity.

The following intermediate steps could be taken toward this goal:

b. Using genome-wide studies, ranging from anonymous dense single nucleotide polymorphism (SNP) scans to specialized arrays of putative functional

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

SNP approaches, of individuals in existing cohort, clinical trial, and other population studies to identify gene variations that influence sensitivity to potentially toxic agents. This will be facilitated by efficient quantitative and qualitative assessment of individual exposures to multiple compounds, which may be aided by toxicogenomics (see exposure assessment section).

c. Focusing more attention on investigating context-dependent genetic effects (that is, gene-gene interactions as well as interactions with other biologic contexts such as developmental age, sex, and life course factors) that reflect the state of the biologic networks underlying responses to toxicologic agents. For example, animal models can be used to better understand polygenic effects.

The following long-term steps could be taken toward this goal:

d. Conducting research on the influence of exposure, genetic differences, and their interaction on epigenetic modification of the human genome.

e. Better characterizing the influence of epigenetic modifications on disease processes that are associated with exposure to toxicologic agents.

f. Developing an animal model resource that mimics the genetic heterogeneity of human populations to study the distribution of gene-gene interactions and gene-epigenetic interactions and that can serve as a model for understanding population risk.

MECHANISTIC INFORMATION

Toxicogenomic studies are improving our knowledge of the underlying biology and the regulatory networks that integrate the signaling cascades involved in toxicity and thus may advance the introduction of mechanistic insight into risk assessment and fulfill the promise of more accurate and expedited elucidation of class-related biologic effects.

An immediate need in the field of toxicogenomics is for more accurate identification of orthologous genes or proteins across species. (There is also a need for cross-species extrapolation, as discussed below.) There are pressing needs to develop algorithms that combine and interpret multiple types of data (for example, gene expression, proteomic, and metabonomic data); better approaches to probe the complexity of toxic responses are also needed. Finally, additional comprehensive time- and dose-related investigations and the study of exposure paradigms that reproduce the human condition with fidelity are needed.

Recommendation 7: Steps should be taken and tools developed to continue advancing the ability of toxicogenomics to provide useful mechanistic insight.

The following immediate steps could be taken toward this goal:

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

a. Develop richer knowledge bases and models that can integrate knowledge of the mechanisms of toxicity and the complex network information, encouraging the community to use these models to study toxicology as a global response. This requires a new paradigm of data management, integration, and computational modeling and will require the development of algorithms that combine and interpret data across multiple platforms.

b. Encourage detailed mechanistic research that is useful for classifying toxic chemicals and to assess the public health relevance of these toxicity classifications.

c. Facilitate the identification of orthologous genes and proteins in different laboratory species and the development of algorithms for comparing different species.

The following intermediate steps could be taken toward this goal:

d. Advance proteomic and metabolomic analyses by promoting the integration of peptide and metabolite separation technologies into toxicologic investigations and advancing proteomic and metabolomic databases.

e. Implement educational programs to help the toxicology and risk assessment communities incorporate data-rich mechanistic information into their professional practice.

The following long-term step could be taken toward this goal:

f. When appropriate, encourage a shift in the focus of mechanistic investigations from single genes to more integrated analyses that embrace the complexity of biologic systems as a whole as well as the multidimensionality of dose- and time-related effects of toxic agents.

DOSE-RESPONSE RELATIONSHIPS

Toxicogenomics has the potential to improve understanding of dose-response relationships, particularly at low doses. Collecting information on dose-response relationships for a range of doses appropriately linked to time will be essential to fully integrate toxicogenomics into risk assessment decision making. To effectively address questions about risks associated with human exposures to environmental chemicals, which may be much lower than doses currently used in toxicology studies, attention must focus on characterizing toxicogenomic responses at low doses. Such efforts will be more valuable when toxicogenomic studies are tied to conventional toxicity responses, such as incorporating toxicogenomics into traditional toxicity-testing programs.

Recommendation 8: Future toxicologic assessment should incorporate dose-response and time-course analyses appropriate to risk assessment. An

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

analysis of known toxic compounds that are well characterized would provide an intellectual framework for future studies.

CROSS-SPECIES EXTRAPOLATION

Toxicogenomic technologies offer the potential to significantly enhance the confidence in animal-to-human toxicity extrapolations that constitute the foundation of risk evaluations. Using toxicogenomics to analyze species differences in toxicity will help explain the molecular basis for the differences, improving the translation of animal observations into credible estimates of potential human risk. In addition, by providing comparisons between humans and other species at the molecular level, toxicogenomics may assist in identifying those animal species and strains that are most relevant for specific assays. Identifying the most relevant strains is clearly advantageous for a number of reasons, and the potential exists to decrease the number of tests required to assess toxicity.

Recommendation 9: Continue to use toxicogenomics to study differences in toxicant responses between animal models and humans and continue to use genotyped and genetically altered animal model strains as experimental tools to better extrapolate results from animal tests to human health. Algorithms must be developed to facilitate accurate identification of genes and proteins that serve the same function in different organisms and species—orthologous genes and proteins—used in toxicologic research.

DEVELOPMENTAL EXPOSURES

Although recognized to be important in a number of disorders, relatively little is known about the health impacts of fetal and early-life exposures to many chemicals in current use. Because of their sensitivity, toxicogenomic technologies are expected to reveal more than previously was possible about the molecules involved in development and the critical molecular level events that can be perturbed by toxicants. Toxicogenomics may also enable screening for chemicals that cause gene expression changes associated with adverse developmental effects.

Recommendation 10: Use toxicogenomics to investigate how exposure during early development conveys susceptibility to drug and chemical toxicities.

MIXTURES

Although much toxicology focuses on the study of single chemicals, humans are frequently exposed to multiple chemicals. It is difficult to decipher

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

how exposure to many chemicals will influence the effects of each one. It is unlikely that toxicogenomic signatures will be able to decipher all interactions among complex mixtures, but it should be possible to use mechanism-of-action data to design informative toxicogenomic experiments, including screening chemicals for potential points of biologic conversion (overlap) such as shared activation and detoxification pathways, enhancing identification and exploration of potential interactions, and moving beyond empirical experiments.

Recommendation 11: Use toxicogenomic approaches to test the validity of methods for estimating potential risks associated with mixtures of environmental chemicals.

VALIDATION

Validation is a key step in moving toxicogenomic technologies from the laboratory to real-world applications. Although the community has learned that careful experiments using genomic approaches can provide results that are comparable among laboratories and that provide insight into the biology of the system under study, the need for standards to assess the quality of particular experiments remains, as do other levels of validation discussed in Chapter 9.

Recommendation 12: Actions should be taken to facilitate the technical and regulatory validation of toxicogenomics.

The following specific steps facilitate validation:

a. Developing objective standards for assessing sample and data quality from different technology platforms, including the development of standardized materials such as those developed by the External RNA Control Consortium.

b. Developing appropriate criteria for using toxicogenomic technologies for different applications, such as hazard screening and exposure assessment.

c. Having regulatory agencies establish clear, transparent, and flexible criteria for the regulatory validation of toxicogenomic technologies.

ETHICAL, LEGAL, AND SOCIAL ISSUES

As toxicogenomic data with clinical and epidemiologic annotation are amassed, it is critical to ensure adequate protections on the privacy, confidentiality, and security of toxicogenomic information in health records. Safeguarding this information will further important individual and societal interests. It will also prevent individuals from being dissuaded from participating in research or undergoing the genetic testing that is the first step in individualized risk assessment and risk reduction. The potential consequences of disclosure of toxicogenomic information are greater with the growth of electronic health records.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

The decision to learn one's toxicogenomic risk should rest with the individual, including risks at the workplace. Employers have the responsibility to provide a safe and healthful workplace and to provide nondiscriminatory employment and reasonable accommodations for individuals with disabilities. Additionally, risk communication is an essential component of translating toxicogenomic information into reduced health risks for the public.

Toxicogenomics is also likely to play a role in occupational, environmental, and pharmaceutical regulation and litigation. Regulatory agencies and courts should give appropriate weight to validation, replication, consistency, sensitivity, and specificity when deciding whether to rely on toxicogenomic data.

Recommendation 13: Address the ethical, legal, and social issues that affect the use of toxicogenomic data and the collection of data and samples needed for toxicogenomic research.

The following are important areas to address and specific steps that could be taken (see Chapter 11):

a. The lack of comprehensive legislation protecting the privacy, confidentiality, and security of health information, including genetic information—especially information relevant to vulnerable populations. This could be addressed by legislative improvements to enhance individual protection and to minimize unnecessary barriers to research while continuing to protect the privacy and welfare of human subjects.

b. The Department of Health and Human Services should explore new approaches to facilitate large-scale biorepository and database research while protecting the welfare and privacy of human subjects.

c. Regulatory agencies and courts should give appropriate weight to a number of important factors in deciding to rely on toxicogenomic data and be careful not to apply more stringent evidentiary standards for toxicogenomic data than other types of toxicologic evidence. Factors a regulatory agency or court should consider when deciding whether to rely on toxicogenomic data include validation, replication, consistency, sensitivity, and specificity, as described in Chapter 11.

d. In toxicogenomic research, especially involving or affecting socially vulnerable populations, special efforts should be directed toward community engagement and consultation about the nature, methods, and consequences of the research. To minimize the risk of adverse impacts on socially vulnerable populations from toxicogenomic research and implementation, access to adequate health care for diagnostic and treatment purposes will be critical and should be a priority for funding agencies and legislators.

e. The appropriate federal agencies should sponsor or facilitate research on ethical, legal, social, and communication issues in applying toxicogenomic technologies, including public attitudes about risk, social impacts of personal-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

ized genetic information about increased risk, the most effective methods of informed consent and data sharing, and what needs to be communicated.

EDUCATION AND TRAINING IN TOXICOGENOMICS

Given the complexity of toxicogenomics, the generation, analysis, and interpretation of toxicogenomic information represents a challenge even in the scientific community and requires the collaborative interdisciplinary efforts of scientific teams of specialists. Therefore, it is essential that education and training in toxicogenomics become a continuous process that reflects the rapid developments in these new technologies. There is a need to develop education and training programs for health professionals, regulators, attorneys and judges, persons communicating to the public, and scientists in training.

Recommendation 14: Develop education and training programs relevant to toxicogenomic applications to predictive toxicology.

The following specific steps meet this recommendation:

a. Conduct educational initiatives to raise awareness of the general public, vulnerable subgroups, and health professionals about toxicogenomic findings that can affect health.

b. Establish a training program for regulators, attorneys, and judges to ensure a basic understanding of the generation and interpretation of toxicogenomic datasets, as applied in regulatory decision making.

c. For media and experts in communication, provide training that may include short courses on what types of toxicogenomic information will be helpful for the public to understand and how to explain the technical information in an understandable way.

d. For new scientists, it may be appropriate to develop programs at the master's and Ph.D. levels that include bioinformatic and toxicogenomic applications in toxicology curricula.

e. For scientists not specializing in toxicogenomics—such as epidemiologists, environmental scientists, and physicians—and for institutions, include didactic instruction in degree programs and curricula, with the goal of educating them on the principles and practice of toxicogenomics.

f. Appropriate federal agencies should develop "points to consider" that identify and discuss ethical, legal, and social issues relevant to individual researchers, institutional review boards, research institutes, companies, and funding agencies participating in toxicogenomic research and applications.

Copyright © National Academy of Sciences. All rights reserved.

References

Aebersold, R., and M. Mann. 2003. Mass spectrometry-based proteomics. Nature 422(6928):198-207.

Ahn, N.S., H. Hu, J.S. Park, J.S. Park, J.S. Kim, S. An, G. Kong, O.I. Aruoma, Y.S. Lee, and K.S. Kang. 2005. Molecular mechanisms of the 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced inverted U-shaped dose responsiveness in anchorage independent growth and cell proliferation of human breast epithelial cells with stem cell characteristics. Mutat. Res. 579(1-2):189-199.

AHS (Agricultural Health Study). 2006. Agricultural Health Study [online]. Available: http://www.aghealth.org/ [accessed Aug. 24, 2006].

Akutsu, T., S. Miyano, and S. Kuhara. 2000. Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function. J. Comput. Biol. 7(3-4):331-343.

Alban, A., S.O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, and I. Currie. 2003. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics 3(1):36-44.

Alexander, J., R. Reistad, S. Hegstad, H. Frandsen, K. Ingebrigtsen, J.E. Paulsen, and G. Becher. 2002. Biomarkers of exposure to heterocyclic amines: Approaches to improve the exposure assessment. Food Chem. Toxicol. 40(8):1131-1137.

Allen, A.L. 1997. Genetic privacy: Emerging concepts and values. Pp. 31-59 in Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era, M.A. Rothstein, ed. New Haven: Yale University Press.

Allen, E.E., J.S. Fetrow, L.W. Daniel, S.J. Thomas, and D.J. John. 2006. Algebraic dependency models of protein signal transduction networks from time-series data. J. Theor. Biol. 238(2):317-330.

Ames, R.G., S.K. Brown, D.C. Mengle, E. Kahn, J.W. Stratton, and R.J. Jackson. 1989. Cholinesterase activity depression among California agricultural pesticide applicators. Am. J. Ind. Med. 15(2):143-150.

Amin, R.P., A.E. Vickers, F. Sistare, K.L. Thompson, R.J. Roman, M. Lawton, J. Kramer, H.K. Hamadeh, J. Collins, S. Grissom, L. Bennett, C.J. Tucker, S. Wild, C. Kind, V. Oreffo, J.W. Davis II, S. Curtiss, J.M. Naciff, M. Cunningham, R. Tennant, J. Stevens, B. Car, T.A. Bertram, and C.A. Afshari. 2004. Identification of putative gene based markers of renal toxicity. Environ. Health Perspect. 112(4):465-479.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Anazawa, Y., H. Nakagawa, M. Furihara, S. Ashida, K. Tamura, H. Yoshioka, T. Shuin, T. Fujioka, T. Katagiri, and Y. Nakamura. 2005. PCOTH, a novel gene overexpressed in prostate cancers, promotes prostate cancer cell growth through phosphorylation of oncoprotein TAF-Ibeta/SET. Cancer Res. 65(11):4578-4586.

Andersen, M.E., H.J. Clewell III, M.L. Gargas, F.A. Smith, and R.H. Reitz. 1987. Physiologically based pharmacokinetics and the risk assessment process for methylene chloride. Toxicol. Appl. Pharmacol. 87(2):185-205.

Anderson, N.L., and N.G. Anderson. 2002. The human plasma proteome: History, character, and diagnostic prospects. Mol. Cell Proteomics 1(11):845-867.

Anderson, N.L., N.G. Anderson, L.R. Haines, D.B. Hardie, R.W. Olafson, and T.W. Pearson. 2004. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J. Proteome Res. 3(2):235-244.

Ando, Y., and Y. Hayashizaki. 2006. Restriction landmark genomic scanning. Nat. Protoc. 1(6):2774-2783.

Andrew, A.S., A.J. Warren, A. Barchowsky, K.A. Temple, L. Klei, N.V. Soucy, K.A. O'Hara, and J.W. Hamilton. 2003. Genomic and proteomic profiling of responses to toxic metals in human lung cells. Environ. Health Perspect. 111(6):825-835.

Andrews, L.B. 2005. Harnessing the benefits of biobanks. J. Law Med. Ethics 33(1):22-30.

Antoniadis, A., S. Lambert-Lacroix, and F. Leblanc. 2003. Effective dimension reduction methods for tumor classification using gene expression data. Bioinformatics 19(5):563-570.

Anway, M.D., A.S. Cupp, M. Uzumcu, and M.K. Skinner. 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 308(5727):1466-1469.

Arbour, N.C., E. Lorenz, B.C. Schutte, J. Zabner, J.N. Kline, M. Jones, K. Frees, J.L. Watt, and D.A. Schwartz. 2000. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat. Genet. 25(2):187-191.

Ausness, R.C. 2002. Will more aggressive marketing practices lead to greater tort liability for prescription drug manufacturers? Wake Forest Law Rev. 37(1):97-139.

Azmi, J., J.L. Griffin, H. Antti, R.F. Shore, E. Johansson, J.K. Nicholson, and E. Holmes. 2002. Metabolic trajectory characterisation of xenobiotic-induced hepatotoxic lesions using statistical batch processing of NMR data. Analyst 127(2):271-276.

Bae, D.S., W.H. Hanneman, R.S. Yang, and J.A. Campain. 2002. Characterization of gene expression changes associated with MNNG, arsenic, or metal mixture treatment in human keratinocytes: Application of cDNA microarray technology. Environ. Health Perspect. 110(Suppl. 6):931-941.

Baggerly, K.A., K.R. Coombes, K.R. Hess, D.N. Stivers, L.V. Abruzzo, and W. Zhang. 2001. Identifying differentially expressed genes in cDNA microarray experiments. J. Comput. Biol. 8(6):639-659.

Baggerly, K.A., J.S. Morris, and K.R. Coombes. 2004. Reproducibility of SELDI-TOF protein patterns in serum: Comparing datasets from different experiments. Bioinformatics 20(5):777-785.

Baggerly, K.A., J.S. Morris, S.R. Edmonson, and K.R. Coombes. 2005. Signal in noise: Evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J. Natl. Cancer Inst. 97(4):307-309.

Bailey, W.J., and R. Ulrich. 2004. Molecular profiling approaches for identifying novel biomarkers. Expert. Opin. Drug Saf. 3(2):137-151.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Baker, V.A., H.M. Harries, J.F. Waring, C.M. Duggan, H.A. Ni, R.A. Jolly, L.W. Yoon, A.T. De Souza, J.E. Schmid, R.H. Brown, R.G. Ulrich, and J.C. Rockett. 2004. Clofibrate-induced gene expression changes in rat liver: A cross-laboratory analysis using membrane cDNA arrays. Environ. Health Perspect. 112(4):428-438.

Ball, C.A, G. Sherlock, H. Parkinson, P. Rocca-Sera, C. Brooksbank, H.C. Causton, D. Cavalieri, T. Gaasterland, P. Hingamp, F. Holstege, M. Ringwald, P. Spellman, C.J. Stoeckert, Jr., J.E. Stewart, R. Taylor, A. Brazma, and J. Quackenbush. 2002a. Standards for microarray data [letter]. Science 298(5593):539.

Ball, C.A., G. Sherlock, H. Parkinson, P. Rocca-Sera, C. Brooksbank, H.C. Causton, D. Cavalieri, T. Gaasterland, P. Hingamp, F. Holstege, M. Ringwald, P. Spellman, C.J. Stoeckert, Jr., J.E. Stewart, R. Taylor, A. Brazma, and J. Quackenbush. 2002b. The underlying principles of scientific publication [review]. Bioinformatics 18(11):1409.

Ball, C.A., A. Brazma, H.C. Causton, S. Chervitz, R. Edgar, P. Hingamp, J.C. Matese, H. Parkinson, J. Quackenbush, M. Ringwald, S.A. Sansone, G. Sherlock, P. Spellman, C. Stoeckert, Y. Tateno, R. Taylor, J. White, and N. Winegarden. (MGED Society). 2004a. Standards for microarray data: An open letter. Environ. Health Perspect. 112(12):A666-A667.

Ball, C.A., A. Brazma, H. Causton, S. Chervitz, R. Edgar, P. Hingamp, J.C. Matese, C. Icahn, H. Parkinson, J. Quackenbush, M. Ringwald, S.A. Sansone, G. Sherlock, P. Spellman, C. Stoeckert, Y. Tateno, R. Taylor, J. White, and N. Winegarden. (MGED Society). 2004b. An open letter on microarray data from the MGED Society. Microbiology 150(Pt 11):3522-3524.

Bammler, T., R.P. Beyer, S. Bhattacharya, G.A. Boorman, A. Boyles, B.U. Bradford, R.E. Bumgarner, P.R. Bushel, K. Chaturvedi, D. Choi, M.L. Cunningham, S. Deng, H.K. Dressman, R.D. Fannin, F.M. Farin, J.H. Freedman, R.C. Fry, A. Harper, M.C. Humble, P. Hurban, T.J. Kavanagh, W.K. Kaufmann, K.F. Kerr, L. Jing, J.A. Lapidus, M.R. Lasarev, J. Li, Y.J. Li, E.K. Lobenhofer, X. Lu, R.L. Malek, S. Milton, S.R. Nagalla, J.P. O'Malley, V.S. Palmer, P. Pattee, R.S. Paules, C.M. Perou, K. Phillips, L.X. Qin, Y. Qiu, S.D. Quigley, M. Rodland, I. Rusyn, L.D. Samson, D.A. Schwartz, Y. Shi, J.L. Shin, S.O. Sieber, S. Slifer, M.C. Speer, P.S. Spencer, D.I. Sproles, J.A. Swenberg, W.A. Suk, R.C. Sullivan, R. Tian, R.W. Tennant, S.A. Todd, C.J. Tucker, B. Van Houten, B.K. Weis, S. Xuan, and H. Zarbl. 2005. Standardizing global gene expression analysis between laboratories and across platforms. Nat. Methods 2(5):351-356.

Bandara, L.R., M.D. Kelly, E.A. Lock, and S. Kennedy. 2003. A correlation between a proteomic evaluation and conventional measurements in the assessment of renal proximal tubular toxicity. Toxicol. Sci. 73(1):195-206.

Baram, M. 2001. Genetic testing for susceptibility to disease from exposure to toxic chemicals: Implications for public and worker health policies. Jurimetrics 41(2):165-176.

Barczak**,** A., M.W. Rodriguez, K. Hanspers, L.L. Koth, T.C. Tai, B.M. Bolstad, T.P. Speed, and D.J. Erle. 2003. Spotted long oligonucleotide arrays for human gene expression analysis. Genome Res. 13(7):1775-1785.

Battuello, K., R. Fenske, C. Furlong, M. Austin, and W. Burke. 2002. CEEH ILSI Core Case Study 1: Genetic Variation in PON1 and Pesticide Sensitivity: Implications for Workplace Screening. March 13, 2002 [online]. Available: http://depts.washington.edu/ceeh/issues/PON.pdf [accessed March 6, 2006].

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

BDW Group (Biomarkers Definitions Working Group). 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 69(3):89-95.

Beckwith-Hall, B.M., J.K. Nicholson, A.W. Nicholls, P.J. Foxall, J.C. Lindon, S.C. Connor, M. Abdi, J. Connelly, and E. Holmes. 1998. Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. Chem. Res. Toxicol. 11(4):260-272.

Begley, T.J., and L.D. Samson. 2004. Network responses to DNA damaging agents. DNA Repair (Amst). 3(8-9):1123-1132.

Begley, T.J., A.S. Rosenbach, T. Ideker, and L.D. Samson. 2002. Damage recovery pathways in Saccharomyces cerevisiae revealed by genomic phenotyping and interactome mapping. Mol. Cancer Res. 1(2):103-112.

Begley, T.J., A.S. Rosenbach, T. Ideker, and L.D. Samson. 2004. Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping. Mol. Cell. 16(1):117-125.

Bellamy, R., C. Ruwende, T. Corrah, K.P. McAdam, H.C. Whittle, and A.V. Hill. 1998. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N. Engl. J. Med. 338(10):640-644.

Bellamy, R., N. Beyers, K.P. McAdam, C. Ruwende, R. Gie, P. Samaai, D. Bester, M. Meyer, T. Corrah, M. Collin, D.R. Camidge, D. Wilkinson, E. Hoal-Van Helden, H.C. Whittle, W. Amos, P. van Helden, and A.V. Hill. 2000. Genetic susceptibility to tuberculosis in Africans: A genome-wide scan. Proc. Natl. Acad. Sci. USA 97(14):8005-8009.

Berra, E., A. Ginouves, and J. Pouyssegur. 2006. The hypoxia-inducible-factor hydroxylases bring fresh air into hypoxia signalling. EMBO Rep. 7(1):41-45.

Bibikova, M., Z. Lin, L. Zhou, E. Chudin, E.W. Garcia, B. Wu, D. Doucet, N.J. Thomas, Y. Wang, E. Vollmer, T. Goldmann, C. Seifart, W. Jiang, D.L. Barker, M.S. Chee, J. Floros, and J.B. Fan. 2006. High-throughput DNA methylation profiling using universal bead arrays. Genome Res. 16(3):383-393.

Birnbaum, L.S., and S.E. Fenton. 2003. Cancer and developmental exposure to endocrine disruptors. Environ. Health Perspect. 111(4):389-394.

Bjornsson, H.T., M.D. Fallin, and A.P. Feinberg. 2004. An integrated epigenetic and genetic approach to common human disease. Trends Genet. 20(8):350-358.

Blackwell, J.M., C.H. Barton, J.K. White, S. Searle, A.M. Baker, H. Williams, and M.A. Shaw. 1995. Genomic organization and sequence of the human NRAMP gene: Identification and mapping of a promoter region polymorphism. Mol. Med. 1(2):194-205.

Blondell, J. 1999. Review of Poison Control Center Data for Residential Exposures to Organophosphate Pesticides, 1993-1996. Memorandum to Dennis Utterback, Team Leader, Organophosphate Residential Team, from Jerome Blondell, Health Statistician, Chemistry and Exposure Branch 2 Health Effects Division, Office of Prevention, Pesticides and Toxic Substances, U.S. Environmental Protection Agency, Washington, DC. February 11, 1999 [online]. Available: http://www.epa.gov/pesticides/op/disulfoton/ppcdata.pdf [accessed May 8, 2007].

Bloom, G., I.V. Yang, D. Boulware, K.Y. Kwong, D. Coppola, S. Eschrich, J. Quackenbush, and T.J. Yeatman. 2004. Multi-platform, multi-site, microarray-based human tumor classification. Am. J. Pathol. 164(1):9-16.

Boess, F., M. Kamber, S. Romer, R. Gasser, D. Muller, S. Albertini, and L. Suter. 2003. Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

slices compared to the in vivo liver gene expression in rats: Possible implications for toxicogenomics use of in vitro systems. Toxicol. Sci. 73(2):386-402.

Bogaerts, J., F. Cardoso, M. Buyse, S. Braga, S. Loi, J.A. Harrison, J. Bines, S. Mook, N. Decker, P. Ravdin, P. Therasse, E. Rutgers, L.J. van 't Veer, and M. Piccart. 2006. Gene signature evaluation as a prognostic tool: Challenges in the design of the MINDACT trial. Nat. Clin. Pract. Oncol. 3(10):540-551.

Bollard, M.E., J. Xu, W. Purcell, J.L. Griffin, C. Quirk, E. Holmes, and J.K. Nicholson. 2002. Metabolic profiling of the effects of D-galactosamine in liver spheroids using (1)H NMR and MAS-NMR spectroscopy. Chem. Res. Toxicol. 15(11):1351-1359.

Bollard, M.E., H.C. Keun, O. Beckonert, T.M. Ebbels, H. Antti, A.W. Nicholls, J.P. Shockcor, G.H. Cantor, G. Stevens, J.C. Lindon, E. Holmes, and J.K. Nicholson. 2005. Comparative metabonomics of differential hydrazine toxicity in the rat and mouse. Toxicol. Appl. Pharmacol. 204(2):135-151.

Bombail, V., J.G. Moggs, and G. Orphanides. 2004. Perturbation of epigenetic status by toxicants. Toxicol. Lett. 149(1-3):51-58.

Bookman, E.B., A.A. Langehorne, J.H. Eckfeldt, K.C. Glass, G.P. Jarvik, M. Klag, G. Koski, A. Motulsky, B. Wilfond, T.A. Manolio, R.R. Fabsitz, and R.V. Luepker. 2006. Reporting genetic results in research studies: Summary and recommendations of an NHLBI working group. Am. J. Med. Genet. A 140(10): 1033-1040.

Boulesteix, A.L., G. Tutz, and K. Strimmer. 2003. A CART-based approach to discover emerging patterns in microarray data. Bioinformatics 19(18):2465-2472.

Boverhof, D.R., K.C. Fertuck, L.D. Burgoon, J.E. Eckel, C. Gennings, and T.R. Zacharewski. 2004. Temporal- and dose-dependent hepatic gene expression changes in immature ovariectomized mice following exposure to ethynyl estradiol. Carcinogenesis. 25(7):1277-1291.

Boverhof, D.R., L.D. Burgoon, C. Tashiro, B. Chittim, J.R. Harkema, D.B. Jump, and T.R. Zacharewski. 2005. Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-Mediated hepatotoxicity. Toxicol. Sci. 85(2):1048-1063.

Bradshaw, R.A. 2005a. DRAFT Proposed Publication Guidelines for the Analysis and Documentation of Peptide and Protein Identifications. July 14, 2005 [online]. Available: http://www.mcponline.org/misc/ParisReport.shtml [accessed Nov. 8, 2005].

Bradshaw, R.A. 2005b. Revised draft guidelines for proteomic data publication. Mol.Cell Proteomics. 4(9):1223-1225.

Branca, A.F., and A. Peck. 2004. Toxicogenomics: A Brenchmark Assessment of Current Practices and Future Expectations. Cambridge Healthtech Advisors, Cambridge, MA.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, T. Gaasterland, P. Glenisson, F.C. Holstege, I.F. Kim, V. Markowitz, J.C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. 29(4):365-371.

Brazma, A., H. Parkinson, U. Sarkans, M. Shojatalab, J. Vilo, N. Abeygunawardena, E. Holloway, M. Kapushesky, P. Kemmeren, G.G. Lara, A. Oezcimen, P. Rocca-Serra, and S.A. Sansone. 2003. ArrayExpress: A public repository for microarray gene expression data at the EBI. Nucleic Acids Res. 31(1):68-71.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Breyer, S.G. 1995. Breaking the Vicious Cycle: Toward Effective Risk Regulation. Cambridge, MA: Harvard University Press.

Brown, M.P., W.N. Grundy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M. Ares, Jr., and D. Haussler. 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc. Natl. Acad. Sci. USA 97(1):262-267.

Brown, S.C., G. Kruppa, and J.L. Dasseux. 2005. Metabolomics applications of FT-ICR mass spectrometry. Mass Spectrom. Rev. 24(2):223-231.

Budinsky, R.A. 2005. A summary of the 2005 winter toxicology forum session on dioxin toxic equivalency factors. Workshop report. Regul. Toxicol. Pharmacol. 43(3): 324-326.

Burke, W., L. Acheson, J. Botkin, K. Bridges, A. Davis, J. Evans, J. Frias, J. Hanson, N. Kahn, R. Kahn, D. Lanier, L.E. Pinsky, N. Press, M.A. Lloyd-Puryear, E. Rich, N. Stevens, E. Thomson, S. Wartman, and M. Wilson. 2002. Genetics in primary care: A USA faculty development initiative. Community Genet. 5(2):138-146.

Bus, J.S., R.A. Canady, T.K. Collier, W. Owens, S.D. Pettit, N.L. Scholz, and A.C. Street. 2007. The extension of molecular and computational information to risk assessment and regulatory decision making. Pp. 151-177 in Genomic Approaches for Cross-Species Extrapolation in Toxicology, W.H. Benson, and R.T. Di Guiulio, eds. Pensacola, FL: SETAC Press.

Butcher, E.C. 2005. Can cell systems biology rescue drug discovery? Nat. Rev. Drug Discov. 4(6):461-467.

Butte, A.J., and I.S. Kohane. 1999. Unsupervised knowledge discovery in medical databases using relevance networks. Proc. AMIA Symp. (1999):711-715.

Buyse, M., S. Loi, L. van't Veer, G. Viale, M. Delorenzi, A.M. Glas, M.S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. Harris, J. Bogaerts, P. Therasse, A. Floore, M. Amakrane, F. Piette, E. Rutgers, C. Sotiriou, F. Cardoso, and M.J. Piccart. 2006. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J. Natl. Cancer Inst. 98(17):1183-1192.

Carruth, R.S., and B.D. Goldstein. 2001. Relative risk greater than two in proof of causation in toxic tort litigation. Jurimetrics 41(2):195-210.

Carter, M.G., T. Hamatani, A.A. Sharov, C.E. Carmack, Y. Qian, K. Aiba, N.T. Ko, D.B. Dudekula, P.M. Brzoska, S.S. Hwang, and M.S. Ko. 2003. In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. Genome Res. 13(5):1011-1021.

Castoldi, M, S. Schmidt, V. Benes, M. Noerholm, A.E. Kulozik, M.W. Hentze, and M.U. Muckenthaler. 2006. A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). RNA 12(5):913-920.

CEBS (Chemical Effects in Biological Systems). 2007. Chemical Effects in Biological Systems. National Institute of Environmental Health Sciences, U.S. National Institutes of Health [online]. Available: http://cebs.niehs.nih.gov/cebs-browser/cebsHome.do [accessed May 30, 2007].

Chadwick, R.F. 2004. The right not to know: A challenge for accurate self-assessment. Philos. Psychiatry Psychology 11(4):299-301.

Chan, E.Y. 2005. Advances in sequencing technology. Mutat Res. 573(1-2):13-40.

Chan, E.Y., N.M. Goncalves, R.A. Haeusler, A.J. Hatch, J.W. Larson, A.M. Maletta, G.R. Yantz, E.D. Carstea, M. Fuchs, G.G. Wong, S.R. Gullans, and R. Gilmanshin. 2004. DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. Genome Res. 14(6):1137-1146.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Chaurand, P., M. Stoeckli, and R.M. Caprioli. 1999. Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. Anal. Chem. 71(23):5263-5270.

Chellappan, S.P. 2001. HOG on the promoter: Regulation of the osmotic stress response. Sci. STKE (93):PE1

Cheung, S.T., X. Chen, X.Y. Guan, S.Y. Wong, L.S. Tai, I.O. Ng, S. So, and S.T. Fan. 2002. Identify metastasis-associated genes in hepatocellular carcinoma through clonality delineation for multinodular tumor. Cancer Res. 62(16):4711-4721.

Cheung, V.G., L.K. Conlin, T.M. Weber, M. Arcaro, K.Y. Jen, M. Morley, and R.S. Spielman. 2003. Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet. 33(3):422-425.

Christiani, D.C., R.R. Sharp, G.W. Collman, and W.A. Suk. 2001. Applying genomic technologies in environmental health research: Challenges and opportunities. J. Occup. Environ. Med. 43(6):526-533.

Chu, T.M., S. Deng, R. Wolfinger, R.S. Paules, and H.K. Hamadeh. 2004. Cross-site comparison of gene expression data reveals high similarity. Environ. Health Perspect. 112(4):449-455.

Churchill, G.A. 2002. Fundamentals of experimental design for cDNA microarrays. Nat. Genet. 32(Suppl.):490-495.

Churchill, G.A., D.C. Airey, H. Allayee, J.M. Angel, A.D. Attie, J. Beatty, W.D. Beavis, J.K. Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K.W. Broman, K.J. Buck, E. Buckler, M. Burmeister, E.J. Chesler, J.M. Cheverud, S. Clapcote, M.N. Cook, R.D. Cox, J.C. Crabbe, W.E. Crusio, A. Darvasi, C.F. Deschepper, R.W. Doerge, C.R. Farber, J. Forejt, D. Gaile, S.J. Garlow, H. Geiger, H. Gershenfeld, T. Gordon, J. Gu, W. Gu, G. de Haan, N.L. Hayes, C. Heller, H. Himmelbauer, R. Hitzemann, K. Hunter, H.C. Hsu, F.A. Iraqi, B. Ivandic, H.J. Jacob, R.C. Jansen, K.J. Jepsen, D.K. Johnson, T.E. Johnson, G. Kempermann, C. Kendziorski, M. Kotb, R.F. Kooy, B. Llamas, F. Lammert, J.M. Lassalle, P.R. Lowenstein, L. Lu, A. Lusis, K.F. Manly, R. Marcucio, D. Matthews, J.F. Medrano, D.R. Miller, G. Mittleman, B.A. Mock, J.S. Mogil, X. Montagutelli, G. Morahan, D.G. Morris, R. Mott, J.H. Nadeau, H. Nagase, R.S. Nowakowski, B.F. O'Hara, A.V. Osadchuk, G.P. Page, B. Paigen, K. Paigen, A.A. Palmer, H.J. Pan, L. Peltonen-Palotie, J. Peirce, D. Pomp, M. Pravenec, D.R. Prows, Z. Qi, R.H. Reeves, J. Roder, G.D. Rosen, E.E. Schadt, L.C. Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M.J. Sillanpaa, L.D. Siracusa, H.W. Snoeck, J.L. Spearow, K. Svenson, L.M. Tarantino, D. Threadgill, L.A. Toth, W. Valdar, F.P. de Villena, C. Warden, S. Whatley, R.W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, F. Zou, Complex Trait Consortium. 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 36(11):1133-1137.

Ciesielski, S., D.P. Loomis, S.R. Mims, and A. Auer. 1994. Pesticide exposures, cholinesterase depression, and symptoms among North Carolina migrant farmworkers. Am. J. Public Health 84(3):446-451.

Clayton, E.W. 2003. Ethical, legal, and social implications of genomic medicine. N. Engl. J. Med. 349(6):562-569.

Clayton, E.W. 2005. Informed consent and biobanks. J. Law Med. Ethics 33(1):15-21.

Clifford, R.J., M.N. Edmonson, C. Nguyen, and K.H. Buetow. 2004. Large-scale analysis of non-synonymous coding region single nucleotide polymorphisms. Bioinformatics 20(7):1006-1014.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Coen, M., E.M. Lenz, J.K. Nicholson, I.D. Wilson, F. Pognan, and J.C. Lindon. 2003. An integrated metabonomic investigation of acetaminophen toxicity in the mouse using NMR spectroscopy. Chem. Res. Toxicol. 16(3):295-303.

Coen, M., S.U. Ruepp, J.C. Lindon, J.K. Nicholson, F. Pognan, E.M. Lenz, and I.D. Wilson. 2004. Integrated application of transcriptomics and metabonomics yields new insight into the toxicity due to paracetamol in the mouse. J. Pharm. Biomed. Anal. 35(1):93-105.

Cohen, S.D., N.R. Pumford, E.A. Khairallah, K. Boekelheide, L.R. Pohl, H.R. Amouzadeh, and J.A. Hinson. 1997. Selective protein covalent binding and target organ toxicity. Toxicol. Appl. Pharmacol. 143(1):1-12.

Colburn, W.A. 2003. Biomarkers in drug discovery and development: From target identification through drug marketing. J. Clin. Pharmacol. 43(4):329-341.

Cole, T.B., B.J. Walter, D.M. Shih, A.D. Tward, A.J. Lusis, C. Timchalk, R.J. Richter, L.G. Costa, and C.E. Furlong. 2005. Toxicity of chlorpyrifos and chlorpyrifos oxon in a transgenic mouse model of the human paraoxonase (PON1) Q192R polymorphism. Pharmacogenet. Genomics. 15(8):589-598.

Conrads, T.P., B.L. Hood, H.J. Issaq, and T.D. Veenstra. 2004. Proteomic patterns as a diagnostic tool for early-stage cancer: A review of its progress to a clinically relevant tool. Mol. Diagn. 8(2):77-85.

Cook, D.N., S. Wang, Y. Wang, G.P. Howles, G.S. Whitehead, K.G. Berman, T.D. Church, B.C. Frank, R.M. Gaspard, Y. Yu, J. Quackenbush, and D.A. Schwartz. 2004. Genetic regulation of endotoxin-induced airway disease. Genomics 83(6):961-969.

Coon, J. J., B. Ueberheide, J.E. Syka, D.D. Dryhurst, J. Ausio, J. Shabanowitz, and D.F. Hunt. 2005. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc. Natl. Acad. Sci. U.S.A 102(27):9463-9468.

Corvi, R., H.J. Ahr, S. Albertini, D.H. Blakey, L. Clerici, S. Coecke, G.R. Douglas, L. Gribaldo, J.P. Groten, B. Haase, K. Hamernik, T. Hartung, T. Inoue, I. Indans, D. Maurici, G. Orphanides, D. Rembges, S.A. Sansone, J.R. Snape, E. Toda, W. Tong, J.H. van Delft, B. Weis, and L.M. Schechtman. 2006. Validation of Toxicogenomics-Based Test Systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. Environ. Health Perspect. 114(3):420-429.

Cosma, G.N., P. Toniolo, D. Currie, B.S. Pasternack, and S.J. Garte. 1992. Expression of the CYP1A1 gene in peripheral lymphocytes as a marker of exposure to creosote in railroad workers. Cancer Epidemiol. Biomarkers Prev. 1(2):137-142.

Craig, R., and R.C. Beavis. 2004. TANDEM: Matching proteins with tandem mass spectra. Bioinformatics 20(9):1466-1467.

Craig, R., J.P. Cortens, and R.C. Beavis. 2004. Open source system for analyzing, validating, and storing protein identification data. J. Proteome Res. 3(6):1234-1242.

Cranor, C.F. 1997. Eggshell skulls and loss of hair from fright: Some moral and legal principles that protect susceptible subpopulations. Environ. Toxicol. Pharmacol. 4(3-4):239-245.

Crespi, C.L., and V.P. Miller. 1999. The use of heterologously expressed drug metabolizing enzymes--state of the art and prospects for the future. Pharmacol. Ther. 84(2):121-131.

Crespi, C.L., B.W. Penman, D.T. Steimel, T. Smith, C.S. Yang, and T.R. Sutter. 1997. Development of a human lymphoblastoid cell line constitutively expressing human CYP1B1 cDNA: Substrate specificity with model substrates and promutagens. Mutagenesis 12(2):83-89.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Cronin, M., K. Ghosh, F. Sistare, J. Quackenbush, V. Vilker, and C. O'Connell. 2004. Universal RNA reference materials for gene expression. Clin. Chem. 50(8):1464-1471.

Cupp, A.S., M. Uzumcu, H. Suzuki, K. Dirks, B. Phillips, and M.K. Skinner. 2003. Effect of transient embryonic in vivo exposure to the endocrine disruptor methoxychlor on embryonic and postnatal testis development. J. Androl. 24(5):736-745.

Daston, G.P. 2004. Developmental toxicity evaluation of butylparaben in Sprague-Dawley rats. Birth Defects Res. B Dev. Reprod. Toxicol 71(4):296-302.

Daston, G.P., and J.M. Naciff. 2005. Gene expression changes related to growth and differentiation in the fetal and juvenile reproductive system of the female rat: Evaluation of microarray results. Reprod. Toxicol. 19(3):381-394.

Datta, A., A. Choudhary, M.L. Bittner, and E.R. Dougherty. 2004. External control in Markovian genetic regulatory networks: The imperfect information case. Bioinformatics 20(6):924-930.

De Sanctis, G.T., and J.M. Drazen. 1997. Genetics of native airway responsiveness in mice. Am. J. Respir. Crit. Care Med. 156(4 Pt 2):S82-S88.

Deschênes, M., and C. Sallée. 2005. Accountability in population biobanking: Comparative approaches. J. Law Med. Ethics 33(1):40-53.

Desiere, F., E.W. Deutsch, A.I. Nesvizhskii, P. Mallick, N.L. King, J.K. Eng, A. Aderem, R. Boyle, E. Brunner, S. Donohoe, N. Fausto, E. Hafen, L. Hood, M.G. Katze, K.A. Kennedy, F. Kregenow, H. Lee, B. Lin, D. Martin, J.A. Ranish, D.J. Rawlings, L.E. Samelson, Y. Shiio, J.D. Watts, B. Wollscheid, M.E. Wright, W. Yan, L. Yang, E.C. Yi, H. Zhang, and R. Aebersold. 2005. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. Genome Biol. 6(1):R9.

Desiere, F., E.W. Deutsch, N.L. King, A.I. Nesvizhskii, P. Mallick, J. Eng, S. Chen, J. Eddes, S.N. Loevenich, and R. Aebersold. 2006. The PeptideAtlas project. Nucleic Acids Res. 34:D655-D658.

de Wildt, R.M., C.R. Mundy, B.D. Gorick, and I.M. Tomlinson. 2000. Antibody arrays for high-throughput screening of antibody-antigen interactions. Nat. Biotechnol. 18(9):989-994.

DHHS (U.S. Department of Health and Human Services). 1997. Issues to Consider in the Research Use of Stored Data or Tissues. Office for Protection from Research Risks, U.S. Department of Health and Human Services [online]. Available: www.hhs.gov/ohrp/humansubjects/guidance/reposit.htm [accessed Nov. 10, 2005].

DHHS (U.S. Department of Health and Human Services). 2004. Guidance on Research Involving Coded Private Information or Biological Specimens. Office for Human Research Protections, U.S. Department of Health and Human Services [online]. Available: http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf [accessed Nov. 10, 2005].

DHHS (U.S. Department of Health and Human Services). 2006. Medical Privacy – National Standards to Protect the Privacy of Personal Health Information. Office for Civil Rights – HIPAA, U.S. Department of Health and Human Services, Washington, DC [online]. Available: http://www.hhs.gov/ocr/hipaa/ [accessed Aug. 24, 2006].

Diamandis, E.P. 2004. Analysis of serum proteomic patterns for early cancer diagnosis: Drawing attention to potential problems. J. Natl. Cancer Inst. 96(5):353-356.

Dickinson, D.A., G.R. Warnes, G. Quievryn, J. Messer, A. Zhitkovich, E. Rubitski, and J. Aubrecht. 2004. Differentiation of DNA reactive and non-reactive genotoxic mechanisms using gene expression profile analysis. Mutat. Res. 549(1-2):29-41.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Dobbin, K.K., and R.M. Simon. 2005. Sample size determination in microarray experiments for class comparison and prognostic classification. Biostat. 6(1):27-38.

Dobbin, K.K., D.G. Beer, M. Meyerson, T.J. Yeatman, W.L. Gerald, J.W. Jacobson, B. Conley, K.H. Buetow, M. Heiskanen, R.M. Simon, J.D. Minna, L. Girard, D.E. Misek, J.M. Taylor, S. Hanash, K. Naoki, D.N. Hayes, C. Ladd-Acosta, S.A. Enkemann, A. Viale, and T.J. Giordano. 2005. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. Clin. Cancer Res. 11(2 Pt. 1):565-572.

Doll, R., and R. Wakeford. 1997. Risk of childhood cancer from fetal irradiation. Br. J. Radiol. 70:130-139.

Doniger, S.W., N. Salomonis, K.D. Dahlquist, K. Vranizan, S.C. Lawlor, and B.R. Conklin. 2003. MAPPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. Genome Biol. 4(1):R7.

Dorne, J.L., K. Walton, W. Slob, and A.G. Renwick. 2002. Human variability in polymorphic CYP2D6 metabolism: Is the kinetic default uncertainty factor adequate? Food Chem. Toxicol. 40(11):1633-1656.

Dorne, J.L., K. Walton, and A.G. Renwick. 2005. Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: A review. Food Chem. Toxicol. 43(2):203-216.

Dudoit, S., R.C. Gentleman, and J. Quackenbush. 2003. Open source software for the analysis of microarray data. Biotechniques (Suppl.):45-51.

EDF (Environmental Defense Fund). 1997. Toxic Ignorance: The Continuing Absence of Basic Health Testing for Top-Selling Chemicals in the United States. New York: Environmental Defense Fund.

Edgar, R., M. Domrachev, and A.E. Lash. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30(1):207-210.

Ehrenberg, L., F. Granath, and M. Tornqvist. 1996. Macromolecule adducts as biomarkers of exposure to environmental mutagens in human populations. Environ. Health Perspect. 104(Suppl. 3):423-428.

Eiseman, E., and S.B. Haga. 1999. Handbook of Human Tissue Sources: A National Resource of Human Tissue Samples. Santa Monica, CA: RAND.

Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95(25):14863-14868.

Elias, J.E., W. Haas, B.K. Faherty, and S.P. Gygi. 2005. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat. Methods 2(9):667-675.

Ellinger-Ziegelbauer, H., B. Stuart, B. Wahle, W. Bomann, and H.J. Ahr. 2004. Characteristic expression profiles induced by genotoxic carcinogens in rat liver. Toxicol. Sci. 77(1):19-34.

Ellis, M., N. Davis, A. Coop, M. Liu, L. Schumaker, R.Y. Lee, R. Srikanchana, C.G. Russell, B. Singh, W.R. Miller, V. Stearns, M. Pennanen, T. Tsangaris, A. Gallagher, A. Liu, A. Zwart, D.F. Hayes, M.E. Lippman, Y. Wang, and R. Clarke. 2002. Development and validation of a method for using breast core needle biopsies for gene expression microarray analyses. Clin. Cancer Res. 8(5):1155-1166.

El-Masri, H.A., D.A. Bell, and C.J. Portier. 1999. Effects of glutathione transferase theta polymorphism on the risk estimates of dichloromethane to humans. Toxicol. Appl. Pharmacol. 158(3):221-230.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Eng, J.K., A.L. McCormack, and J.R.Yates III. 1994. An approach to correlate tandem mass spectra data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5(11):976-989.

EPA (U.S. Environmental Protection Agency). 1986. Guidelines for Carcinogen Risk Assessment. EPA/630/R-00/004. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.

EPA (U.S. Environmental Protection Agency). 2002. Interim Genomics Policy (June 25, 2002). Science Policy Council, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/osa/spc/genomics.htm [accessed Nov. 14, 2005].

EPA (U.S. Environmental Protection Agency). 2003. Exposure and Human Health Reassessment of 2,3,7,8-Tetrachlorodibenzo-*p*-Dioxin (TCDD) and Related Compounds. NAS Review Draft. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC. December 2003 [online]. Available: http://www.epa.gov/ncea/pdfs/dioxin/nas-review/ [accessed Aug. 29, 2006].

EPA (U.S. Environmental Protection Agency). 2004. Potential Implications of Genomics for Regulatory and Risk Assessment Applications at EPA. EPA 100/B-04/002. Genomics Task Force Workgroup, Science Policy Council, U.S. Environmental Protection Agency, Washington, DC. December 2004 [online]. Available: http://www.epa.gov/osa/pdfs/EPA-Genomics-White-Paper.pdf [accessed Nov. 28, 2005].

EPA (U.S. Environmental Protection Agency). 2006a. The HPV Voluntary Challenge Chemical List. U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/chemrtk/pubs/update/hpvchmlt.htm [accessed May 1, 2007].

EPA (U.S. Environmental Protection Agency). 2006b. TSCA Regulations. Civil Enforcement, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/compliance/civil/tsca/tscaenfstatreq.html#regs [accessed May 1, 2007].

Famili, I., R. Mahadevan, and B.O. Palsson. 2005. k-Cone analysis: Determining all candidate values for kinetic parameters on a network scale. Biophys. J. 88(3):1616-1625.

Farber, D.A. 1987. Toxic causation. Minn. L. Rev. 71:1219-1261.

FDA (Food and Drug Administration). 2005a. Guidance for Industry: Pharmacogenomic Data Submissions. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. March 2005 [online]. Available: http://www.fda.gov/cber/gdlns/pharmdtasub.htm [accessed Feb. 3, 2006].

FDA (Food and Drug Administration). 2005b. Drug-Diagnostic Co-Development Concept Paper. U.S. Department of Health and Human Services, Food and Drug Administration. April 2005 [online]. Available: http://www.fda.gov/cder/genomics/pharmacoconceptfn.pdf [accessed Feb. 10, 2006].

Fella, K., M. Gluckmann, J. Hellmann, M. Karas, P.J. Kramer, and M. Kroger. 2005. Use of two-dimensional gel electrophoresis in predictive toxicology: Identification of potential early protein biomarkers in chemically induced hepatocarcinogenesis. Proteomics 5(7):1914-1927.

Fenyo, D. 2000. Identifying the proteome: Software tools. Curr. Opin. Biotechnol. 1(4):391-395.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Fielden, M.R., B.P. Eynon, G. Natsoulis, K. Jarnagin, D. Banas, and K.L. Kolaja. 2005. A gene expression signature that predicts the future onset of drug-induced renal tubular toxicity. Toxicol. Pathol. 33(6):675-683.

Finkel, T., and N.J. Holbrook. 2000. Oxidants, oxidative stress and the biology of ageing. Nature 408(6809):239-247.

Fletcher, N., D. Wahlstrom, R. Lundberg, C.B. Nilsson, K.C. Nilsson, K. Stockling, H. Hellmold, and H. Hakansson. 2005. 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) alters the mRNA expression of critical genes associated with cholesterol metabolism, bile acid biosynthesis, and bile transport in rat liver: A microarray study. Toxicol. Appl. Pharmacol. 207(1):1-24.

Fodor, S.P., R.P. Rava, X.C. Huang, A.C. Pease, C.P. Holmes, and C.L. Adams. 1993. Multiplexed biochemical assays with biological chips. Nature 364(6437):555-556.

Fontenot, A.P., and L.A. Maier. 2005. Genetic susceptibility and immune-mediated destruction in beryllium-induced disease. Trends Immunol. 26(10):543-549.

Forand, P.E., S.J. Kunselman, J.M. Drazen, E. Israel, A. Pillari, T.J. Armstrong, and T.B. Britton. 2001. Genetic analysis in the Asthma Clinical Research Network. Control. Clin. Trials 22(Suppl. 6):S196-S206.

Fornace Jr., A.J., I. Alamo Jr., and M.C. Hollander. 1988. DNA damage-inducible transcripts in mammalian cells. Proc. Natl. Acad. Sci. USA 85(23):8800-8804.

Forrest, M.S., Q. Lan, A.E. Hubbard, L. Zhang, R. Vermeulen, X. Zhao, G. Li, Y.Y. Wu, M. Shen, S. Yin, S.J. Chanock, N. Rothman, and M.T. Smith. 2005. Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. Environ. Health Perspect. 113(6):801-807.

Fortier, A., G. Min-Oo, J. Forbes, S. Lam-Yuk-Tseung, and P. Gros. 2005. Single gene effects in mouse models of host: Pathogen interactions. J. Leukoc. Biol. 77(6):868-877.

Foster, M.W., and R.R. Sharp. 2002. Race, ethnicity, and genomics: Social classifications as proxies of biological heterogeneity. Genome Res. 12(6):844-850.

Fountoulakis, M., P. Berndt, U.A. Boelsterli, F. Crameri, M. Winter, S. Albertini, and L. Suter. 2000. Two-dimensional database of mouse liver proteins: Changes in hepatic protein levels following treatment with acetaminophen or its nontoxic regioisomer 3-acetamidophenol. Electrophoresis 21(11):2148-2161.

Frank, R., and R. Hargreaves. 2003. Clinical biomarkers in drug discovery and development. Nat. Rev. Drug Discov. 2(7):566-580.

Friedman, N., M. Linial, I. Nachman, and D. Pe'er. 2000. Using Bayesian networks to analyze expression data. J. Comput. Biol. 7(3-4):601-620.

Frueh, F.W., K.C. Hayashibara, P.O. Brown, and J.P. Whitlock Jr. 2001. Use of cDNA microarrays to analyze dioxin-induced changes in human liver gene expression. Toxicol. Lett. 122(3):189-203.

Furlong, C.E., T.B. Cole, G.P. Jarvik, C. Pettan-Brewer, G.K. Geiss, R.J. Richter, D.M. Shih, A.D. Tward, A.J. Lusis, and L.G. Costa. 2005a. Role of paraoxonase (PON1) status in pesticide sensitivity: Genetic and temporal determinants. Neurotoxicology 26(4):651-659.

Furlong, C.E., T.B. Cole, B.J. Walter, D.M. Shih, A. Tward, A.J. Lusis, C. Timchalk, R.J. Richter, and L.G. Costa. 2005b. Paraoxonase 1 (PON1) status and risk of insecticide exposure. J. Biochem. Mol. Toxicol. 19(3):182-183.

Gallagher, K., W.H. Benson, M. Brody, A. Fairbrother, J. Hasan, R. Klaper, D. Lattier, S. Lundquist, N. McCarroll, G. Miller, J. Preston, P. Sayre, J. Seed, B. Smith, A. Street, R. Troast, V. Vu, L. Reiter, W. Farland, and K. Dearfield. 2006. Genomics:

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Applications, challenges, and opportunities for the U.S. Environmental Protection Agency. Hum. Ecol. Risk Assess. 12(3):572-590.

Gant, T.W., P.R. Baus, B. Clothier, J. Riley, R. Davies, D.J. Judah, R.E. Edwards, E. George, P. Greaves, and A.G. Smith. 2003. Gene expression profiles associated with inflammation, fibrosis, and cholestasis in mouse liver after griseofulvin. EHP Toxicogenomics 111(1T):37-43.

Gao, J., G.J. Opiteck, M.S. Friedrichs, A.R. Dongre, and S.A. Hefta. 2003. Changes in the protein expression of yeast as a function of carbon source. J. Proteome Res. 2(6):643-649.

Garner, J.M. et al. 2000. Medical monitoring: The evolution of a cause of action. Environ. Law Rep. 30(1):10,024-10,033.

Garosi, P., C. De Filippo, M. van Erk, P. Rocca-Serra, S.A. Sansone, and R. Elliott. 2005. Defining best practice for microarray analyses in nutrigenomic studies. Br. J. Nutr. 93(4):425-432.

Garrod, S., E. Humpher, S.C. Connor, J.C. Connelly, M. Spraul, J.K. Nicholson, and E. Holmes. 2001. High-resolution (1)H NMR and magic angle spinning NMR spectroscopic investigation of the biochemical effects of 2-bromoethanamine in intact renal and hepatic tissue. Magn. Reson. Med. 45(5):781-790.

Ge, Y., B.G. Lawhorn, M. ElNaggar, E. Strauss, J.H. Park, T.P. Begley, and F.W. McLafferty. 2002. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. J. Am. Chem. Soc. 124(4):672-678.

Geller, G., J.R. Botkin, M.J. Green, N. Press, B.B. Biesecker, B. Wilfond, G. Grana, M.B. Daly, K. Schneider, and M.J. Kahn. 1997. Genetic testing for susceptibility to adult-onset cancer. The process and content of informed consent. JAMA 277(18):1467-1474.

Gene Logic, Inc. 2005. ToxExpress Program. Gene Logic, Inc. [online]. Available: http://www.genelogic.com/docs/pdfs/toxExpressProductSheet.pdf [accessed Nov. 8, 2005].

Gerber, S.A., J. Rush, O. Stemman, M.W. Kirschner, and S. P. Gygi. 2003. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. U.S.A 100(12):6940-6945.

Giaever, G., A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A.P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.D. Entian, P. Flaherty, F. Foury, D.J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J.H. Hegemann, S. Hempel, Z. Herman, D.F. Jaramillo, D.E. Kelly, S.L. Kelly, P. Kotter, D. LaBonte, D.C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S.L. Ooi, J.L. Revuelta, C.J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D.D. Shoemaker, S. Sookhai-Mahadeo, R.K. Storms, J.N. Strathern, G. Valle, M. Voet, G. Volckaert, C.Y. Wang, T.R. Ward, J. Wilhelmy, E.A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J.D. Boeke, M. Snyder, P. Philippsen, R.W. Davis, and M. Johnston. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418(6896):387-391.

Ginsburg, G.S., and M. Angrist. 2006. The future may be closer than you think: A response from the Personalized Medicine Coalition to the Royal Society's report on personalized medicine. Personalized Medicine 3(2):119-123.

Golub, T.R., D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Lander. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286(5439):531-537.

Grabher, C., and A.T. Look. 2006. Fishing for cancer models. Nat. Biotechnol. 24(1):45-46.

Gray, J.P., T.L. Leas, E. Obert, D. Brown, G.C. Clark, and J.P. Vanden Heuvel. 2003. Evidence of aryl hydrocarbon receptor ligands in Presque Isle Bay of Lake Erie. Aquat. Toxicol. 64(3):343-358.

Green, M.J., S.K. Peterson, M.W. Baker, G.R. Harper, L.C. Friedman, W.S. Rubinstein, and D.T. Mauger. 2004. Effect of a computer-based decision aid on knowledge, perceptions, and intentions about genetic testing for breast cancer susceptibility: A randomized controlled trial. JAMA 292(4):442-452.

Greenberg, D.S. 2001. Science, Money, and Politics: Political Triumph and Ethical Erosion. Chicago: University of Chicago Press.

Griffin, J.L. 2004. Metabolic profiles to define the genome: Can we hear the phenotypes? Philos. Trans. R Soc. Lond. B Biol. Sci. 359(1446):857-871.

Griffin, J.L., S.A. Bonney, C. Mann, A.M. Hebbachi, G.F. Gibbons, J.K. Nicholson, C.C. Shoulders, and J. Scott. 2004. An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver. Physiol. Genomics 17(2):140-149.

Griffin, R.J., S. Dunwoody, and K. Neuwirth. 1999. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ. Res. 80(2 Pt 2):S230-S245.

Grodsky, J.A. 2005. Genetics and environmental law: Redefining public health. Calif. Law Rev. 93(1):171-270.

Gunderson, K.L., F.J. Steemers, G. Lee, L.G. Mendoza, and M.S. Chee. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. Nat. Genet. 37(5):549-554.

Guo, Y., L.L. Breeden, H. Zarbl, B.D. Preston, and D.L. Eaton. 2005. Expression of a human cytochrome p450 in yeast permits analysis of pathways for response to and repair of aflatoxin-induced DNA damage. Mol. Cell Biol. 25(14):5823-5833.

Guo, Y., L.L. Breeden, W. Fan, L.P. Zhao, D.L. Eaton, and H. Zarbl. 2006. Analysis of cellular responses to aflatoxin B(1) in yeast expressing human cytochrome P450 1A2 using cDNA microarrays. Mutat. Res. 593(1-2):121-142.

Gut, I.G. 2004. DNA analysis by MALDI-TOF mass spectrometry. Hum. Mutat. 23(5):437-441.

Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17(10):994-999.

Gygi, S.P., G.L. Corthals, Y. Zhang, Y. Rochon, and R. Aebersold. 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc. Natl. Acad. Sci. USA 97(17):9390-9395.

Haber, L.T., A. Maier, P.R. Gentry, H.J. Clewell, and M.L. Dourson. 2002. Genetic polymorphisms in assessing interindividual variability in delivered dose. Regul. Toxicol. Pharmacol. 35(2 Pt.1):177-197.

Hadley, D.W., J.F. Jenkins, E. Dimond, M. de Carvalho, I. Kirsch, and C.G. Palmer. 2004. Colon cancer screening practices after genetic counseling and testing for hereditary nonpolyposis colorectal cancer. J. Clin. Oncol. 22(1):39-44.

Hamadeh, H.K., P.R. Bushel, S. Jayadev, K. Martin, O. DiSorbo, S. Sieber, L. Bennett, R. Tennant, R. Stoll, J.C. Barrett, K. Blanchard, R.S. Paules, and C.A. Afshari.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

2002a. Gene expression analysis reveals chemical-specific profiles. Toxicol. Sci. 67(2):219-231.
Hamadeh, H.K., P.R. Bushel, S. Jayadev, O. DiSorbo, L. Bennett, L. Li, R. Tennant, R. Stoll, J.C. Barrett, R.S. Paules, K. Blanchard, and C.A. Afshari. 2002b. Prediction of compound signature using high density gene expression profiling. Toxicol. Sci. 67(2):232-240.
Hamadeh, H.K., S. Jayadev, E.T. Gaillard, Q. Huang, R. Stoll, K. Blanchard, J. Chou, C.J. Tucker, J. Collins, R. Maronpot, P. Bushel, and C.A. Afshari. 2004. Integration of clinical and gene expression endpoints to explore furan-mediated hepatotoxicity. Mutat. Res. 549(1-2):169-183.
Hanahan, D., and R.A. Weinberg. 2000. The hallmarks of cancer. Cell 100(1):57-70.
Hannon, G.J., and J.J. Rossi. 2004. Unlocking the potential of the human genome with RNA interference. Nature 431(7006):371-378.
Hansen, B.T., J.A. Jones, D.E. Mason, and D.C. Liebler. 2001. SALSA: A pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. Anal. Chem. 73(8):1676-1683.
Hansen, B.T., S.W. Davey, A.J. Ham, and D.C. Liebler. 2005. P-Mod: An algorithm and software to map modifications to peptide sequences using tandem MS data. J. Proteome Res. 4(2):358-368.
Harada, M. 1978. Congenital Minamata disease: Intrauterine methylmercury poisoning. Teratology 18(2):285-258.
Hard, M.L., M. Abdolell, B.H. Robinson, and G. Koren. 2005. Gene-expression analysis after alcohol exposure in the developing mouse. J. Lab. Clin. Med. 145(1):47-54.
Hardiman, G. 2004. Microarray platforms--comparisons and contrasts. Pharmacogenomics 5(5):487-502.
Hashimoto, R.F., S. Kim, I. Shmulevich, W. Zhang, M.L. Bittner, and E.R. Dougherty. 2004. Growing genetic regulatory networks from seed genes. Bioinformatics 20(8):1241-1247.
Haugen, A.C., R. Kelley, J.B. Collins, C.J. Tucker, C. Deng, C.A. Afshari, J.M. Brown, T. Ideker, and B. Van Houten. 2004. Integrating phenotypic and expression profiles to map arsenic-response networks. Genome Biol. 5(12):R95.
Hayashi, H., G. Nagae, S. Tsutsumi, K. Kaneshiro, T. Kozaki, A. Kaneda, H. Sugisaki, and H. Aburatani. 2007. High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array. Hum. Genet. 120(5):701-711.
Heijne, W.H., R.H. Stierum, M. Slijper, P.J. van Bladeren, and B. van Ommen. 2003. Toxicogenomics of bromobenzene hepatotoxicity: A combined transcriptomics and proteomics approach. Biochem. Pharmacol. 65(5):857-875.
Heinloth, A.N., R.D. Irwin, G.A. Boorman, P. Nettesheim, R.D. Fannin, S.O. Sieber, M.L. Snell, C.J. Tucker, L. Li, G.S. Travlos, G. Vansant, P.E. Blackshear, R.W. Tennant, M.L. Cunningham, and R.S. Paules. 2004. Gene expression profiling of rat livers reveals indicators of potential adverse effects. Toxicol. Sci. 80(1):193-202
Heinloth, A.N., G.A. Boorman, J.F. Foley, N.D. Flagler, and R.S. Paules. 2007. Gene expression analysis offers unique advantages to histopathology in liver biopsy evaluations. Toxicol. Pathol. 35(2):276-283.
Herrero, J., A. Valencia, and J. Dopazo. 2001. A hierarchical unsupervised growing neural network for clustering gene expression patterns. Bioinformatics 17(2):126-136.
Hinds, D.A., L.L. Stuve, G.B. Nilsen, E. Halperin, E. Eskin, D.G. Ballinger, K.A. Frazer, and D.R. Cox. 2005. Whole-genome patterns of common DNA variation in three human populations. Science 307(5712):1072-1079.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Hirschhorn, J.N., K. Lohmueller, E. Byrne, and K. Hirschhorn. 2002. A comprehensive review of genetic association studies. Genet. Med. 4(2):45-61.
Hoffmann, R., T. Seidl, and M. Dugas. 2002. Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. Genome Biol. 3(7):RESEARCH0033.
Hogstrand, C., S. Balesaria, and C.N. Glover. 2002. Application of genomics and proteomics for study of the integrated response to zinc exposure in a non-model fish species, the rainbow trout. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 133(4):523-535.
Holladay, S.D., and R.J. Smialowicz. 2000. Development of the murine and human immune system: Differential effects of immunotoxicants depend on time of exposure. Environ Health Perspect. 108(Suppl. 3):463-473.
Holmes, E., A.W. Nicholls, J.C. Lindon, S. Ramos, M. Spraul, P. Neidig, S.C. Connor, J. Connelly, S.J. Damment, J. Haselden, and J.K. Nicholson. 1998. Development of a model for classification of toxin-induced lesions using 1H NMR spectroscopy of urine combined with pattern recognition. NMR Biomed. 11(4-5):235-244.
Holtzman, N.A. 1996. Medical and ethical issues in genetic screening: An academic view. Environ. Health Perspect. 104(Suppl. 5):987-990.
Homola, J. 2003. Present and future of surface plasmon resonance biosensors. Anal. Bioanal. Chem. 377(3):528-539.
Hopwood, P., A. Howell, F. Lalloo, and G. Evans. 2003. Do women understand the odds? Risk perceptions and recall of risk information in women with a family history of breast cancer. Community Genet. 6(4):214-223.
Hosack, D.A., G. Dennis, Jr, B.T. Sherman, H.C. Lane, and R.A. Lempicki. 2003. Identifying biological themes within lists of genes with EASE. Genome Biol. 4(10):R70.
Hu, Q., R.J. Noll, H. Li, A. Makarov, M. Hardman, and C.R. Graham. 2005. The Orbitrap: A new mass spectrometer. J. Mass Spectrom. 40(4):430-443.
Hughes, T.R., M. Mao, A.R. Jones, J. Burchard, M.J. Marton, K.W. Shannon, S.M. Lefkowitz, M. Ziman, J.M. Schelter, M.R. Meyer, S. Kobayashi, C. Davis, H. Dai, Y.D. He, S.B. Stephaniants, G. Cavet, W.L. Walker, A. West, E. Coffey, D.D. Shoemaker, R. Stoughton, A.P. Blanchard, S.H. Friend, and P.S. Linsley. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. 19(4):342-347.
HUGO (Human Genome Organisation). 1997. Statement on Patenting Issues Related to Early Release of Raw Sequence Data. HUGO Intellectual Property Committee. May 1997 [online]. Available: http://www.hugo-international.org/committee_ip.htm [accessed Aug. 30, 2006].
HUGO (Human Genome Organisation). 2000. Statement on Benefit-Sharing. HUGO Ethics Committee. April 9, 2000 [online]. Available: http://www.hugo-international.org/Statement_on_Benefit_Sharing.htm [accessed Aug. 30, 2006].
IARC (International Agency for Research on Cancer). 1987. P. 58 in IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Supplement 7. Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs Volumes 1-42. Lyon: International Agency for Research on Cancer.
Iconix Pharmaceuticals. 2005. Bridging Genomics and Chemistry. Products. Applications [online]. Available: http://www.iconixpharm.com/products/products_applications.html [accessed Nov 8, 2005].
Iida, M., C.H. Anna, J. Hartis, M. Bruno, B. Wetmore, J.R. Dubin, S. Sieber, L. Bennett, M.L. Cunningham, R.S. Paules, K.B. Tomer, C.D. Houle, A.B. Merrick, R.C. Sills, and T.R. Devereux. 2003. Changes in global gene and protein expression

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

during early mouse liver carcinogenesis induced by non-genotoxic model carcinogens oxazepam and Wyeth-14,643. Carcinogenesis 24(4):757-770.

ILSI HESI. 2006. International Life Sciences Institute- Health and Environmental Sciences Institute [online]. Available: http://www.hesiglobal.org/ [accessed Jan. 12, 2006].

Imoto, S., S. Kim, T. Goto, S. Miyano, S. Aburatani, K. Tashiro, and S. Kuhara. 2003. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. J. Bioinform. Comput. Biol. 1(2):231-252.

Infante-Rivard, C. 2004. Drinking water contaminants, gene polymorphisms, and fetal growth. Environ. Health Perspect. 112(11):1213-1216.

IOM (Institute of Medicine). 1996. Fetal Alcohol Syndrome: Diagnosis, Epidemiology, Prevention, and Treatment, K. Stratton, S. Howe, and F. Battaglia, eds. Washington, DC: National Academy Press.

IOM (Institute of Medicine). 2004. Health Literacy: A Prescription to End Confusion. Washington, DC: The National Academies Press.

Iorio, M.V., M. Ferracin, C.G. Liu, A. Veronese, R. Spizzo, S. Sabbioni, E. Magri, M. Pedriali, M. Fabbri, M. Campiglio, S. Menard, J.P. Palazzo, A. Rosenberg, P. Musiani, S. Volinia, I. Nenci, G.A. Calin, P. Querzoli, M. Negrini, and C.M. Croce. 2005. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 65(16):7065-7070.

Irizarry, R.A., D. Warren, F. Spencer, I.F. Kim, S. Biswal, B.C. Frank, E. Gabrielson, J.G. Garcia, J. Geoghegan, G. Germino, C. Griffin, S.C. Hilmer, E. Hoffman, A.E. Jedlicka, E. Kawasaki, F. Martinez-Murillo, L. Morsberger, H. Lee, D. Petersen, J. Quackenbush, A. Scott, M. Wilson, Y. Yang, S.Q. Ye, and W. Yu. 2005. Multiple-laboratory comparison of microarray platforms. Nat. Methods 2(5):345-350.

Ishimura, R., S. Ohsako, T. Kawakami, M. Sakaue, Y. Aoki, and C. Tohyama. 2002. Altered protein profile and possible hypoxia in the placenta of 2,3,7,8-tetrachlorodibenzo-p-dioxin-exposed rats. Toxicol. Appl. Pharmacol. 185(3):197-206.

Jaenisch, R., and A. Bird. 2003. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. Nat. Genet. 33(Suppl.):245-254.

Janger, E.J. 2005. Genetic information, privacy and insolvency. J. Law Med. Ethics 33(1):79-88.

Janzi, M., J. Odling, Q. Pan-Hammarstrom, M. Sundberg, J. Lundeberg, M. Uhlen, L. Hammarstrom, and P. Nilsson. 2005. Serum microarrays for large scale screening of protein levels. Mol. Cell. Proteomics 4(12):1942-1947.

Jarvinen, A.K., S. Hautaniemi, H. Edgren, P. Auvinen, J. Saarela, O.P. Kallioniemi, and O. Monni. 2004. Are data from different gene expression microarray platforms comparable? Genomics 83(6):1164-1168.

Jelinsky, S.A., and L.D. Samson. 1999. Global response of *Saccharomyces cerevisiae* to an alkylating agent. Proc. Natl. Acad. Sci. USA 96(4):1486-1491.

Jelinsky, S.A., P. Estep, G.M. Church, and L.D. Samson. 2000. Regulatory networks revealed by transcriptional profiling of damaged Saccharomyces cerevisiae cells: Rpn4 links base excision repair with proteasomes. Mol. Cell Biol. 20(21):8157-8167.

Jessen, B.A., J.S. Mullins, A. De Peyster, and G.J. Stevens. 2003. Assessment of hepatocytes and liver slices as in vitro test systems to predict in vivo gene expression. Toxicol. Sci. 75(1):208-222.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Johnson, C.D., Y. Balagurunathan, K.P. Lu, M. Tadesse, M.H. Falahatpisheh, R.J. Carroll, E.R. Dougherty, C.A. Afshari, and K.S. Ramos. 2003. Genomic profiles and predictive biological networks in oxidant-induced atherogenesis. Physiol. Genomics 13(3):263-275.

Johnson, C.D., Y. Balagurunathan, M.G. Tadesse, M.H. Falahatpisheh, M. Brun, M.K. Walker, E.R. Dougherty, K.S. Ramos. 2004. Unraveling gene-gene interactions regulated by ligands of the aryl hydrocarbon receptor. Environ. Health Perspect. 112(4):403-412.

Johnson, D.E., and G.H. Wolfgang. 2000. Predicting human safety: Screening and computational approaches. Drug Discov. Today 5(10):445-454.

Jolly, R.A., K.M. Goldstein, T. Wei, H. Gao, P. Chen, S. Huang, J.M. Colet, T.P. Ryan, C.E. Thomas, and S.T. Estrem. 2005. Pooling samples within microarray studies: A comparative analysis of rat liver transcription response to prototypical toxicants. Physiol. Genomics 22(3):346-355.

Jones, J.A., L. Kaphalia, M. Treinen-Moslen, and D.C. Liebler. 2003. Proteomic characterization of metabolites, protein adducts, and biliary proteins in rats exposed to 1,1-dichloroethylene or diclofenac. Chem. Res. Toxicol. 16(10):1306-1317.

Julka, S., and F. Regnier. 2004. Quantification in proteomics through stable isotope coding: A review. J. Proteome Res. 3(3):350-363.

Kahn, J.D. 2005. From disparity to difference: How race-specific medicines may undermine policies to address inequalities in health care. Southern California Interdisciplinary Law Journal 15:105-130.

Kane, M.D., T.A. Jatkoe, C.R. Stumpf, J. Lu, J.D. Thomas, and S.J. Madore. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res. 28(22):4552-4557.

Kanno, J. K.I. Aisaki, K. Igarashi, N. Nakatsu, S. Kitajima, and Y. Kodama. 2005. Percellome" and "Mille-Feuille data" System for Toxicogenomics [abstract]. 5th World Congress on Alternatives & Animal Use in the Life Sciences, 5.12 Workshop: Toxicogenomics - Potenial, Validation, and Case Studies, August 24, 2005, Berlin, Germany [online]. Available: http://www.ctw-congress.de/scripts/abs-view/abstract.php?id=25649&topic=383&kid=37-%20http://www.ctw-congress.de/scripts/abs-view/abstract.php?id=25649&topic=383&kid=37 [accessed Nov. 8, 2005].

Kardia, S.L., and C. Wang. 2005. The role of health education and behavior in public health genetics. Health Educ. Behav. 32(5):583-588.

Kelleher, N.L. 2004. Top-down proteomics. Anal. Chem. 76(11):197A-203A.

Keller, A., A.I. Nesvizhskii, E. Kolker, and R. Aebersold. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74(20):5383-5392.

Keller-Seitz, M.U., U. Certa, C. Sengstag, F.E. Wurgler, M. Sun, and M. Fasullo. 2004. Transcriptional response of yeast to aflatoxin B1: Recombinational repair involving RAD51 and RAD1. Mol. Biol. Cell. 15(9):4321-4336.

Kendziorski, C., R.A. Irizarry, K.S. Chen, J.D. Haag, and M.N. Gould. 2005. On the utility of pooling biological samples in microarray experiments. Proc. Natl. Acad Sci. USA 102(12):4252-4257.

Keun, H.C., T.M. Ebbels, M.E. Bollard, O. Beckonert, H. Antti, E. Holmes, J.C. Lindon, and J.K. Nicholson. 2004. Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles. Chem. Res. Toxicol. 17(5):579-587.

Copyright © National Academy of Sciences. All rights reserved.

Kiechl, S., E. Lorenz, M. Reindl, C.J. Wiedermann, F. Oberhollenzer, E. Bonora, J. Willeit, and D.A. Schwartz. 2002. Toll-like receptor 4 polymorphisms and atherogenesis. N. Engl. J. Med. 347(3):185-192.

Kikkawa, R., T. Yamamoto, T. Fukushima, H. Yamada, and I. Horii. 2005. Investigation of a hepatotoxicity screening system in primary cell cultures –"What biomarkers would need to be addressed to estimate toxicity in conventional and new approaches?" J. Toxicol. Sci. 30(1):61-72.

Kim, H-L. 2003. Comparison of oligonucleotide-microarray and serial analysis of gene expression (SAGE) in transcript profiling analysis of megakaryocytes derived from CD34+ cells. EMM 35(5):460-466.

Kim, J.Y., J. Kwon, J.E. Kim, W.S. Koh, M.K. Chung, S. Yoon, C.W. Song, and M. Lee. 2005. Identification of potential biomarkers of genotoxicity and carcinogenicity in L5178Y mouse lymphoma cells by cDNA microarray analysis. Environ. Mol. Mutagen. 45(1):80-89.

Kim, S.K., J. Lund, M. Kiraly, K. Duke, M. Jiang, J.M. Stuart, A. Eizinger, B.N. Wylie, and G.S. Davidson. 2001. A gene expression map for *Caenorhabditis elegans*. Science 293(5537):2087-2092.

Kirby, R., E.J. Cho, B. Gehrke, T. Bayer, Y.S. Park, D.P. Neikirk, J.T. McDevitt, and A.D. Ellington. 2004. Aptamer-based sensor arrays for the detection and quantitation of proteins. Anal. Chem. 76(14):4066-4075.

Kirkpatrick, D.S., S.F. Weldon, G. Tsaprailis, D.C. Liebler, and A.J. Gandolfi. 2005. Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin. Proteomics 5(8):2104-2111.

Kleeberger, S.R. 1991. Understanding biological variability in susceptibility to respiratory disease: Animal models. Pharmacogenetics 1(2):114-118.

Kleeberger, S.R. 2003. Genetic aspects of susceptibility to air pollution. Eur. Respir. J Suppl. 40:52s-56s.

Kleeberger, S.R., and B.B. Hudak. 1992. Acute ozone-induced change in airway permeability: Role of infiltrating leukocytes. J. Appl. Physiol. 72(2):670-676.

Kleeberger, S.R., R.C. Levitt, and L.Y. Zhang. 1993a. Susceptibility to ozone-induced inflammation. I. Genetic control of the response to subacute exposure. Am. J. Physiol. 264(1 Pt 1):L15-L20.

Kleeberger, S.R., R.C. Levitt, and L.Y. Zhang. 1993b. Susceptibility to ozone-induced inflammation. II. Separate loci control responses to acute and subacute exposures. Am. J. Physiol. 264(1 Pt 1):L21-L26.

Kleeberger, S.R., R.C. Levitt, L.Y. Zhang, M. Longphre, J. Harkema, A. Jedlicka, S.M. Eleff, D. DiSilvestre, and K.J. Holroyd. 1997. Linkage analysis of susceptibility to ozone-induced lung inflammation in inbred mice. Nat. Genet. 17(4):475-478.

Klein, A.R. 1999. A model for enhanced risk recovery in tort. Wash. Lee Law Rev. 56(4):1173-1210.

Klein, R.J., C. Zeiss, E.Y. Chew, J.Y. Tsai, R.S. Sackler, C. Haynes, A.K. Henning, J.P. SanGiovanni, S.M. Mane, S.T. Mayne, M.B. Bracken, F.L. Ferris, J. Ott, C. Barnstable, and J. Hoh. 2005. Complement factor H polymorphism in age-related macular degeneration. Science 308(5720):385-389.

Knoppers, B.M. 2005. Biobanking: International norms. J. Law Med. Ethics 33(1):7-14.

Koskinen, H., P. Pehkonen, E. Vehniainen, A. Krasnov, C. Rexroad, S. Afanasyev, H. Molsa, and A. Oikari. 2004. Response of rainbow trout transcriptome to model chemical contaminants. Biochem. Biophys. Res. Commun. 320(3):745-753.

Kramer, J.A., S.D. Pettit, R.P. Amin, T.A. Bertram, B. Car, M. Cunningham, S.W. Curtiss, J.W. Davis, C. Kind, M. Lawton, J.M. Naciff, V. Oreffo, R.J. Roman, F.D.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Sistare, J. Stevens, K. Thompson, A.E. Vickers, S. Wild, and C.A. Afshari. 2004. Overview on the application of transcription profiling using selected nephrotoxicants for toxicology assessment. Environ. Health Perspect. 112(4):460-464.

Kultima, K., A.M. Nystrom, B. Scholz, A.L. Gustafson, L. Dencker, and M. Stigson. 2004. Valproic acid teratogenicity: A toxicogenomics approach. Environ. Health Perspect. 112(12):1225-1235.

Kuo, W.P., T.K. Jenssen, A.J. Butte, L. Ohno-Machado, and I.S. Kohane. 2002. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 18(3):405-412.

Kurachi, M., S. Hashimoto, A. Obata, S. Nagai, T. Nagahata, H. Inadera, H. Sone, C. Tohyama, S. Kaneko, K. Kobayashi, and K. Matsushima. 2002. Identification of 2,3,7,8-tetrachlorodibenzo-p-dioxin-responsive genes in mouse liver by serial analysis of gene expression. Biochem. Biophys. Res. Commun. 292(2):368-377.

La Caze, A. 2005. Does pharmacogenomics provide an ethical challenge to the utilisation of cost-effectiveness analysis by public health systems? Pharmacoeconomics 23(5):445-447.

Lakin, N.D., and S.P. Jackson. 1999. Regulation of p53 in response to DNA damage. Oncogene 18(53):7644-7655.

Lam, S.H., Y.L. Wu, V.B. Vega, L.D. Miller, J. Spitsbergen, Y. Tong, H. Zhan, K.R. Govindarajan, S. Lee, S. Mathavan, K.R. Murthy, D.R. Buhler, E.T. Liu, and Z. Gong. 2006. Conservation of gene expression signatures between zebrafish and human liver tumors and tumor progression. Nat. Biotechnol. 24(1):73-75.

Lame, M.W., A.D. Jones, D.W. Wilson, S.K. Dunston, and H.J. Segall. 2000. Protein targets of monocrotaline pyrrole in pulmonary artery endothelial cells. J. Biol. Chem. 275(37):29091-29099.

Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121(1):185-199.

Larkin, J.E., B.C. Frank, H. Gavras, R. Sultana, and J. Quackenbush. 2005. Independence and reproducibility across microarray platforms. Nat. Methods 2(5):337-344.

Le, Q.T., P.D. Sutphin, S. Raychaudhuri, S.C. Yu, D.J. Terris, H.S. Lin, B. Lum, H.A. Pinto, A.C. Koong, and A.J. Giaccia. 2003. Identification of osteopontin as a prognostic plasma marker for head and neck squamous cell carcinomas. Clin. Cancer Res. 9(1):59-67.

Le Corre, P., R.J. Parmer, M.T. Kailasam, B.P. Kennedy, T.P. Skaar, H. Ho, R. Leverge, D.W. Smith, M.G. Ziegler, P.A. Insel, N.J. Schork, D.A. Flockhart, and D.T. O'Connor. 2004. Human sympathetic activation by α_2-adrenergic blockade with yohimbine: Bimodal, epistatic influence of cytochrome P450-mediated drug metabolism. Clin. Pharmacol. Ther. 76(2):139-153.

Lee, J.S., I.S. Chu, A. Mikaelyan, D.F. Calvisi, J. Heo, J.K. Reddy, and S.S. Thorgeirsson. 2004. Application of comparative functional genomics to identify best-fit mouse models to study human cancer. Nat. Genet. 36(12):1306-1311.

Lerman, C., K. Gold, J. Audrain, T.H. Lin, N.R. Boyd, C.T. Orleans, B. Wilfond, G. Louben, and N. Caporaso. 1997. Incorporating biomarkers of exposure and genetic susceptibility into smoking cessation treatment: Effects on smoking-related cognitions, emotions, and behavior change. Health Psychol. 16(1):87-99.

Li, W.F., L.G. Costa, and C.E. Furlong. 1993. Serum paraoxonase status: A major factor in determining resistance to organophosphates. J. Toxicol. Environ. Health. 40(2-3):337-346.

Liebler, D.C. 2002a. Introduction to Proteomics: Tools for the New Biology. Totowa, NJ: Humana Press.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Liebler, D.C., and F.P. Guengerich. 2005. Elucidating mechanisms of drug-induced toxicity. Nat. Rev. Drug Discov. 4(5):410-420.

Liebler, D.C., B.T. Hansen, S.W. Davey, L. Tiscareno, and D.E. Mason. 2002. Peptide sequence motif analysis of tandem MS data with the SALSA algorithm. Anal. Chem. 74(1):203-210.

Liebler, D.C., B.T. Hansen, J.A. Jones, H. Badghisi, and D.E. Mason. 2003. Mapping protein modifications with liquid chromatography-mass spectrometry and the SALSA algorithm. Adv. Protein Chem. 65:195-216.

Liedberg, B., C. Nylander, and I. Lundstrom. 1995. Biosensing with surface plasmon resonance--how it all started. Biosens. Bioelectron. 10(8):i-ix.

454 Life Sciences. 2006. The 454 Genome Sequencer 20 Sequencing System [online]. Available: http://www.454.com/ [accessed Jan. 3, 2006].

Lindahl, M., B. Stahlbom, and C. Tagesson. 1995. Two-dimensional gel electrophoresis of nasal and bronchoalveolar lavage fluids after occupational exposure. Electrophoresis 16(7):1199-1204.

Lindahl, M., B. Stahlbom, and C. Tagesson. 1999. Newly identified proteins in human nasal and bronchoalveolar lavage fluids: Potential biomedical and clinical applications. Electrophoresis 20(18):3670-3676.

Lindon, J.C., J.K. Nicholson, E. Holmes, H. Antti, M.E. Bollard, H. Keun, O. Beckonert, T.M. Ebbels, M.D. Reily, D. Robertson, G.J. Stevens, P. Luke, A.P. Breau, G.H. Cantor, R.H. Bible, U. Niederhauser, H. Senn, G. Schlotterbeck, U.G. Sidelmann, S.M. Laursen, A. Tymiak, B.D. Car, L. Lehman-McKeeman, J.M. Colet, A. Loukaci, and C. Thomas. 2003. Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. Toxicol. Appl. Pharmacol. 187(3):137-146.

Lindon, J.C., J.K. Nicholson, E. Holmes, H.C. Keun, A. Craig, J.T. Pearce, S.J. Bruce, N. Hardy, S.A. Sansone, H. Antti, P. Jonsson, C. Daykin, M. Navarange, R.D. Beger, E.R. Verheij, A. Amberg, D. Baunsgaard, G.H. Cantor, L. Lehman-McKeeman, M. Earll, S. Wold, E. Johansson, J.N. Haselden, K. Kramer, C. Thomas, J. Lindberg, I. Schuppe-Koistinen, I.D. Wilson, M.D. Reily, D.G. Robertson, H. Senn, A. Krotzky, S. Kochhar, J. Powell, F. van der Ouderaa, R. Plumb, H. Schaefer, and M. Spraul. 2005a. Summary recommendations for standardization and reporting of metabolic analyses. Nat. Biotechnol. 23(7):833-838.

Lindon, J.C., H.C. Keun, T.M. Ebbels, J.M. Pearce, E. Holmes, and J.K. Nicholson. 2005b. The Consortium for Metabonomic Toxicology (COMET): Aims, activities and achievements. Pharmacogenomics 6(7):691-699.

Link, A.J., J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, and J.R. Yates III. 1999. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17(7):676-682.

Lipshutz, R.J., S.P. Fodor, T.R. Gingeras, and D.J. Lockhart. 1999. High density synthetic oligonucleotide arrays. Nat. Genet. 21(Suppl. 1):20-24.

Liu, H., R.G. Sadygov, and J.R. Yates III. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. 76(14):4193-4201.

Liu, J., T.M. Fujiwara, N.T. Buu, F.O. Sanchez, M. Cellier, A.J. Paradis, D. Frappier, E. Skamene, P. Gros, K. Morgan, and E. Schurr. 1995. Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene. Am. J. Hum. Genet. 56(4):845-853.

Livak, K.J. 1999. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal. 14(5-6):143-149.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Livingston, R.J., A. von Niederhausern, A.G. Jegga, D.C. Crawford, C.S. Carlson, M.J. Rieder, S. Gowrisankar, B.J. Aronow, R.B. Weiss, and D.A. Nickerson. 2004. Pattern of sequence variation across 213 environmental response genes. Genome Res. 14(10A):1821-1831.

Lo, H.S., Z. Wang, Y. Hu, H.H. Yang, S. Gere, K.H. Buetow, and M.P. Lee. 2003. Allelic variation in gene expression is common in the human genome. Genome Res.13(8):1855-1862.

Lobenhofer, E.K., X. Cui, L. Bennett, P.L. Cable, B.A. Merrick, G.A. Churchill, and C.A. Afshari. 2004. Exploration of low-dose estrogen effects: Identification of No Observed Transcriptional Effect Level (NOTEL). Toxicol. Pathol. 32(4):482-492.

Long, A.D., H.J. Mangalam, B.Y. Chan, L. Tolleri, G.W. Hatfield, and P. Baldi. 2001. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12. J. Biol. Chem. 276(23):19937-19944.

Lu, X., and D.P. Lane. 1993. Differential induction of transcriptionally active p53 following UV or ionizing radiation: defects in chromosome instability syndromes? Cell 75(4):765-778.

Lum, P.Y., C.D. Armour, S.B. Stepaniants, G. Cavet, M.K. Wolf, J.S. Butler, J.C. Hinshaw, P. Garnier, G.D. Prestwich, A. Leonardson, P. Garrett-Engele, C.M. Rush, M. Bard, G. Schimmack, J.W. Phillips, C.J. Roberts, and D.D. Shoemaker. 2004. Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. Cell 116(1):121-137.

Luyendyk, J.P., W.B. Mattes, L.D. Burgoon, T.R. Zacharewski, J.F. Maddox, G.N. Cosma, P.E. Ganey, and R.A. Roth. 2004. Gene expression analysis points to hemostasis in livers of rats cotreated with lipopolysaccharide and ranitidine. Toxicol. Sci. 80(1):203-213.

MacCoss, M.J. 2005. Computational analysis of shotgun proteomics data. Curr. Opin. Chem. Biol. 9(1):88-94.

MacCoss, M.J., W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark, and J.R. Yates III. 2002. Shotgun identification of protein modifications from protein complexes and lens tissue. Proc. Natl. Acad. Sci. USA 99(12):7900-7905.

MacDonald, N., S. Chevalier, R. Tonge, M. Davison, R. Rowlinson, J. Young, S. Rayner, and R. Roberts. 2001. Quantitative proteomic analysis of mouse liver response to the peroxisome proliferator diethylhexylphthalate (DEHP). Arch. Toxicol. 75(7): 415-424.

MacPhee, M., K.P. Chepenik, R.A. Liddell, K.K. Nelson, L.D. Siracusa, and A.M. Buchberg. 1995. The secretory phospholipase A2 gene is a candidate for the Mom1 locus, a major modifier of ApcMin-induced intestinal neoplasia. Cell 81(6):957-966.

Mah, N., A. Thelin, T. Lu, S. Nikolaus, T. Kuhbacher, Y. Gurbuz, H. Eickhoff, G. Kloppel, H. Lehrach, B. Mellgard, C.M. Costello, and S. Schreiber. 2004. A comparison of oligonucleotide and cDNA-based microarray systems. Physiol. Genomics 16(3):361-370.

Mahadevan, B., C. Keshava, T. Musafia-Jeknic, A. Pecaj, A. Weston, and W.M. Baird. 2005. Altered gene expression patterns in MCF-7 cells induced by the urban dust particulate complex mixture standard reference material 1649a. Cancer Res. 65(4):1251-1258.

Maitra, A., N.V. Adsay, P. Argani, C. Iacobuzio-Donahue, A. De Marzo, J.L. Cameron, C.J. Yeo, and R.H. Hruban. 2003. Multicomponent analysis of the pancreatic ade-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

nocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray. Mod. Pathol. 16(9):902-912.
Majumder, M.A. 2005. Cyberbanks and other virtual research repositories. J. Law Med. Ethics 33(1):31-39.
Malek, R.L., R.B. Irby, Q.M. Guo, K. Lee, S. Wong, M. He, J. Tsai, B. Frank, E.T. Liu, J. Quackenbush, R. Jove, T.J. Yeatman, and N.H. Lee. 2002. Identification of Src transformation fingerprint in human colon cancer. Oncogene 21(47):7256-7265.
Malinowski, M.J. 2005. Technology transfer in biobanking: Credits, debits, and population health futures. J. Law Med. Ethics 33(1):54-69.
Malo, D., K. Vogan, S. Vidal, J. Hu, M. Cellier, E. Schurr, A. Fuks, N. Bumstead, K. Morgan, and P. Gros. 1994. Haplotype mapping and sequence analysis of the mouse Nramp gene predict susceptibility to infection with intracellular parasites. Genomics 23(1):51-61.
Manly, K.F., and J.M. Olson. 1999. Overview of QTL mapping software and introduction to map manager QT. Mamm. Genome 10(4):327-334.
Mann, M., and O.N. Jensen. 2003. Proteomic analysis of post-translational modifications. Nat. Biotechnol. 21(3):255-261.
Manza, L.L., S.G. Codreanu, S.L. Stamer, D.L. Smith, K.S. Wells, R.L. Roberts, and D.C. Liebler. 2004. Global shifts in protein sumoylation in response to electrophile and oxidative stress. Chem. Res. Toxicol. 17(12):1706-1715.
Marchant, G.E. 2000. Genetic susceptibility and biomarkers in toxic injury litigation. Jurimetrics 41(1):67-109.
Marchant, G.E. 2002. Toxicogenomics and toxic torts. Trends Biotechnol. 20(8):329-332.
Marchant, G.E. 2003a. Genomics and toxic substances: Part I – Toxicogenomics. Environ. Law Rep. News Anal. 33(1):10071-10093.
Marchant, G.E. 2003b. Genomics and toxic substances: Part II – Toxicogenomics. Environ. Law Rep. News Anal. 33(9):10641-10667.
Marchant, G.E. 2005. Property rights and benefit-sharing for DNA donors? Jurimetrics J. 45(2):153–178.
Marchant, G.E. 2006. Genetic data in toxic tort litigation. J. Law Policy14:7-37.
Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, and J.M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376-380.
Marras, S.A., F.R. Kramer, and S. Tyagi. 1999. Multiplex detection of single-nucleotide variations using molecular beacons. Genet. Anal. 14(5-6):151-156.
Marsh, S., and H.L. McLeod. 2004. Pharmacogenetics of irinotecan toxicity. Pharmacogenomics 5(7):835-843.
Marsh, S., M. Xiao, J. Yu, R. Ahluwalia, M. Minton, R.R. Freimuth, P.Y. Kwok, and H.L. McLeod. 2004. Pharmacogenomic assessment of carboxylesterases 1 and 2. Genomics 84(4):661-668.
Marshall, J., P. Kupchak, W. Zhu, J. Yantha, T. Vrees, S. Furesz, K. Jacks, C. Smith, I. Kireeva, R. Zhang, M. Takahashi, E. Stanton, and G. Jackowski. 2003. Processing

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. J. Proteome Res. 2(4):361-372.
Mattes, W.B. 2004. Annotation and cross-indexing of array elements on multiple platforms. Environ. Health Perspect. 112(4):506-510.
Mattes, W.B., S.D. Pettit, S.A. Sansone, P.R. Bushel, and M.D. Waters. 2004. Database development in toxicogenomics: Issues and efforts. Environ. Health Perspect. 112(4):495-505.
Mattingly, C.J., G.T. Colby, J.N. Forrest, and J.L. Boyer. 2003. The Comparative Toxicogenomics Database (CTD). Environ. Health. Perspect. 111(6):793-795.
McCanlies, E.C., K. Kreiss, M. Andrew, and A. Weston. 2003. HLA-DPB1 and chronic beryllium disease: A HuGE review. Am. J. Epidemiol. 157(5):388-398.
McGuire, A.L., and R.A. Gibbs. 2006. Genetics: No longer de-identified. Science 312(5772):370-371.
McKeown-Eyssen, G., C. Baines, D.E. Cole, N. Riley, R.F. Tyndale, L. Marshall, and V. Jazmaji. 2004. Case-control study of genotypes in multiple chemical sensitivity: CYP2D6, NAT1, NAT2, PON1, PON2, and MTHFR. Int. J. Epidemiol. 33(5):971-978.
Medeiros, A.M., M.G. Bird, and G. Witz. 1997. Potential biomarkers of benzene exposure. J. Toxicol. Environ. Health 51(6):519-539.
Meek, M.E., J.R. Bucher, S.M. Cohen, V. Dellarco,R.N. Hill, L.D. Lehman-McKeeman, D.G. Longfellow, T. Pastoor, J. See, and D.E. Patto. 2003a. A framework for human relevance analysis of information on carcinogenic modes of action. Crit. Rev. Toxicol. 33(6):591-653.
Meek, B., A. Renwick, and C. Sonich-Mullin. 2003b. Practical application of kinetic data in risk assessment—an IPCS initiative. International Programme on Chemical Safety. Toxicol. Lett. 138(1-2):151-160.
MGED Society. 2005. Microarray Gene Expression Data Society [online]. Available: http://www.mged.org [accessed Nov. 9, 2005].
MIAME. 2002. Minimum Information About a Microarray Experiment – MIAME 1.1 DRAFT 6. Version 1.1. April 1, 2002 [online]. Available: http://www.mged.org/Workgroups/MIAME/miame_1.1.html [accessed Jan. 13, 2005].
MIAME. 2003. Minimum Information About a Microarray Experiment – MIAME for Toxicogenomics (MIAME/Tox). February 12, 2003 [online]. Available: http://www.mged.org/MIAME1.1-DenverDraft.DOC [accessed Jan. 13, 2006].
MIAME. 2005. Minimum Information About a Microarray Experiment – MIAME. Microarray Gene Expression Data Society [online]. Available: http://www.mged.org/Workgroups/MIAME/miame.html [accessed Nov. 8, 2005].
Miller, D.P., G. Liu, I. De Vivo, T.J. Lynch, J.C. Wain, L. Su, and D.C. Christiani. 2002. Combinations of the variant genotypes of *GSTP1*, *GSTM1* and *p53* are associated with an increased lung cancer risk. Cancer Res. 62(10):2819-2823.
Miller, J.C., H. Zhou, J. Kwekel, R. Cavallo, J. Burke, E.B. Butler, B.S. Teh, and B.B. Haab. 2003. Antibody microarray profiling of human prostate cancer sera: Antibody screening and identification of potential biomarkers. Proteomics 3(1):56-63.
Mimura, J., K. Yamashita, K. Nakamura, M. Morita, T.N. Takagi, K. Nakao, M. Ema, K. Sogawa, M. Yasuda, M. Katsuki, and Y. Fujii-Kuriyama. 1997. Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. Genes Cells 2(10):645-654.
Moggs, J.G., H. Tinwell, T. Spurway, H.S. Chang, I. Pate, F.L. Lim, D.J. Moore, A. Soames, R. Stuckey, R. Currie, T. Zhu, I. Kimber, J. Ashby, and G. Orphanides.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

2004. Phenotypic anchoring of gene expression changes during estrogen-induced uterine growth. Environ. Health Perspect. 112(16):1589-1606.

Moggs, J.G., T.C. Murphy, F.L. Lim, D.J. Moore, R. Stuckey, K. Antrobus, I. Kimber, and G. Orphanides. 2005. Anti-proliferative effect of estrogen in breast cancer cells that re-express ERalpha is mediated by aberrant regulation of cell cycle genes. J. Mol. Endocrinol. 34(2):535-551.

Mohlke, K.L., W.C. Nichols, R.J. Westrick, E.K. Novak, K.A. Cooney, R.T. Swank, and D. Ginsburg. 1996. A novel modifier gene for plasma von Willebrand factor level maps to distal mouse chromosome 11. Proc. Natl. Acad. Sci. USA 93(26):15352-15357.

Montesano, R., A.E. Pegg, and G.P. Margison. 1980. Alkylation of DNA and carcinogenicity of N-nitroso compounds. J. Toxicol. Environ. Health 6(5-6): 1001-1008.

Monticelli, S., K.M. Ansel, C. Xiao, N.D. Socci, A.M. Krichevsky, T.H. Thai, N. Rajewsky, D.S. Marks, C. Sander, K. Rajewsky, A. Rao, and K.S. Kosik. 2005. MicroRNA profiling of the murine hematopoietic system. Genome Biol. 6(8):R71.

Mortishire-Smith, R.J., G.L. Skiles, J.W. Lawrence, S. Spence, A.W. Nicholls, B.A. Johnson, and J.K. Nicholson. 2004. Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of a drug-induced toxicity. Chem. Res. Toxicol. 17(2):165-173.

MSU (Michigan State University). 2007. dbZach System. Zacharewski Laboratory Bioinformatics Group, National Food Safety & Toxicology Center, Michigan State University, East Lansing MI [online]. Available: http://dbzach.fst.msu.edu/ [accessed May 9, 2007].

Mukherjee, S., P. Tamayo, S. Rogers, R. Rifkin, A. Engle, C. Campbell, T.R. Golub, and J.P. Mesirov. 2003. Estimating dataset size requirements for classifying DNA microarray data. J. Comput. Biol. 10(2):119-142.

Myers, T.G., E.C. Dietz, N.L. Anderson, E.A. Khairallah, S.D. Cohen, and S.D. Nelson. 1995. A comparative study of mouse liver proteins arylated by reactive metabolites of acetaminophen and its nonhepatotoxic regioisomer, 3'-hydroxyacetanilide. Chem. Res. Toxicol. 8(3):403-413.

Naciff, J.M., M.L. Jump, S.M. Torontali, G.J. Carr, J.P. Tiesman, G.J. Overmann, and G.P. Daston. 2002. Gene expression profile induced by 17alpha-ethynyl estradiol, bisphenol A, and genistein in the developing female reproductive system of the rat. Toxicol. Sci. 68(1):184-199.

Naciff, J.M., K.A. Hess, G.J. Overmann, S.M. Torontali, G.J. Carr, J.P. Tiesman, L.M. Foertsch, B.D. Richardson, J.E. Martinez, and G.P. Daston. 2005a. Gene expression changes induced in the testis by transplacental exposure to high and low doses of 17{alpha}-ethynyl estradiol, genistein, or bisphenol A. Toxicol. Sci. 86(2):396-416.

Naciff, J.M., S.M. Torontali, G.I. Overmann, G.J. Carr, J.P. Tiesman, G.P. Daston. 2005b. Evaluation of the gene expression changes induced by 17-alpha-ethynyl estradiol in the immature uterus/ovaries of the rat using high density oligonucleotide arrays. Birth Defects Res. B Dev. Reprod. Toxicol. 74(2):164-184.

Natsoulis, G., L. El Ghaoui, G.R. Lanckriet, A.M. Tolley, F. Leroy, S. Dunlea, B.P. Eynon, C.I. Pearson, S. Tugendreich, and K. Jarnagin. 2005. Classification of a large microarray data set: Algorithm comparison and analysis of drug signatures. Genome Res. 15(5):724-736.

NBAC (National Bioethics Advisory Commission). 1999. Research Involving Human Biological Materials: Ethical Issues and Policy Guidance, Vol. 1. Report and Recommendations of the National Bioethics Advisory Commission. National Bio-

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

ethics Advisory Commission, Rockville, MD. August 1999 [online]. Available: http://www.georgetown.edu/research/nrcbl/nbac/hbm.pdf [accessed Aug. 29, 2006].

NCI (National Cancer Institute). 2006a. Occupational and Environmental Epidemiology Branch Overview. Division of Cancer, Epidemiology and Genetics, National Cancer Institute [online]. Available: http://dceg.cancer.gov/occu-overview.html [accessed Aug. 24, 2006].

NCI (National Cancer Institute). 2006b. Clinical Proteomic Technologies for Cancer. National Cancer Institute [online]. Available: http://proteomics.cancer.gov/ [accessed Aug. 24, 2006].

NCI (National Cancer Institute). 2006c. The Mouse Models of Human Cancers Consortium (MMHCC). National Cancer Institute [online]. Available: http://emice.nci.nih.gov/ [accessed Aug. 13, 2006].

NCI (National Cancer Institute). 2006d. Cancer Biomedical Informatics Grid (caBIG). National Cancer Institute [online]. Available: https://cabig.nci.nih.gov/ [accessed Aug. 24, 2006].

NCSL (National Conference of State Legislatures). 2005a. State Genetic Privacy Laws [online]. Available: http://www.ncsl.org/programs/health/genetics/prt.htm [accessed Nov. 8, 2005].

NCSL (National Conference of State Legislatures). 2005b. Genetic and Health Insurance State Anti-Discrimination Laws [online]. Available: http://www.ncsl.org/programs/health/genetics/ndishlth.htm [accessed Nov. 14, 2005].

NCSL (National Conference of State Legislatures). 2005c. State Genetic Employment Laws [online]. Available: http://www.ncsl.org/programs/health/genetics/ndiscrim.htm [accessed Nov. 14, 2005].

NCVHS (National Committee on Vital and Health Statistics). 2006. Privacy and Confidentiality in the Nationwide Health Information Network. Letter to Michael O. Leavitt, Secretary, U.S. Department of Health and Human Services, Washington, DC, from Simon P. Cohn, Chairman, National Committee on Vital and Health Statistics. June 22, 2006 [online]. Available: http://www.ncvhs.hhs.gov/060622lt.htm [accessed Aug. 30, 2006].

Nemeth, K.A., A.V. Singh, and T.B. Knudsen. 2005. Searching for biomarkers of developmental toxicity with microarrays: Normal eye morphogenesis in rodent embryos. Toxicol. Appl. Pharmacol. 206(2):219-228.

Nesvizhskii, A.I., and R. Aebersold. 2004. Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. Drug Discov. Today 9(4):173-181.

Nesvizhskii, A.I., A. Keller, E. Kolker, and R. Aebersold. 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75(17):4646-4658.

Newton, R.K., M. Aardema, and J. Aubrecht. 2004. The utility of DNA microarrays for characterizing genotoxicity. Environ Health Perspect. 112(4):420-422.

Nguyen, D.V., and D.M. Rocke. 2002. Tumor classification by partial least squares using microarray gene expression data. Bioinformatics 18(1):39-50.

Nicholls, A.W., E. Holmes, J.C. Lindon, J.P. Shockcor, R.D. Farrant, J.N. Haselden, S.J. Damment, C.J. Waterfield, and J.K. Nicholson. 2001. Metabonomic investigations into hydrazine toxicity in the rat. Chem. Res. Toxicol. 14(8):975-987.

Nicholls, A.W., R.J. Mortishire-Smith, and J.K. Nicholson. 2003. NMR spectroscopic-based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. Chem. Res. Toxicol. 16(11):1395-1404.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Nicholson, J.K., J. Connelly, J.C. Lindon, and E. Holmes. 2002. Metabonomics: A platform for studying drug toxicity and gene function. Nat. Rev. Drug Discov. 1(2):153-161.

Nicholson, J.K., E. Holmes, and I.D. Wilson. 2005. Gut microorganisms, mammalian metabolism and personalized health care. Nat. Rev. Microbiol. 3(5):431-438.

NIH (National Institute of Health). 2007. NIH Mission [online]. Available: http://www.nih.gov/about/ [accessed April 13, 2007].

Noble, H.B. 2000. Concerns grow over reactions to lyme shots. N.Y. Times Nov. 21, 2000, F1.

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press.

NRC (National Research Council). 1984. Toxicity Testing: Strategies to Determine Needs and Priorities. Washington, DC: National Academy Press.

NRC (National Research Council). 1989. Improving Risk Communication. Washington, DC: National Academy Press.

NRC (National Research Council). 1994. Science and Judgment in Risk Assessment. Washington, DC: National Academy Press.

NRC (National Research Council). 2005a. Health Implications of Perchlorate Ingestion. Washington, DC: National Academy Press.

NRC (National Research Council). 2005b. Reaping the Benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health. Washington, DC: The National Academies Press.

NRC (National Research Council). 2005c. Toxicogenomic Technologies and Risk Assessment of Environmental Carcinogens, A Workshop Summary. Washington, DC: The National Academies Press.

NRC (National Research Council). 2005d. Communicating Toxicogenomics Information to Nonexperts. Washington, DC: The National Academies Press.

NRC (National Research Council). 2006a. Toxicity Testing for Assessment of Environmental Agents: Interim Report. Washington, DC: The National Academies Press.

NRC (National Research Council). 2006b. Health Risk from Dioxin and Related Compounds: Evaluation of the EPA Reassessment. Washington, DC: The National Academies Press.

NRC (National Research Council). 2006d. Human Biomonitoring for Environmental Chemicals. Washington, DC: The National Academies Press.

Nuwaysir, E.F., M. Bittner, J. Trent, J.C. Barrett, and C.A. Afshari. 1999. Microarrays and toxicology: The advent of toxicogenomics. Mol. Carcinog. 24(3):153-159.

OECD (Organisation for Economic Co-operation and Development). 2001. Draft Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment. Series on Testing and Assessment. OECD Environmental Health and Safety Publications, Paris.

OECD (Organisation for Economic Co-operation and Development). 2006. Manual for Investigation of HPV Chemicals, Chapter 2: SIDS, the SIDS Plan and the SIDS Dossier. Organisation for Economic Co-operation and Development [online]. Available: http://www.oecd.org/dataoecd/13/18/36045056.pdf [accessed May 1, 2007].

Okey, A.B., L.M. Vella, and P.A. Harper. 1989. Detection and characterization of a low affinity form of cytosolic Ah receptor in livers of mice nonresponsive to induction of cytochrome P1-450 by 3-methylcholanthrene. Mol. Pharmacol. 35(6):823-830.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Okey, A.B., M.A. Franc, I.D. Moffat, N. Tijet, P.C. Boutros, M. Korkalainen, J. Tuomisto, and R. Pohjanvirta. 2005. Toxicological implications of polymorphisms in receptors for xenobiotic chemicals: The case of the aryl hydrocarbon receptor. Toxicol. Appl. Pharmacol. 207(Suppl. 2):43-51.

Oliver, S.G., M.K. Winson, D.B. Kell, and F. Baganz. 1998. Systematic functional analysis of the yeast genome. Trends Biotechnol. 16(9):373-378.

Olle, E.W., A. Sreekumar, R.L. Warner, S.D. McClintock, A.M. Chinnaiyan, M.R. Bleavins, T.D. Anderson, and K.J. Johnson. 2005. Development of an internally controlled antibody microarray. Mol. Cell Proteomics 4(11):1664-1672.

Ong, S.E., B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, and M. Mann. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell Proteomics 1(5):376-386.

Orchard, S., H. Hermjakob, R.K. Julian, Jr., K. Runte, D. Sherman, J. Wojcik, W. Zhu, and R. Apweiler. 2004. Common interchange standards for proteomics data: Public availability of tools and schema. Proteomics 4(2):490-491.

Orentlicher, D. 1997. Genetic privacy in the patient-physician relationship. Pp. 77-92 in Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era, M.A. Rothstein, ed. New Haven: Yale University Press.

Orr, M.S., and U. Scherf. 2002. Large-scale gene expression analysis in molecular target discovery. Leukemia 16(4):473-477.

Osmond, C., and D.J. Barker. 2000. Fetal, infant, and childhood growth are predictors of coronary heart disease, diabetes, and hypertension in adult men and women. Environ. Health Perspect. 108(Suppl. 3):545-553.

Ossorio, P.N. 2004. Societal and ethical issues in pharmacogenomics. Module 2: Fundamentals of Pharmacogenomics in Pharmacogenomics: Applications to Patient Care, R.T. Scheife, ed. Thousand Oaks, CA: Sage.

Ossorio, P., and T. Duster. 2005. Race and genetics: Controversies in biomedical, behavioral, and forensic sciences. Am. Psychol. 60(1):115-128.

Palmer, L.I. 2003. Medical liability for pharmacogenomics. Pp. 187-206 in Pharmacogenomics: Social, Ethical, and Clinical Dimensions, M.A. Rothstein, ed. Hoboken, NJ: John-Wiley & Sons.

Palmer, L.I. 2005. Should liability play a role in social control of biobanks? J. Law Med. Ethics 33(1):70–78.

Park, P.J., Y.A. Cao, S.Y. Lee, J.W. Kim, M.S. Chang, R. Hart, and S. Choi. 2004. Current issues for DNA microarrays: Platform comparison, double linear amplification, and universal RNA reference. J. Biotechnol. 112(3):225-245.

Paweletz, C.P., L. Charboneau, V.E. Bichsel, N.L. Simone, T. Chen, J.W. Gillespie, M.R. Emmert-Buck, M.J. Roth, E.F. Petricoin III, and L.A. Liotta. 2001. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene 20(16):1981-1989.

Pease, A.C., D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, and S.P. Fodor. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. U.S.A. 91(11):5022-5026.

Peden, D.B. 2000. Development of atopy and asthma: Candidate environmental influences and important periods of exposure. Environ. Health Perspect. 108(Suppl 3):475-482.

Pedrioli, P.G., J.K. Eng, R. Hubley, M. Vogelzang, E.W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R.H. Angeletti, R. Apweiler, K. Cheung, C.E. Costello, H. Hermjakob, S. Huang, R.K. Julian, E. Kapp, M.E. McComb, S.G. Oliver, G. Omenn, N.W. Paton,

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

R. Simpson, R. Smith, C.F. Taylor, W. Zhu, and R. Aebersold. 2004. A common open representation of mass spectrometry data and its application to proteomics research. Nat. Biotechnol. 22(11):1459-1466.

Pegg, A.E. 2000. Repair of O(6)-alkylguanine by alkyltransferases. Mutat. Res. 462(2-3):83-100.

Peng, J., J.E. Elias, C.C. Thoreen, L.J. Licklider, and S.P. Gygi. 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large scale protein analysis: the yeast proteome. J.Proteome Res. 2(1):43-50.

Pennie, W., S.D. Pettit, and P.G. Lord. 2004. Toxicogenomics in risk assessment: An overview of an HESI collaborative research program. Environ. Health Perspect. 112(4):417-419.

Pepe, M.S., R. Etzioni, Z. Feng, J.D. Potter, M.L. Thompson, M. Thornquist, M. Winget, and Y. Yasui. 2001. Phases of biomarker development for early detection of cancer. J. Natl. Cancer Inst, 93(14):1054-1061.

Perera, F., R. Santella, and M. Poirier. 1986. Biomonitoring of workers exposed to carcinogens: Immunoassays to benzo(a)pyrene-DNA adducts as a prototype. J. Occup. Med. 28(10):1117-1123.

Perkins, D.N., D.J. Pappin, D.M. Creasy, and J.S. Cottrell. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18):3551-3567.

Pesavento, J.J., Y.B. Kim, G.K. Taylor, and N.L. Kelleher. 2004. Shotgun annotation of histone modifications: A new approach for streamlined characterization of proteins by top down mass spectrometry. J. Am. Chem. Soc. 126(11):3386-3387.

Petricoin, E.F., A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, and L.A. Liotta. 2002a. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359(9306):572-527.

Petricoin, E.F., D.K. Ornstein, C.P. Paweletz, A. Ardekani, P.S. Hackett, B.A. Hitt, A. Velassco, C. Trucco, L. Wiegand, K. Wood, C.B. Simone, P.J. Levine, W.M. Linehan, M.R. Emmert-Buck, S.M. Steinberg, E.C. Kohn, and L.A. Liotta. 2002b. Serum proteomic patterns for detection of prostate cancer. J. Natl. Cancer Inst. 94(20):1576-1578.

Petricoin, E.F., J.L. Hackett, L.J. Lesko, R.K. Puri, S.I. Gutman, K. Chumakov, J. Woodcock, D.W. Feigal, Jr., K.C. Zoon, and F.D. Sistare. 2002c. Medical applications of microarray technologies: A regulatory science perspective. Nat. Genet. 32(Suppl.):474-479.

Phillips, M.S., R. Lawrence, R. Sachidanandam, A.P. Morris, D.J. Balding, M.A. Donaldson, J.F. Studebaker, W.M. Ankener, S.V. Alfisi, F.S. Kuo, A.L. Camisa, V. Pazorov, K.E. Scott, B.J. Carey, J. Faith, G. Katari, H.A. Bhatti, J.M. Cyr, V. Derohannessian, C. Elosua, A.M. Forman, N.M. Grecco, C.R. Hock, J.M. Kuebler, J.A. Lathrop, M.A. Mockler, E.P. Nachtman, S.L. Restine, S.A. Varde, M.J. Hozza, C.A. Gelfand, J. Broxholme, G.R. Abecasis, M.T. Boyce-Jacino, and L.R. Cardon. 2003. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. Nat. Genet. 33(3):382-387.

Pinkerton, K.E., and J.P. Joad. 2000. The mammalian respiratory system and critical windows of exposure for children's health. Environ. Health Perspect. 108(Suppl. 3):457-462.

Plumb, R.S., C.L. Stumpf, M.V. Gorenstein, J.M. Castro-Perez, G.J. Dear, M. Anthony, B.C. Sweatman, S.C. Connor, and J.N. Haselden. 2002. Metabonomics: The use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

shows potential for the screening of rat urine in drug development. Rapid Commun. Mass Spectrom. 16(20):1991-1996.
Potter, J.D. 2004. Toward the last cohort. Cancer Epidemiol. Biomarkers Prev. 13(6):895-897.
Prows, D.R., and G.D. Leikauf. 2001. Quantitative trait analysis of nickel-induced acute lung injury in mice. Am. J. Respir. Cell. Mol. Biol. 24(6):740-746.
Prows, D.R., H.G. Shertzer, M.J. Daly, C.L. Sidman, and G.D. Leikauf. 1997. Genetic analysis of ozone-induced acute lung injury in sensitive and resistant strains of mice. Nat. Genet. 17(4):471-474.
Prows, D.R., M.J. Daly, H.G. Shertzer, and G.D. Leikauf. 1999. Ozone-induced acute lung injury: Genetic analysis of F(2) mice generated from A/J and C57BL/6J strains. Am. J. Physiol. 277(2 Pt 1):L372-L380.
Puga, A., A. Maier, and M. Medvedovic. 2000. The transcriptional signature of dioxin in human hepatoma HepG2 cells. Biochem. Pharmacol. 60(8):1129-1142.
Qian, X., B.W. Scheithauer, K. Kovacs, and R.V. Lloyd. 2005. DNA microarrays: Recent developments and applications to the study of pituitary tissues. Endocrine 28(1):49-56.
Qiu, Y., L.Z. Benet, and A.L. Burlingame. 1998. Identification of the hepatic protein targets of reactive metabolites of acetaminophen in vivo in mice using two-dimensional gel electrophoresis and mass spectrometry. J. Biol. Chem. 273(28): 17940-17953.
Radich, J.P., M. Mao, S. Stepaniants, M. Biery, J. Castle, T. Ward, G. Schimmack, S. Kobayashi, M. Carleton, J. Lampe, and P.S. Linsley. 2004. Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics 83(6):980-988.
Ramachandran, N., E. Hainsworth, B. Bhullar, S. Eisenstein, B. Rosen, A.Y. Lau, J.C. Walter, and J. LaBaer. 2004. Self-assembling protein microarrays. Science 305(5680):86-90.
Ramaswamy, S., P. Tamayo, R. Rifkin, S. Mukherjee, C.H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, and T.R. Golub. 2001. Multiclass cancer diagnosis using tumor gene expression signatures. Proc. Natl. Acad. Sci. USA 98(26):15149-15154.
Ramsahoye, B.H. 2002. Measurement of genome wide DNA methylation by reversed-phase high-performance liquid chromatography. Methods 27(2):156-161.
Ranish, J.A., E.C. Yi, D.M. Leslie, S.O. Purvine, D.R. Goodlett, J. Eng, and R. Aebersold. 2003. The study of macromolecular complexes by quantitative proteomics. Nat. Genet. 33(3):349-355.
Raychaudhuri, S., J.M. Stuart, and R.B. Altman. 2000. Principal components analysis to summarize microarray experiments: Application to sporulation time series. Pac. Symp. Biocomput. (2000):455-466.
Reeder, C.E., and W.M. Dickson. 2003. Economic implications on pharmacogenomics. Pp. 229-250 in Pharmacogenomics: Social, Ethical, and Clinical Dimensions, M.A. Rothstein, ed. Hoboken, NJ: John-Wiley & Sons.
Renegar, G., C.J. Webster, S. Stuerzebecher, L. Harty, S.E. Ide, B. Balkite, T.A. Rogalski-Salter, N. Cohen, B.B. Spear, D.M. Barnes, and C. Brazell. 2006. Returning genetic research results to individuals: Points-to-consider. Bioethics 20(1):24-36.
Renwick, A.G., and N.P. Lazarus. 1998. Human variability and noncancer risk assessment-an analysis of the default uncertainty factor. Regul. Toxicol. Pharmacol. 27(1 Pt. 1):3-20.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Righetti, P.G., and A. Bossi. 1997. Isoelectric focusing in immobilized pH gradients: Recent analytical and preparative developments. Anal. Biochem. 247(1):1-10.

Robosky, L.C., D.F. Wells, L.A. Egnash, M.L. Manning, M.D. Reily, and D.G. Robertson. 2005. Metabonomic identification of two distinct phenotypes in Sprague-Dawley (Crl:CD[SD]) rats. Toxicol. Sci. 87(1):277-284.

Rogojina, A.T., W.E. Orr, B.K. Song, and E.E. Geisert, Jr. 2003. Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. Mol. Vis. 9:482-496.

Rose, G. 1994. The Strategy of Preventive Medicine. New York: Oxford University Press.

Rothstein, M.A. 1996. Preventing the discovery of plaintiff genetic profiles by defendants seeking to limit damages in personal injury litigation. Indiana Law J. 71(4):877-910.

Rothstein, M.A. 1997. Genetic secrets: A policy framework. Pp. 461-495 in Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era, M.A. Rothstein, ed. New Haven: Yale University Press.

Rothstein, M.A. 1998. Genetic privacy and confidentiality: Why they are so hard to protect. J. Law Med. Ethics 26(3):198-201.

Rothstein, M.A. 2000a. Ethical guidelines for medical research on workers. J. Occup. Environ. Med. 42(12):1166-1171.

Rothstein, M.A. 2000b. Genetics and the work force of the next hundred years. Columbia Bus. Law Rev. 3:371-402.

Rothstein, M.A. 2003. Epilogue: Policy prescriptions. Pp. 319-336 in Pharmacogenomics: Social, Ethical, and Clinical Dimensions, M.A. Rothstein, ed. Hoboken, NJ: John-Wiley & Sons.

Rothstein, M.A. 2005a. Currents in contemporary ethics: Research privacy under HIPAA and the common rule. J. Law Med. Ethics 33(1):154-159.

Rothstein, M.A. 2005b. Expanding the ethical analysis of biobanks. J. Law Med. Ethics 33(1):89-101.

Rothstein, M.A., and P.G. Epps. 2001a. Pharmacogenomics and the (ir)relevance of race. Pharmacogenomics J. 1(2):104-108.

Rothstein, M.A., and P.G. Epps. 2001b. Ethical and legal implications of pharmacogenomics. Nat. Rev. Genet. 2(3):228-231.

Rothstein, M.A., and C.A. Hornung. 2004. Public attitudes. Pp. 1-25 in Genetics and Life Insurance: Medical Underwriting and Social Policy, M.A. Rothstein, ed. Cambridge, MA: MIT Press.

Rothstein, M.A., and M.K. Talbott. 2006. Compelled disclosure of health information: Protecting against the greatest potential threat to privacy. JAMA 295(24):2882-2885.

Royal Society. 2005. Personalized Medicine: Hopes and Realities. Policy report. The Royal Society, London. September 2005 [online]. Available: http://www.royal soc.ac.uk/displaypagedoc.asp?id=17570.

Ruepp, S.U., R.P. Tonge, J. Shaw, N. Wallis, and F. Pognan. 2002. Genomics and proteomics analysis of acetaminophen toxicity in mouse liver. Toxicol. Sci. 65(1):135-150.

Ruepp, S., F. Boess, L. Suter, M.C. de Vera, G. Steiner, T. Steele, T. Weiser, and S. Albertini. 2005. Assessment of hepatotoxic liabilities by transcript profiling. Toxicol. Appl. Pharmacol. 207(2 Suppl.):161-170.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Saal, L.H., C. Troein, J. Vallon-Christersson, S. Gruvberger, A. Borg, and C. Peterson. 2002. BioArray Software Environment (BASE): A platform for comprehensive management and analysis of microarray data. Genome Biol. 3(8):SOFTWARE 0003.

Sadygov, R.G., D. Cociorva, and J.R. Yates III. 2004. Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. Nat. Methods 1(3):195-202.

Saeed, A.I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush. 2003. TM4: A free, open-source system for microarray data management and analysis. Biotechniques 34(2):374-378.

Safe, S. 1990. Polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs) and related compounds: Environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). CRC Crit. Rev. Toxicol. 21(1):51-88.

Said, M.R., T.J. Begley, A.V. Oppenheim, D.A. Lauffenburger, and L.D. Samson. 2004. Global network analysis of phenotypic effects: Protein networks and toxicity modulation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 101(52): 18006-18011.

Salerno, R.A., and L.J. Lesko. 2004. Pharmacogenomics in Drug Development and Regulatory Decision-making: The Genomic Data Submission (GDS) Proposal. Pharmacogenomics 5(1):25-30.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74(12):5463-5467.

Sansone, S.A., N. Morrison, P. Rocca-Serra, and J. Fostel. 2004. Standardization initiatives in the (eco)toxicogenomics domain: A review. Comp. Funct. Genomics 5(8):633-641.

Saul, S. 2005. U.S. to Review Drug Intended for One Race. N.Y. Times, P.1. June 13, 2005.

Savoie, C.J., S. Aburatani, S. Watanabe, Y. Eguchi, S. Muta, S. Imoto, S. Miyano, S. Kuhara, and K. Tashiro. 2003. Use of gene networks from full genome microarray libraries to identify functionally relevant drug-affected genes and gene regulation cascades. DNA Res. 10(1):19-25.

Sawada, H., K. Takami, and S. Asahi. 2005. A toxicogenomic approach to drug-induced phospholipidosis: Analysis of its induction mechanism and establishment of a novel in vitro screening system. Toxicol. Sci. 83(2):282-292.

Scarano, M.I., M. Strazzullo, M.R. Matarazzo, and M. D'Esposito. 2005. DNA methylation 40 years later: Its role in human health and disease. J. Cell Physiol. 204(1):21-35.

Schadt, E.E., C. Li, B. Ellis, and W.H. Wong. 2001. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. J. Cell Biochem. (Suppl 37):120-125.

Schechtman, L.M. 2005. Toxicogenomics: ICCVAM Fundamentals for Validation and Regulatory Acceptance. Presentation at NAS Workshop onValidation of Toxicogenomic Technologies: A Focus on Chemical Classification Strategies, July 7-8, 2005, Washington, DC.

Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270(5235):467-470.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Schena, M., D. Shalon, R. Heller, A. Chai, P.O. Brown, and R.W. Davis. 1996. Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. U.S.A. 93(20):10614-10619.

Schubert, E.L., L. Hsu, L.A. Cousens, J. Glogovac, S. Self, B.J. Reid, P.S. Rabinovitch, and P.L. Porter. 2002. Single nucleotide polymorphism array analysis of flow-sorted epithelial cells from frozen versus fixed tissues for whole genome analysis of allelic loss in breast cancer. Am. J. Pathol. 160(1):73-79.

Searfoss, G.H., W.H. Jordan, D.O. Calligaro, E.J. Galbreath, L.M. Schirtzinger, B.R. Berridge, H. Gao, M.A. Higgins, P.C. May, and T.P. Ryan. 2003. Adipsin, a biomarker of gastrointestinal toxicity mediated by a functional gamma-secretase inhibitor. J. Biol. Chem. 278(46):46107-46116.

Searle, S., and J.M. Blackwell. 1999. Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility. J. Med. Genet. 36(4):295-299.

Searfoss, G.H., Jordan, W.H., Calligaro, D.O., Galbreath, E.J., Schirtzinger, L.M., Berridge, B.R., Gao, H., Higgins, M.A., May, P.C., and Ryan, T.P. 2003. Adipsin, a Biomarker of Gastrointestinal Toxicity Mediated by a Functional -Secretase Inhibitor. J. Biol. Chem. 278:46107-46116.

Sen, B., A. Wang, S.D. Hester, JL. Robertson, and D.C. Wolf. 2005. Gene expression profiling of responses to dimethylarsinic acid in female F344 rat urothelium. Toxicology 215(3):214-226.

Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13(11):2498-2504.

Shaw, M.A., D. Clayton, S.E. Atkinson, H. Williams, N. Miller, D. Sibthorpe, and J.M. Blackwell. 1996. Linkage of rheumatoid arthritis to the candidate gene NRAMP1 on 2q35. J. Med. Genet. 33(8):672-677.

Shendure, J., G.J. Porreca, N.B. Reppas, X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K. Zhang, R.D. Mitra, and G.M. Church. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309(5741):1728-1732.

Shiloh, S. 1996. Decision-making in the context of genetic risk. Pp. 82-103 in The Troubled Helix: Social and Psychological Implications of the New Human Genetics, T. Marteau, and M. Richards, eds. Cambridge: Cambridge University Press.

Shingara, J., K. Keiger, J. Shelton, W. Laosinchai-Wolf, P. Powers, R. Conrad, D. Brown, and E. Labourier. 2005. An optimized isolation and labeling platform for accurate microRNA expression profiling. RNA 11(9):1461-1470.

Shippy, R., T.J. Sendera, R. Lockner, C. Palaniappan, T. Kaysser-Kranich, G. Watts, and J. Alsobrook. 2004. Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. BMC Genomics 5(1):61.

Shmulevich, I., E.R. Dougherty, S. Kim, and W. Zhang. 2002a. Probabilistic Boolean Networks: A rule-based uncertainty model for gene regulatory networks. Bioinformatics 18(2):261-274.

Shmulevich, I., E.R. Dougherty, and W. Zhang. 2002b. Gene perturbation and intervention in probabilistic Boolean networks. Bioinformatics 18(10):1319-1331.

Shugart, L., J.M. Holland, and R.O. Rahn. 1983. Dosimetry of PAH skin carcinogenesis: Covalent binding of benzo[a]pyrene to mouse epidermal DNA. Carcinogenesis 4(2):195-198.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Shukla, S.J., and M.E. Dolan. 2005. Use of CEPH and non-CEPH lymphoblast cell lines in pharmacogenetic studies. Pharmacogenomics 6(3):303-310.

Sidransky, D., R. Irizarry, J.A. Califano, X. Li, H. Ren, N. Benoit, and L. Mao. 2003. Serum protein MALDI profiling to distinguish upper aerodigestive tract cancer patients from control subjects. J. Natl. Cancer Inst. 95(22):1711-1717.

Silver, L.M. 1995. Mouse Genetics: Concepts and Applications. New York: Oxford University Press.

Simon, R., M.D. Radmacher, and K. Dobbin. 2002. Design of studies using DNA microarrays. Genet. Epidemiol. 23(1):21-36.

Simon, R., M.D. Radmacher, K. Dobbin, and L.M. McShane. 2003. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J. Natl. Cancer. Inst. 95(1):14-18.

Skipper, P.L., X. Peng, C.K. Soohoo, and S.R. Tannenbaum. 1994. Protein adducts as biomarkers of human carcinogen exposure. Drug Metab. Rev. 26(1-2):111-124.

SMD. 2006. Stanford MicroArray Database. Departments of Biochemistry and Genetics at the School of Medicine, Stanford University [online]. Available: http://genome-www5.stanford.edu/ [accessed Aug. 24, 2006].

Smith, D., B.D. Collins, J. Heil, and T.H. Koch. 2003. Sensitivity and specificity of photoaptamer probes. Mol. Cell. Proteomics 2(1):11-18.

Smith, R.D. 2002. Trends in mass spectrometry instrumentation for proteomics. Trends Biotechnol. 20(Suppl. 12):S3-S7.

Soinov, L.A. 2003. Supervised classification for gene network reconstruction. Biochem. Soc. Trans. 31(Pt 6):1497-1502.

Sorenson, J.R., and B. Cheuvront. 1993. The Human Genome Project and health behavior and health education research. Health Educ. Res. 8(4):589-593.

Soukas, A., P. Cohen, N.D. Socci, and J.M. Friedman. 2000. Leptin-specific patterns of gene expression in white adipose tissue. Genes. Dev. 14(8):963-980.

Spellman, P.T., M. Miller, J. Stewart, C. Troup, U. Sarkans, S. Chervitz, D. Bernhart, G. Sherlock, C. Ball, M. Lepage, M. Swiatek, W.L. Marks, J. Goncalves, S. Markel, D. Iordan, M. Shojatalab, A. Pizarro, J. White, R. Hubley, E. Deutsch, M. Senger, B.J. Aronow, A. Robinson, D. Bassett, C.J. Stoeckert, Jr., and A. Brazma. 2002. Design and implementation of microarray gene expression markup language (MAGE-ML). Genome Biol. 3(9):RESEARCH0046.

Stahl, D.C., K.M. Swiderek, M.T. Davis, and T.D. Lee. 1996. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. J. Am. Soc. Mass Spectrom. 7(6):532-540.

Stearman, R.S., L. Dwyer-Nield, L. Zerbe, S.A. Blaine, Z. Chan, P.A. Bunn, Jr., G.L. Johnson, F.R. Hirsch, D.T. Merrick, W.A. Franklin, A.E. Baron, R.L. Keith, R.A. Nemenoff, A.M. Malkinson, and M.W. Geraci. 2005. Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model. Am. J. Pathol. 167(6):1763-1775.

Steiner, G., L. Suter, F. Boess, R. Gasser, M.C. de Vera, S. Albertini, and S. Ruepp. 2004. Discriminating different classes of toxicants by transcript profiling. Environ. Health Perspect. 112(12):1236-1248.

Stoyanova, R., J.K. Nicholson, J.C. Lindon, and T.R. Brown. 2004. Sample classification based on Bayesian spectral decomposition of metabonomic NMR data sets. Anal. Chem. 76(13):3666-3674.

Stribinskis, V., and K.S. Ramos. 2006. Activation of human long interspersed nuclear element 1 retrotransposition by benzo(a)pyrene, an ubiquitous environmental carcinogen. Cancer Res. 66(5):2616-2620.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Strouse, D.S. 2005. Informed consent to genetic research on banked human tissue. Jurimetrics 45:135-152.

Stuart, J.M., E. Segal, D. Koller, and S.K. Kim. 2003. A gene-coexpression network for global discovery of conserved genetic modules. Science 302(5643):249-255.

Suchard, M.A., P. Yudkin, and J.S. Sinsheimer. 1999. Are general practitioners willing and able to provide genetic services for common diseases? J. Genet. Couns. 8(5):301-311.

Sun, Y.V., D.R. Boverhof, L.D. Burgoon, M.R. Fielden, and T.R. Zacharewski. 2004. Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences. Nucl. Acids Res. 32(15):4512-4523.

Suter, L., M. Haiker, M.C. De Vera, and S. Albertini. 2003. Effect of two 5-HT6 receptor antagonists on the rat liver: A molecular approach. Pharmacogenomics J. 3(6):320-334.

Suter, L., L.E. Babiss, and E.B. Wheeldon. 2004. Toxicogenomics in predictive toxicology in drug development. Chem. Biol. 11(2):161-171.

Sutherland, J.E., and M. Costa. 2003. Epigenetics and the environment. Ann. N.Y. Acad. Sci. 983:151-160.

Sweet-Cordero, A., S. Mukherjee, A. Subramanian, H. You, J.J. Roix, C. Ladd-Acosta, J. Mesirov, T.R. Golub, and T. Jacks. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. Nat. Genet. 37(1):48-55.

Syka, J.E., J.A. Marto, D.L. Bai, S. Horning, M.W. Senko, J.C. Schwartz, B. Ueberheide, B. Garcia, S. Busby, T. Muratore, J. Shabanowitz, and D.F. Hunt. 2004a. Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. J. Proteome Res. 3(3):621-626.

Syka, J.E., J.J. Coon, M.J. Schroeder, J. Shabanowitz, and D.F. Hunt. 2004b. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. USA 101(26):9528-9533.

Syvanen, A.C. 2005. Toward genome-wide SNP genotyping. Nat. Genet. 37 (Suppl.):S5-S10.

Tamada, Y., S. Kim, H. Bannai, S. Imoto, K. Tashiro, S. Kuhara, and S. Miyano. 2003. Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection. Bioinformatics 19(Suppl. 2):II227-II236.

Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA 96(6):2907-2912.

Tan, P.K., T.J. Downey, E.L. Spitznagel, Jr., P. Xu, D. Fu, D.S. Dimitrov, R.A. Lempicki, B.M. Raaka, and M.C. Cam. 2003. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res. 31(19):5676-5684.

Taylor, C.F., N.W. Paton, K.L. Garwood, P.D. Kirby, D.A. Stead, Z. Yin, E.W. Deutsch, L. Selway, J. Walker, I. Riba-Garcia, S. Mohammed, M.J. Deery, J.A. Howard, T. Dunkley, R. Aebersold, D.B. Kell, K.S. Lilley, P. Roepstorff, J.R. Yates III, A. Brass, A.J. Brown, P. Cash, S.J. Gaskell, S.J. Hubbard, and S.G. Oliver. 2003. A systematic approach to modeling, capturing, and disseminating proteomics experimental data. Nat. Biotechnol. 21(3):247-254.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Tennant, R.W. 2002. The National Center for Toxicogenomics: Using new technologies to inform mechanistic toxicology [editorial]. Environ. Health Perspect.110(1):A8-A10.

Teuschler, L., J. Klaunig, E. Carney, J. Chambers, R. Conolly, C. Gennings, J. Giesy, R. Hertzberg, C. Klaassen, R. Kodell, D. Paustenbach, and R. Yang. 2002. Support of science-based decisions concerning the evaluation of the toxicology of mixtures: A new beginning. Regul. Toxicol. Pharmacol. 36(1):34-39.

Thackaberry, E.A., Z. Jiang, C.D. Johnson, K.S. Ramos, and M.K. Walker. 2005. Toxicogenomic profile of 2,3,7,8-tetrachlorodibenzo-p-dioxin in the murine fetal heart: Modulation of cell cycle and extracellular matrix genes. Toxicol. Sci. 88(1):231-241.

Theilhaber, J., T. Connolly, S. Roman-Roman, S. Bushnell, A. Jackson, K. Call, T. Garcia, and R. Baron. 2002. Finding genes in the C2C12 osteogenic pathway by k-nearest-neighbor classification of expression data. Genome Res. 12(1):165-176.

Thomas, D.B., D.L. Gao, S.G. Self, C.J. Allison, Y. Tao, J. Mahloch, R. Ray, Q. Qin, R. Presley, and P. Porter. 1997. Randomized trial of breast self-examination in Shanghai: Methodology and preliminary results. J. Natl. Cancer Inst. 89(5):355-365.

Thomas, R.S., D.R. Rank, S.G. Penn, G.M. Zastrow, K.R. Hayes, K. Pande, E. Glover, T. Silander, M.W. Craven, J.K. Reddy, S.B. Jovanovich, and C.A. Bradfield. 2001. Identification of toxicologically predictive gene sets using cDNA microarrays. Mol. Pharmacol. 60(6):1189-1194.

Thompson, K.L., C.A. Afshari, R.P. Amin, T.A. Bertram, B. Car, M. Cunningham, C. Kind, J.A. Kramer, M. Lawton, M. Mirsky, J.M. Naciff , V. Oreffo, P.S. Pine, and F.D. Sistare. 2004. Identification of platform-independent gene expression markers of cisplatin nephrotoxicity. Environ. Health Perspect. 112(4):488-494.

Thorgeirsson, S.S., J.S. Lee, and J.W. Grisham. 2006. Functional genomics of hepatocellular carcinoma. Hepatology 43(2 Suppl. 1):S145-S150.

Thukral, S.K., P.J. Nordone, R. Hu, L. Sullivan, E. Galambos, V.D. Fitzpatrick, L. Healy, M.B. Bass, M.E. Cosenza, and C.A. Afshari. 2005. Prediction of nephrotoxicant action and identification of candidate toxicity-related biomarkers. Toxicol. Pathol. 33(3):343-355.

Timchalk, C., A. Kousba, and T. S. Poet. 2002. Monte Carlo analysis of the human chlorpyrifos-oxonase (PON1) polymorphism using a physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model. Toxicol. Lett. 135(1-2):51-59.

Todd, J.A., T.J. Aitman, R.J. Cornall, S. Ghosh, J.R. Hall, C.M. Hearne, A.M. Knight, J.M. Love, M.A. McAleer, J.B. Prins, M. Rodriguez, A. Lathrop, A. Pressey, N.H. DeLarato, L.B. Peterson, and L.S. Wicker. 1991. Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature 351(6327):542-547.

Tonge, R., J. Shaw, B. Middleton, R. Rowlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie, and M. Davison. 2001. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics 1(3):377-396.

Toronen, P., M. Kolehmainen, G. Wong, and E. Castren. 1999. Analysis of gene expression data using self-organizing maps. FEBS Lett. 451(2):142-146.

TRC (Toxicogenomics Research Consortium). 2003. TRC Member Publications. Toxicogenomics Research Consortium. Division of Extramural Research and Training, National Institute of Environmental Health Sciences [online]. Available: http://www.niehs.nih.gov/dert/trc/pub-mem.htm [accessed May 9, 2007].

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

TRC (Toxicogenomics Research Consortium). 2005. Toxicogenomics Research Consortium. Division of Extramural Research and Training, National Institute of Environmental Health Sciences [online]. Available: http://www.niehs.nih.gov/dert/trc/home.htm [accessed Feb. 28, 2006].

Tusher, V.G., R. Tibshirani, and G. Chu. 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA 98(9):5116-5121.

UK Biobank Limited. 2007. Biobank: Improving the Health of Future Generation [online]. Available: http://www.ukbiobank.ac.uk/ [accessed Sept. 28, 2007].

Ulrich, R., and S.H. Friend. 2002. Toxicogenomics and drug discovery: Will new technologies help us produce better drugs? Nat. Rev. Drug Discov. 1(1):84-88.

Ulrich, R.G., J.C. Rockett, G.G. Gibson, and S.D. Pettit. 2004. Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. Environ. Health Perspect. 112(4):423-427.

Urushidani, T., and T. Nagao. 2005. Toxicogenomics: Japanese initiative. Pp. 623-631 in Handbook of Toxicogenomics: Strategies and Applications, J. Borlak, ed. Weinheim: Wiley-VCH.

UW-Madison (University of Wisconsin at Madison). 2006. Edge^2: The Environment, Drugs and Gene Expression. Bradfield Laboratory, University of Wisconsin at Madison [online]. Available: http://edge.oncology.wisc.edu/edge.php [accessed May 9, 2007].

Uzumcu, M., H. Suzuki, and M.K. Skinner. 2004. Effect of the anti-androgenic endocrine disruptor vinclozolin on embryonic testis cord formation and postnatal testis development and function. Reprod. Toxicol. 18(6):765-774.

van Delft, J.H., E. van Agen, S.G. van Breda, M.H. Herwijnen, Y.C. Staal, and J.C. Kleinjans. 2005. Comparison of supervised clustering methods to discriminate genotoxic from non-genotoxic carcinogens by gene expression profiling. Mutat. Res. 575(1-2):17-33.

van de Peppel, J., P. Kemmeren, H. van Bakel, M. Radonjic, D. van Leenen, F.C. Holstege. 2003. Monitoring global messenger RNA changes in externally controlled microarray experiments. EMBO Rep. 4(4):387-393.

van de Vijver, M.J., Y.D. He, L.J. van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E.T. Rutgers, S.H. Friend, and R. Bernards. 2002. A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347(25):1999-2009.

van de Waterbeemd, H., and E. Gifford. 2003. ADMET in silico modelling: Towards prediction paradise? Nat. Rev. Drug Discov. 2(3):192-204.

van Leeuwen, A., P.I. Schrier, M.J. Giphart, I.A. Noordermeer, D.J. Ruiter, P. Rubinstein, and J.J. van Rood. 1986. TCA: A polymorphic genetic marker in leukemias and melanoma cell lines. Blood 67(4):1139-1142.

van 't Veer, L.J., H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, and S.H. Friend. 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530-536.

Venter, J.C., H.O. Smith, and L. Hood. 1996. A new strategy for genome sequencing. Nature 381(6581):364-366.

Vezina, C.M., N.J. Walker, and J.R. Olson. 2004. Subchronic exposure to TCDD, PeCDF, PCB126, and PCB153: Effect on hepatic gene expression. Environ. Health Perspect. 112(16):1636-1644.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Villanueva, J., J. Philip, D. Entenberg, C.A. Chaparro, M.K. Tanwar, E.C. Holland, and P. Tempst. 2004. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. Anal. Chem. 76(6):1560-1570.

Von Eggeling, F., A. Gawriljuk, W. Fiedler, G. Ernst, U. Claussen, J. Klose, and I. Romer. 2001. Fluorescent dual colour 2D-protein gel electrophoresis for rapid detection of differences in protein pattern with standard image analysis software. Int. J. Mol. Med. 8(4):373-377.

Wagner, J.A. 2002. Overview of biomarkers and surrogate endpoints in drug development. Dis. Markers 18(2):41-46.

Walum, E., J. Hedander, and P. Garberg. 2005. Research perspectives for pre-screening alternatives to animal experimentation on the relevance of cytotoxicity measurements, barrier passage determinations and high throughput screening in vitro to select potentially hazardous compounds in large sets of chemicals. Toxicol. Appl. Pharmacol. 207(2 Suppl.):393-397.

Wang, C., R. Gonzalez, and S.D. Merajver. 2004. Assessment of genetic testing and related counseling services: Current research and future directions. Soc. Sci. Med. 58(7):1427-1442.

Wang, C., D.J. Bowen, and S.L. Kardia. 2005. Research and practice opportunities at the intersection of health education, health behavior, and genomics. Health Educ. Behav. 32(5):686-701.

Wang, H.T., J.P. Kong, F. Ding, X.Q. Wang, M.R. Wang, L.X. Liu, M. Wu, and Z.H. Liu. 2003. Analysis of gene expression profile induced by EMP-1 in esophageal cancer cells using cDNA Microarray. World J. Gastroenterol. 9(3):392-398.

Wang, J., J. Delabie, H. Aasheim, E. Smeland, and O. Myklebost. 2002. Clustering of the SOM easily reveals distinct gene expression patterns: Results of a reanalysis of lymphoma study. BMC Bioinformatics 3(1):36.

Wang, W., H. Zhou, H. Lin, S. Roy, T.A. Shaler, L.R. Hill, S. Norton, P. Kumar, M. Anderle, and C.H. Becker. 2003. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal. Chem. 75(18):4818-4826.

Wang, X., D.J. Tomso, X. Liu, and D.A. Bell. 2005. Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. Toxicol. Appl. Pharmacol. 207(Suppl. 2):S84-S90.

Wang, Y., M.E. Bollard, H. Keun, H. Antti, O. Beckonert, T.M. Ebbels, J.C. Lindon, E. Holmes, H. Tang, and J.K. Nicholson. 2003. Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning 1H nuclear magnetic resonance spectroscopy of liver tissues. Anal. Biochem. 323(1):26-32.

Wang, Y., J. Hayakawa, F. Long, Q. Yu, A.H. Cho, G. Rondeau, J. Welsh, S. Mittal, I. De Belle, E. Adamson, M. McClelland, and D. Mercola. 2005a. Promoter array studies identify cohorts of genes directly regulated by methylation, copy number change, or transcription factor binding in human cancer cells. Ann. N.Y. Acad. Sci. 1058:162-185.

Wang, Y., H. Tang, J.K. Nicholson, P.J. Hylands, J. Sampson, and E. Holmes. 2005b. A metabonomic strategy for the detection of the metabolic effects of chamomile (Matricaria recutita L.) ingestion. J. Agric. Food Chem. 53(2):191-196.

Wang, Z., H. Fan, H.H. Yang, Y. Hu, K.H. Buetow, and M.P. Lee. 2004. Comparative sequence analysis of imprinted genes between human and mouse to reveal imprinting signatures. Genomics 83(3):395-401.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Wang, Z., D. Neuburg, C. Li, L. Su, J.Y. Kim, J.C. Chen, and D.C. Christiani. 2005. Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. Environ. Health Perspect. 113(2):233-241.

Waring, J.F., R.A. Jolly, R. Ciurlionis, P.Y. Lum, J.T. Praestgaard, D.C. Morfitt, B. Buratto, C. Roberts, E. Schadt, and R.G. Ulrich. 2001. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicol. Appl. Pharmacol. 175(1):28-42.

Waring, J.F., R. Gum, D. Morfitt, R.A. Jolly, R. Ciurlionis, M. Heindel, L. Gallenberg, B. Buratto, and R.G. Ulrich. 2002. Identifying toxic mechanisms using DNA microarrays: Evidence that an experimental inhibitor of cell adhesion molecule expression signals through the aryl hydrocarbon nuclear receptor. Toxicology 181-182:537-550.

Waring, J.F., R.G. Ulrich, N. Flint, D. Morfitt, A. Kalkuhl, F. Staedtler, M. Lawton, J.M. Beekman, and L. Suter. 2004. Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. Environ. Health Perspect. 112(4):439-448.

Washburn, M.P., D. Wolters, and J.R. Yates III. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19(3):242-247.

Washburn, M.P., R. Ulaszek, C. Deciu, D.M. Schieltz, and J.R. Yates III. 2002. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. Anal. Chem. 74(7):1650-1657.

Washington, HA. 1994. Henrietta Lacks: An unsung hero. Emerge (Oct.):29.

Waterland, R.A., and R.L. Jirtle. 2004. Early nutrition, epigenetic changes at transposons and imprinted genes, and enhanced susceptibility to adult chronic diseases. Nutrition 20(1):63-68.

Waters, M., G. Boorman, P. Bushel, M. Cunningham, R. Irwin, A. Merrick, K. Olden, R. Paules, J. Selkirk, S. Stasiewicz, B. Weis, B. Van Houten, N. Walker, and R. Tennant. 2003b. Systems toxicology and the Chemical Effects in Biological Systems (CEBS) knowledge base. EHP Toxicogenomics 111(1T):15-28.

Waters, N.J, S. Garrod, R.D. Farrant, J.N. Haselden, S.C. Connor, J. Connelly, J.C. Lindon, E. Holmes, and J.K. Nicholson. 2000. High-resolution magic angle spinning (1)H NMR spectroscopy of intact liver and kidney: Optimization of sample preparation procedures and biochemical stability of tissue during spectral acquisition. Anal. Biochem. 282(1):16-23.

Waters, N.J., E. Holmes, A. Williams, C.J. Waterfield, R.D. Farrant, and J.K. Nicholson. 2001. NMR and pattern recognition studies on the time-related metabolic effects of alpha-naphthylisothiocyanate on liver, urine, and plasma in the rat: An integrative metabonomic approach. Chem. Res. Toxicol. 14(10):1401-1412.

Waters, N.J., E. Holmes, C.J. Waterfield, R.D. Farrant, and J.K. Nicholson. 2002. NMR and pattern recognition studies on liver extracts and intact livers from rats treated with alpha-naphthylisothiocyanate. Biochem. Pharmacol. 64(1):67-77.

Waters, N.J., C.J. Waterfield, R.D. Farrant, E. Holmes, and J.K. Nicholson. 2005. Metabonomic deconvolution of embedded toxicity: Application to thioacetamide hepato- and nephrotoxicity. Chem. Res. Toxicol. 18(4):639-654.

Weber, W.W. 1997. Pharmacogenetics. Oxford: Oxford University Press.

Weijer, C., and P.B. Miller. 2004. Protecting communities in pharmacogenetic and pharmacogenomic research. Pharmacogenomics J. 4(1):9-16.

Weinstein, J.N., T.G. Myers, P.M. O'Connor, S.H. Friend, A.J. Fornace Jr, K.W. Kohn, T. Fojo, S.E. Bates, L.V. Rubinstein, N.L. Anderson, J.K. Buolamwini, W.W. van

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Osdol, A.P. Monks, D.A. Scudiero, E.A. Sausville, D.W. Zaharevitz, B. Bunow, V.N. Viswanadhan, G.S. Johnson, R.E. Wittes, and K.D. Paull. 1997. An information-intensive approach to the molecular pharmacology of cancer. Science 275(5298):343-349.

Welch, K.D., B. Wen, D.R. Goodlett, E.C. Yi, H. Lee, T.P. Reilly, S.D. Nelson, and L.R. Pohl. 2005. Proteomic identification of potential susceptibility factors in drug-induced liver disease. Chem. Res. Toxicol. 18(6):924-933.

Wen, X., S. Fuhrman, G.S. Michaels, D.B. Carr, S. Smith, J.L. Barker, and R. Somogyi. 1998. Large-scale temporal gene expression mapping of central nervous system development. Proc. Natl. Acad. Sci USA 95(1):334-339.

Wesselkamper, S.C., L.M. Case, L.N. Henning, M.T. Borchers, J.W. Tichelaar, J.M. Mason, N. Dragin, M. Medvedovic, M.A. Sartor, C.R. Tomlinson, and G.D. Leikauf. 2005. Gene expression changes during the development of acute lung injury: Role of TGF-{beta}. Am. J. Respir. Crit. Care Med. 172(11):1399-1411.

Weston, W.M., E.E. LeClair, W. Trzyna, K.M. McHugh, P. Nugent, C.M. Lafferty, L. Ma, R.S. Tuan, and R.M. Greene. 1999. Differential display identification of plunc, a novel gene expressed in embryonic palate, nasal epithelium, and adult lung. J. Biol. Chem. 274(19):13698-13703.

Wiback, S.J., R. Mahadevan, and B.O. Palsson. 2004. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The Escherichia coli spectrum. Biotechnol. Bioeng. 86(3):317-331.

Wilkinson, C.F., G.R. Christoph, E. Julien, J.M. Kelley, J. Kronenberg, J. McCarthy, and R. Reiss. 2000. Assessing the risks of exposures to multiple chemicals with a common mechanism of toxicity: How to cumulate? Regul. Toxicol. Pharmacol. 31(1):30-43.

Williams, R.E., H. Major, E.A. Lock, E.M. Lenz, and I.D. Wilson. 2005. D-Serine-induced nephrotoxicity: A HPLC-TOF/MS-based metabonomics approach. Toxicology 207(2):179-190.

Wilson, I.D., and J.K. Nicholson. 2003. Topics in xenobiochemistry: Do metabolic pathways exist for xenobiotics? The micro-metabolism hypothesis. Xenobiotica 33(9):887-901.

Wilson, I.D., R. Plumb, J. Granger, H. Major, R. Williams, and E.M. Lenz. 2005a. HPLC-MS-based methods for the study of metabonomics. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 817(1):67-76.

Wilson, I.D., J.K. Nicholson, J. Castro-Perez, J.H. Granger, K.A. Johnson, B.W. Smith, and R.S. Plumb. 2005b. High resolution “ultra performance” liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. J. Proteome Res. 4(2):591-598.

Wiltse, J. 2005. Mode of action: Inhibition of histone deacetylase, altering WNT-dependent gene expression, and regulation of beta-catenin – developmental effects of valproic acid. Crit. Rev. Toxicol. 35(8-9):727-738.

Witte, K. 1994. Fear control and danger control: A test of the Extended Parallel Process Model (EPPM). Commun. Monogr. 61(2):113-134.

Wolpert, D.H., and W.G. MacReady. 1997. No free lunch theorems for optimization, IEEE T. Evolut. Comput. 1(1):67-82.

Wolters, D.A., M.P. Washburn, and J.R.Yates, III. 2001. An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73(23): 5683-5690.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Workman, C.T., H.C. Mak, S. McCuine, J.B. Tagne, M. Agarwal, O. Ozier, T.J. Begley, L.D. Samson, and T. Ideker. 2006. A systems approach to mapping DNA damage response pathways. Science 312(5776):1054-1059.

Wu, M.T., B. Demple, R.A. Bennett, D.C. Christiani, R. Fan, and H. Hu. 2000. Individual variability in the zinc inducibility of metallothionein-IIA mRNA in human lymphocytes. J. Toxicol. Environ. Health A 61(7):553-567.

Wu, S.L., J. Kim, W.S. Hancock, and B. Karger. 2005. Extended Range Proteomic Analysis (ERPA): A new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). J. Proteome Res. 4(4):1155-1170.

Xenarios, I., L. Salwinski, X.J. Duan, P. Higney, S.M. Kim, and D. Eisenberg. 2002. DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions. Nucleic Acids Res. 30(1):303-305.

Xiao, G.G., M. Wang, N. Li, J.A. Loo, and A.E. Nel. 2003. Use of proteomics to demonstrate a hierarchical oxidative stress response to diesel exhaust particle chemicals in a macrophage cell line. J. Biol. Chem. 278(50):50781-50790.

Xu, X.Q., C.K. Leow, X. Lu, X. Zhang, J.S. Liu, W.H. Wong, A. Asperger, S. Deininger, and H.C. Eastwood Leung. 2004. Molecular classification of liver cirrhosis in a rat model by proteomics and bioinformatics. Proteomics 4(10):3235-3245.

Yamamoto, T., R. Kikkawa, H. Yamada, and I. Horii. 2005. Identification of oxidative stress-related proteins for predictive screening of hepatotoxicity using a proteomic approach. J. Toxicol. Sci. 30(3):213-227.

Yang, I., I.Y. Park, S.M. Jang, L.H. Shi, H.K. Ku, and S.R. Park. 2006. Rapid quantification of DNA methylation through dNMP analysis following bisulfite-PCR. Nucleic Acids Res. 34(8):e61.

Yang, I.V., E. Chen, J.P. Hasseman, W. Liang, B.C. Frank, S. Wang, V. Sharov, A.I. Saeed, J. White, J. Li, N.H. Lee, T.J. Yeatman, and J. Quackenbush. 2002. Within the fold: Assessing differential expression measures and reproducibility in microarray assays. Genome Biol. 3(11):RESEARCH0062.

Yang, R.S.H. 1994. Toxicology of Chemical Mixtures: Case Studies, Mechanisms, and Novel Approaches. San Diego: Academic Press.

Yang, Y.H., S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai, and T.P. Speed. 2002. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30(4):e15.

Yao, G., M. Craven, N. Drinkwater, and C.A. Bradfield. 2004. Interaction networks in yeast define and enumerate the signaling steps or the vertebrate aryl hydrocarbon receptor. PloS Biol. 2(3):E65.

Yates, J.R. III. 1998. Mass spectrometry and the age of the proteome. J. Mass. Spectrom. 33(1):1-19.

Yates, J.R. III. 2000. Mass spectrometry: From genomics to proteomics. Trends Genet. 16(1):5-8.

Yauk, C.L., M.L. Berndt, A. Williams, and G.R. Douglas. 2004. Comprehensive comparison of six microarray technologies. Nucleic Acids Res. 32(15):e124.

Yuen, T., E. Wurmbach, R.L. Pfeffer, B.J. Ebersole, and S.C. Sealfon. 2002. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. Nucleic Acids Res. 30(10):E48.

Yuk, J.S., and K.S. Ha. 2005. Proteomic applications of surface plasmon resonance biosensors: Analysis of protein arrays. Exp. Mol. Med. 37(1):1-10.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Yuk, J.S., S.H. Jung, J.W. Jung, D.G. Hong, J.A. Han, Y.M. Kim, and K.S. Ha. 2004. Analysis of protein interactions on protein arrays by a wavelength interrogation-based surface plasmon resonance biosensor. Proteomics 4(11):3468-3476.

Zarbl, H. 2001. DNA microarrays: An overview of technologies and applications to toxicology. Curr. Protocol. Toxicol. (Suppl. 9):1.4.2-1.4.16.

Zeeberg, B.R., W. Feng, G. Wang, M.D. Wang, A.T. Fojo, M. Sunshine, S. Narasimhan, D.W. Kane, W.C. Reinhold, S. Lababidi, K.J. Bussey, J. Riss, J.C. Barrett, and J.N. Weinstein. 2003. GoMiner: A resource for biological interpretation of genomic and proteomic data. Genome Biol. 4(4):R28.

Zeiger, E. 2003. Validation and acceptance of new and revised tests: A flexible but transparent process. Toxicol. Lett. (140-141):31-35.

Zhang, K., Z.S. Qin, J.S. Liu, T. Chen, M.S. Waterman, and F. Sun. 2004. Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies. Genome Res. 14(5):908-916.

Zhou, H., J.A. Ranish, J.D. Watts, and R. Aebersold. 2002. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. Nat. Biotechnol. 20(5):512-515.

Zhou, W., G. Liu, D.P. Miller, S.W. Thurston, L.L. Xu, J.C. Wain, T.J. Lynch, L. Su, and D.C. Christiani. 2003. Polymorphisms in the DNA repair genes *XRCC1* and *ERCC2*, smoking, and lung cancer risk. Cancer Epidemiol. Biomarkers Prev. 12(4):359-365.

Zhou, W., J.J. Ryan, and H. Zhou. 2004. Global analyses of sumoylated proteins in *Saccharomyces cerevisiae.* Induction of protein sumoylation by cellular stresses. J. Biol. Chem. 279(31):32262-32268.

Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R.A. Dean, M. Gerstein, and M. Snyder. 2001. Global analysis of protein activities using proteome chips. Science 293(5537):2101-2105.

Zou, M., and S.D. Conzen. 2005. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics 21(1):71-79.

Copyright © National Academy of Sciences. All rights reserved.

Appendix A

Biographic Information on the Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology

David C. Christiani *(Chair)* is professor of occupational medicine and epidemiology at the Harvard School of Public Health and professor of medicine at the Harvard Medical School. He is also a physician in the pulmonary critical care unit at the Massachusetts General Hospital. Dr. Christiani's research interests focus on the molecular epidemiology of acute and chronic diseases, with an emphasis on the development and application of biologic markers for environmentally induced diseases, and gene-environment interactions for malignant and nonmalignant disorders, susceptibility to environmental cancer, and molecular markers of acute lung injury. He is a member of the editorial board of *Environmental Health Perspectives-Toxicogenomics*, *Environmental Health*, *American Journal of Industrial Medicine*, and *International Journal of Environmental and Occupational Health.* Dr. Christiani has served on the Benzene Study Scientific Advisory Committee for the National Cancer Institute. He is board certified in internal medicine, preventive medicine, and pulmonary diseases. Dr. Christiani earned his M.D. from Tufts University and M.P.H. and M.S. from Harvard University.

Cynthia A. Afshari is an associate director in the toxicology department at Amgen, Inc. Her expertise is in the areas of molecular toxicology, functional genomics/toxicogenomics, in vitro models, and carcinogenesis. At Amgen, she leads the investigative and in vitro screening toxicology groups and conducts research in new predictive toxicity assays. She is also responsible for guiding preclinical safety assessment work for several therapeutic project teams. Previously, she was an adjunct professor of toxicology at the University of North Carolina, Chapel Hill, and was the director of basic research applications at the

Copyright © National Academy of Sciences. All rights reserved.

National Institute of Environmental Health Sciences (NIEHS) Microarray Center for 4 years. At NIEHS, she headed an interdisciplinary group of biologists, engineers, and computer scientists investigating applications of new genomic technologies to mechanistic toxicology. Dr. Afshari currently serves as chair of the Steering Committee on Application of Genomics and Proteomics to Mechanism-based Risk Assessment organized by the Health and Environmental Science Institute (HESI) of the International Life Sciences Institute (ILSI) and was chair of the Nephrotoxicity and Database working groups of the same ILSI-HESI Subcommittee for 2 years. She is also a member of the Pharmaceutical Research and Manufacturers of America Genomics Subcommittee and is an associate editor of *Toxicologic Pathology* and *Toxicological Sciences* and a reviewing editor for *Environmental Health Perspectives*. She earned her Ph.D. in toxicology from the University of North Carolina, Chapel Hill, and is a board-certified toxicologist.

John M. Balbus is the director of the health program at Environmental Defense and is also an adjunct associate professor of environmental health at Johns Hopkins University and George Washington University. His expertise is in epidemiology, toxicology, and risk science. He spent 7 years at George Washington University, where he was the founding director of the Center for Risk Science and Public Health and served as acting chair of the Department of Environmental and Occupational Health. He was also an associate professor of medicine. Dr. Balbus serves as a committee member on the Environmental Protection Agency (EPA) Children's Health Protection Advisory Committee and a peer review panel member on the EPA Voluntary Children's Chemical Exposure Program; he was on EPA Science Advisory Board review subcommittees for the National Air Toxics Research Plan and Computational Toxicology. Dr. Balbus earned his M.D. from the University of Pennsylvania and M.P.H. from Johns Hopkins University.

James S. Bus is director of external technology, toxicology and environmental research and consulting at Dow Chemical Company. His research interests include the mechanism of superoxide radical-mediated paraquat toxicity, the relationship between benzene metabolism and toxicity, metabolic pathways as defense mechanisms to toxicant exposure, and mode of action considerations in the use of transgenic animals for mutagenicity and carcinogenicity evaluations. Dr. Bus is involved with the CIIT Centers for Health Research advisory board and the Health and Environmental Sciences Institute of the International Life Sciences Institute; he also chairs a long-range research committee for the American Chemistry Council and serves on the Environmental Protection Agency Science Advisory Board. He is a member of the National Research Council standing Committee on Emerging Issues and Data on Environmental Contaminants. Dr. Bus is a member of several professional societies, including the Society of Toxicology (serving as president, 1996-1997), the American Society for Pharmacology and Experimental Therapeutics, the American Conference of Governmental

Copyright © National Academy of Sciences. All rights reserved.

and Industrial Hygienists, and the Teratology Society; he is a Diplomate of the American Board of Toxicology. He earned his Ph.D. in pharmacology from Michigan State University.

Bruce F. Demple is professor of toxicology in the Department of Genetics and Complex Disease at the Harvard School of Public Health. His research focuses on the impact of environmental exposures, including oxidants, ionizing radiation, and arsenic on in vitro and in vivo systems. Dr. Demple's research interests include occurrence and repair of oxidative, radiation, and alkylation damage to DNA; cellular responses to free radicals; redox-regulated and nitric-oxide-regulated gene expression; and mutagenesis mechanisms and genetic stability. Dr. Demple is a member of the Harvard National Institute of Environmental Health Sciences Center for Environmental Health and serves on the editorial boards of *DNA Repair* and *BMC Cell Biology*. He earned his Ph.D. in biochemistry from the University of California, Berkeley.

Linda E. Greer is the director of the health program for the Natural Resources Defense Council (NRDC). Dr. Greer's expertise is in environmental toxicology and risk assessment and her focus at NRDC is on toxic chemical pollution policy and analysis. Dr. Greer's largest current project is on global sources of mercury pollution. She has served on numerous National Research Council (NRC) committees including the Committee on Industrial Competitiveness and Environmental Protection, the Committee on Ground Water Cleanup Alternatives, and the Committee on Hazardous Wastes in Highway Rights-of-Way. Dr. Greer also served on the NRC's Board on Life Sciences from 2001 to 2004. She is a member of the NRC standing Committee on Emerging Issues and Data on Environmental Contaminants and is a member of the Environmental Protection Agency Science Advisory Board. Dr. Greer earned a M.S.P.H. in environmental science and engineering from the University of North Carolina School of Public Health in Chapel Hill and a Ph.D. in environmental toxicology from the University of Maryland.

Sharon L. R. Kardia is director of the Public Health Genetic Program and associate professor of Epidemiology in the School of Public Health at the University of Michigan. She is also the co-director for the Life Sciences Values and Society Program and the Michigan Center for Genomics and Public Health at the University of Michigan. She has served as a research scientist and professor at the University of Michigan since 1991, where her work in genetic epidemiology focuses on genetics of common chronic diseases, especially cardiovascular disease and hypertension. She has served on the Institute of Medicine Committee on Genomics and the Public's Health in the 21st Century and the State of Michigan's Cancer Genetics Network Advisory Committee. Dr. Kardia earned her M.A. in statistics and Ph.D. in genetics from the University of Michigan.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

George D. Leikauf is a professor at the University of Pittsburgh Graduate School of Public Health. Previously he served as the director of the Molecular Toxicology Division in the Department of Environmental Health at the University of Cincinnati. He was also an adjunct professor in the Pulmonary and Critical Care Division of the Department of Internal Medicine. Dr. Leikauf's research interests include the role of gene-environment interactions in controlling pulmonary response to inhaled pollutants as well as genomic/proteomic techniques in diagnosis and treatment of environmental lung disease. He is associate director of the Center for Environmental Genetics at the University of Cincinnati and has served on numerous committees, including chairing National Institutes of Health Study Sections, and is on several editorial boards. He earned his Ph.D. in environmental health sciences from New York University Medical Center and completed his postdoctoral training at the Cardiovascular Research Institute at the University of California, San Francisco.

Daniel C. Liebler is a professor of biochemistry and pharmacology at Vanderbilt University's School of Medicine and is director of the Proteomics Laboratory in the Mass Spectrometry Research Center. His research interests are in proteomic analysis of protein modifications and proteomic responses to cell injury. He was at the University of Arizona for 16 years, where he was a professor in the Department of Pharmacology and Toxicology. From 1999 to 2003, he served as director of the Southwest Environmental Health Sciences Center at the University of Arizona. He is the associate editor of *Environmental Health Perspectives: Toxicogenomics and Molecular Carcinogenesis*. He also serves on the editorial boards of *Chemical Research in Toxicology*, *Toxicology and Applied Pharmacology*, *Chemical and Biological Interactions*, and the *Journal of Proteome Research*. Dr. Liebler earned his Ph.D. in pharmacology from Vanderbilt University.

Gary E. Marchant is a professor at Arizona State University (ASU) College of Law and executive director of the ASU Center for the Study of Law, Science and Technology. He is also a professor in the School of Life Sciences and the Lincoln Professor of Emerging Technologies, Law and Ethics at ASU. Dr. Marchant teaches and conducts research in environmental law; risk assessment and risk management; genetics and the law; and law, science, and technology. Before joining the ASU faculty in 1999, he was a partner at the Washington, DC, office of the law firm of Kirkland & Ellis, where his practice focused on environmental and administrative law. He currently serves as a member on the National Research Council Committee on State Practices in Setting Mobile Source Emissions Standards. Dr. Marchant earned a Ph.D. in genetics from the University of British Columbia, a Master of Public Policy from the Kennedy School of Government at Harvard University, and a J.D. from Harvard Law School, where he was editor-in-chief of the *Harvard Journal of Law & Technology* and editor of the *Harvard Environmental Law Review*.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

John Quackenbush is a professor in the Department of Biostatistics and Computational Biology, Dana-Farber Institute and Department of Biostatistics, Harvard University School of Public Health. His primary research areas are functional genomics and bioinformatics, and his work focuses on integrating diverse data types to provide insight into biological systems. He and his group have been investigating gene expression patterns in animal models with the goal of identifying mechanisms underlying a range of human diseases. They have also used microarrays to look for diagnostic and prognostic expression fingerprints in human breast and colon cancer, and he has been active in using plant models to develop methods for integrating functional genomic and metabolomic approaches. Dr. Quackenbush is currently a member of the National Research Council standing Committee on Emerging Issues and Data on Environmental Contaminants. He earned a Ph.D. in theoretical particle physics from the University of California, Los Angeles.

Kenneth S. Ramos is a professor and chair of the Department of Biochemistry and Molecular Biology at the University of Louisville Health Sciences Center. He also serves as director of the Center for Genetics and Molecular Medicine. His research focuses on the study of molecular mechanisms of environmental pathogenesis, transcriptional control of retrotransposons, and inference of gene regulatory networks. He is editor of the National Institute of Environmental Health Sciences journal *Environmental Health Perspectives: Toxicogenomics*. Dr. Ramos has served on numerous National Research Council committees, including the Committee for a Review of Evidence Regarding the Link Between Exposure to Agent Orange and Diabetes, Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides: Second Biennial Update, HHMI Predoctoral Fellowships Panel on Neurosciences and Physiology, and Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides: First Biennial Update. He serves as chair of the Committee on Emerging Issues and Data on Environmental Contaminants. Dr. Ramos received a Ph.D. in biochemical pharmacology and toxicology from the University of Texas at Austin.

Mark A. Rothstein is the Herbert F. Boehl Chair of Law and Medicine and is director of the Institute for Bioethics, Health Policy and Law at the University of Louisville. He has appointments in the Departments of Medicine and Family and Geriatric Medicine at the School of Medicine and at the Louis D. Brandeis School of Law. He earned a J.D. from Georgetown University. Mr. Rothstein's interests include the ethical, legal, and social implications of genetics, privacy, health policy, and employment law. He is chair of the Privacy and Confidentiality Subcommittee of the National Committee on Vital and Health Statistics, the federal advisory committee that advises the Secretary of Health and Human Services on health information policy, including the privacy regulations of the Health Insurance Portability and Accountability Act. He has served on the Institute of Medicine Committee on Assessing Genetic Risks: Issues and Implica-

Copyright © National Academy of Sciences. All rights reserved.

tions for Health and is a member of the National Research Council Committee on Emerging Issues and Data on Environmental Contaminants.

Raymond E. Stoll is president of Stoll and Associates, LLC, a toxicology consulting company specializing in toxicologic, pharmaceutical, and biological issues in compound development. His areas of interest are computers in toxicology, carcinogenesis, alternative models in carcinogenicity assessment, and toxicogenomics and proteomics. Dr. Stoll was director of toxicology and safety assessment at Boehringer Ingelheim Pharmaceuticals from 1992 to 2003 and was a member of their product development committee. Currently, he holds the title of Highly Distinguished Scientist in Toxicology at Boehringer Ingelheim Pharmaceuticals. Before working at Boehringer Ingelheim Pharmaceuticals, he served as senior director of toxicology at Cetus Corporation and as director of toxicology and pathology at the Sandoz Research Institute. He is a past chair of the Pharmaceutical Research and Manufacturers of America preclinical safety committee. He earned his Ph.D. in pharmacology and toxicology from Purdue University.

Roger G. Ulrich is currently chief development officer at Calistoga Pharmaceuticals, Inc. Previously he was senior scientific director with Rosetta Inpharmatics, LLC, Merck & Company, Inc. located in Seattle, Washington. He earned his B.S. and M.S. degrees in biology, chemistry, and biomedical sciences from Western Michigan University and his Ph.D. in cell and molecular biology from West Virginia University. He joined the pathology and toxicology research department at The Upjohn Company in 1983 (currently part of Pfizer) as a research scientist in ultrastructural pathology and cell biology, and was a Diplomate of the American Board of Toxicology from 1993 to 2003. He worked for Abbott Laboratories (Abbott Park, Illinois) from 1997 to 2001 as director of toxicology and director of cellular and molecular toxicology and joined Merck-Rosetta in 2001 as senior director. In his 22-year industrial career, he has contributed to the discovery and development of new drugs in a variety of therapeutic areas, including diabetes, cancer, and infectious diseases; he pioneered the application of gene expression profiling to toxicology. He currently serves on the editorial boards of *Xenobiotica and Toxicology In Vitro* with past service on other editorial boards. He is an active member of several scientific societies and associations, including the American Association for the Study of Liver Diseases, the Health and Environmental Science Institute of the International Life Sciences Institute, and the Pharmaceutical Research and Manufacturers of America Genomics Division. He continues to hold academic appointments. He has authored more than 100 full-length publications in cell biology and toxicology and pursues a research focus on cellular and molecular pharmacology and toxicology.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Helmut Zarbl is a principal investigator in the Department of Environmental and Occupational Medicine at the University of Medicine and Dentistry of New Jersey. Prior to this he was a full member in the Divisions of Human Biology and Public Health Sciences at the Fred Hutchinson Cancer Research Center (FHRCR). He was the director and a principal investigator for the National Institute of Environmental Health Sciences (NIEHS)-sponsored FHFRC/University of Washington Toxicogenomics Research Consortium. His areas of expertise are in DNA microarray technology development and functional genomics and toxicogenomics, in particular understanding tissue, strain, species, and intraindividual differences in response to environmental toxicants through gene expression analysis. He is currently an affiliate professor at the University of Washington (UW) in the School of Medicine (pathology) and the School of Public Health and Community Medicine (toxicology). He is also a member of the UW/NIEHS Center for Ecogenetics and Environmental Health, where he is the director of Functional Genomics Core. Previously, he was an associate professor at MIT He earned his Ph.D. in biochemistry from McGill University.

Copyright © National Academy of Sciences. All rights reserved.

Appendix B

Minimum Information About a Microarray Experiment for Toxicogenomics[1]

With the development of toxicogenomics we believe that it is necessary to move toward harmonization of minimal toxicological data requirements to fully realise the potential of this emerging interdisciplinary field.

Following the very favourable response that the Minimum Information About a Microarray Experiment (MIAME) (Brazma et al. 2001) has received from the microarray community and key scientific journals (Nature 2002; Ball et al. 2002a,b,c) we have initiated the harmonization process as applied to array-based toxicogenomic experiments. MIAME/Tox (MIAME 2002) is a set of guidelines defining the minimum information required to interpret unambiguously and potentially reproduce and verify array-based toxicogenomic experiments. Similarly, MIAME/Tox seeks to provide such a conceptual structure in the context of pharmacogenomics and chemogenomics. Therefore, this harmonization effort in toxicogenomics will have broad application in experimental science as well as clinical medicine.

MIAME/Tox supports a number of different objectives, for example: linking data within a study, and linking several studies from one institution and exchanging toxicogenomics datasets among public databases. In fact, the major objective of MIAME/Tox is to guide the development of toxicogenomics databases and data management software. The breadth, depth, and uniformity of the information a database contains are critical to its utility. To address the last issue, MIAME/Tox content areas for experiment descriptions include information that are recommended to be provided by maximum use of controlled vocabularies or ontologies (such as species taxonomy, cell types, anatomy terms, histopa-

[1]Source: Based on MIAME/Tox 1.1, Draft, August 2003, from MIAME 2002. Reprinted with permission from authors; 2002, Microarray Gene Expression Data Society.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

thology, toxicology, and chemical compound nomenclature). The use of controlled vocabularies is needed to enable database queries and automated data analysis.

MIAME/Tox guidelines have been adopted to guide the development of toxicogenomics databases underway at the NIEHS National Center for Toxicogenomics, USA (NCT 2006), and at the EMBL European Bioinformatics Institute, UK (EMBL EBI 2007), in conjunction with the International Life Sciences Institute's Health and Environmental Sciences Institute, USA (ILSI HESI 2007).

MIAME/Tox is continuously developing in accordance with our understanding of microarray technology and its applications to toxicology and pharmacology. Please join the MIAME/Tox discussion list (mged-tox@lists.source forge.net) and contribute with your ideas and comments.

MIAME/TOX CHECKLIST

Minimum information to be recorded about toxicogenomics experiments is defined in subsequent sections and should include the following data domains:

- Agent description, formulation, purity, solubility, vehicle, separation methods, chemical structure, active moieties, safety and toxicity, storage, half-life.
- Experimental design parameters, genetic background and animal husbandry information or cell line and culture information, exposure parameters, dosing regimen, and dose groups.
- Microarray data, specifying the number and details of replicate array bioassays associated with particular samples, and including PCR transcript analysis if available.
- Biological endpoint data, including animal and organ weights at necropsy or cell counts and doubling times, clinical chemistry and enzyme assays, hematology, urinalysis, other.
- Textual endpoint information such as clinical and gross observations, pathology and microscopy findings.

As with MIAME, MIAME/Tox has two major sections.

- Array design description;
- Gene expression experiment description.

The first section remains identical to the MIAME 1.1 document, and the second section is extended to fulfill the need of this toxicogenomics-specific application of MIAME. The two components of MIAME/Tox are discussed in further detail below.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Array Design Description

This section remains identical to the MIAME 1.1 document.

Experiment Description

By *experiment*, MIAME refers to a set of one or more hybridizations that are in some way related (e.g., related to the same publication or the same study). The minimum information for a toxicogenomic experiment includes a description of the following five parts.

1. Toxicogenomic experimental design
2. Biological materials used, extract preparation and labeling, toxicological assays.
3. Hybridization procedures and parameters
4. Gene expression measurement data and Specifications of data processing

MIAME/Tox recommends the following details on each of these sections.

1. Toxicogenomic Experimental Design

The following information is included in the experimental design.

- Authors, laboratory or clinic, contact
- A brief description of the experiment and its goal and a link to a publication if one exists (Links [URL], citations).
- Indicate the Experiment Design Type

Instances could be:
- compound_treatment_design
- dose_response_design
- injury_design
- stimulus_or_stress_design
- other.

Note that use of multiple entries if of course possible to specify the type of the experiment. One can also propose a term: e.g. Acute, pre-chronic or chronic treatment, or clinical trial

- Experimental factors, i.e. organisms, parameters or conditions tested, for instance:
  - species, strain, genotype, genetic variation
  - age and weight, developmental stage
  - dose(s) in standard units

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

  - route of exposure, vehicle, time of treatments and observations
- Indicate the total number of hybridizations in the experiment
- Quality control steps taken:
  - Replicates done (yes/no), type of replicates (biological, technical) description
  - if pools of extracts (yes/no) were used versus extracts from individual samples, description
  - whether dye swap is used (only for two channel platforms)
  - other.

2. Biological Materials Used, Extract Preparation and Labeling, Toxicological Assays

By *biological material*, MIAME/Tox refers to the material (sample) used in toxicological, pharmacological or clinical investigations and from which nucleic acids were extracted for subsequent labelling and hybridisation. In this section all steps that precede the hybridization are described. We can usually distinguish between:

- Assessment of the source of the sample (biosource properties);
- Treatments applied to the samples (manipulations);
- Toxicological assessments;
- Extract preparation;
- Extract labelling; and
- Hybridization controls.

Biosource properties

- organism (NCBI taxonomy)
- sample source provider
- descriptors relevant to the particular sample, such as
  - strain
  - sex
  - genetic background
  - genetic modifications
  - age
  - weights
  - development stage
  - organism part (tissue) of the organism's anatomy from which the biological material is derived (if samples are cells)
  - cell type
  - animal/plant strain or line
  - genetic variation (e.g., gene knockout, transgenic variation)

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

- individual genetic characteristics (e.g., disease alleles, polymorphisms)
- disease state or normal
- additional clinical information available an individual identifier (for interrelation of the biological materials in the experiment)

Sample manipulations: laboratory protocols and relevant parameters, such as:

- facilities details
- animal husbandry and housing details
- cell culture conditions
- growth conditions (passage level and frequency)
- metabolic competency of cell strains
- treatment (stressor), in vivo, in vitro
- treatment type (e.g., compound, small molecule, heat shock, cold shock, food deprivation, diet)
- treatment compound name and grade formulation, including manufacturer
- type of compound (e.g. chemical, drug or solvent)
- CASRN, chemical structure/molecular formula
- vehicle for chemical treatment
- exposure method (route of administration, e.g. oral, gavage, mucolar, medium, intraperitoneal, intramuscular, intravenous, topical)
- duration
- dose (and unit)
- date/time at death or at sacrifice
- sacrifice method

Toxicological assessments: Laboratory Protocols and Relevant Parameters Measured and Data Files e.g.,

Clinical Observations

- weight
- survival (yes/no)
- signs (e.g., general, behavior)
- site of application
- lesions
- color effects
- other

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

Gross Necropsy Examination

- organs and tissues examination list
- organs and tissues collection list
- organs and tissues weight list
- organs and tissues storage method and location

Histopathology Evaluation

- which biological materials (control and experimental)
- slide preparation, storage method and location
- topography (definite anatomical region)
- system
- organ
- sites
- cell type(s)
- morphology(s)
- qualifier(s) for the morphology(s)

Clinical Pathology

- hematology (e.g., erythrocyte count, mean corpuscular volume, hemoglobin)
- clinical chemistry (e.g.,sorbitol dehydrogenase (SDH), alkaline phosphatase (ALP), creatine kinase (CK)
- other parameters measured, e.g., sperm morphology and vaginal cytology evaluation (SMVCE)
- estrous cycle length
- micronucleated erythrocytes determination
- functional observation battery
- other.

Nucleic Acid Extraction Protocol Applied to the Biological Material

- Type of nucleic acid RNA, mRNA, or genomic DNA is extracted
- extraction method
- amplification methods if any.

Labeling Protocol for Each Labeling Prepared from the Extract, Including

- amount of nucleic acids labeled
- label used (e.g., A-Cy3, G-Cy5, 33P, ….)

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

- label incorporation method
- Facility details (if this part of the experiments has been carried out in facility different from the sample treatment and toxicological assessments steps above, e.g. consortium, contracting out).

External Controls Added to Hybridization Extract(s) (spiking controls)

- element on array expected to hybridize to spiking control
- spike type (e.g., oligonucleotide, plasmid DNA, transcript)
- spike qualifier (e.g., concentration, expected ratio, labelling methods if different than that of the extract)

3. Hybridization Procedures and Parameters

This section remains identical to the MIAME 1.1 document

4. Measurement Data and Specifications of Data Processing

This section remains identical to the MIAME 1.1 document

On the Behalf of EMBL-EBI, NIEHS NCT, NIEHS NTP and ILSI HESI

Susanna-Assunta Sansone1, EMBL—The European Bioinformatics Institute; Michael D. Waters, NIEHS National Center for Toxicogenomics; Syril D. Pettit, ILSI Health and Environmental Sciences Institute; William B. Mattes, ILSI Health and Environmental Sciences Institute, Genomics Committee; Jennifer Fostel, Pfizer and ILSI Health and Environmental Sciences Institute, Genomics Committee; William Pennie, NIEHS National Toxicology Program; Alvis Brazma, EMBL—The European Bioinformatics Institute; Raymond W. Tennant, NIEHS National Center for Toxicogenomics; James K. Selkirk, NIEHS National Center for Toxicogenomics; Richard S. Paules, NIEHS National Center for Toxicogenomics; Pierre R. Bushel, NIEHS National Center for Toxicogenomics; William C. Eastin, NIEHS National Toxicology Program.

REFERENCES

Ball, C.A, G. Sherlock, H. Parkinson, P. Rocca-Sera, C. Brooksbank, H.C. Causton, D. Cavalieri, T. Gaasterland, P. Hingamp, F. Holstege, M. Ringwald, P. Spellman, C.J. Stoeckert, Jr., J.E. Stewart, R. Taylor, A. Brazma, and J. Quackenbush. 2002a. Standards for microarray data [letter]. Science 298(5593):539.

Ball, C.A., G. Sherlock, H. Parkinson, P. Rocca-Sera, C. Brooksbank, H.C. Causton, D. Cavalieri, T. Gaasterland, P. Hingamp, F. Holstege, M. Ringwald, P. Spellman, C.J. Stoeckert, Jr., J.E. Stewart, R. Taylor, A. Brazma, and J. Quackenbush. 2002b.

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

The underlying principles of scientific publication [review]. Bioinformatics 18(11):1409.

Ball, C.A., G. Sherlock, H. Parkinson, P. Rocca-Sera, C. Brooksbank, H.C. Causton, D. Cavalieri, T. Gaasterland, P. Hingamp, F. Holstege, M. Ringwald, P. Spellman, C.J. Stoeckert, Jr., J.E. Stewart, R. Taylor, A. Brazma, and J. Quackenbush. 2002c. A guide to microarray experiments-an open letter to the scientific journals. The Lancet 360(9338):1019.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, T. Gaasterland, P. Glenisson, F.C. Holstege, I.F. Kim, V. Markowitz, J.C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. 29(4):365-371.

EMBL EBI (European Bioinformatics Institute). 2007. NET Project at the EBI. European Bioinformatics Institute [online]. Available: http://www.ebi.ac.uk/net-project/index.html [accessed April 26, 2007].

ILSI HESI. 2007. International Life Sciences Institute- Health and Environmental Sciences Institute [online]. Available: http://www.hesiglobal.org/ [accessed April 26, 2007].

MIAME. 2002. Minimum Information About a Microarray Experiment—MIAME 1.1. DRAFT 6. Version 1.1. April 1, 2002 [online]. Available: http://www.mged.org/Workgroups/MIAME/miame_1.1.html [accessed April 26, 2007].

MIAME. 2003. A MIAME for Toxicogenomics-MIAME/TOX. Draft August 2003-Based on MIAME/Tox 1.1. [online]. Available: http://www.mged.org/Workgroups/rsbi/MIAME-Tox-Checklist.doc [accessed April 26, 2007].

Nature (Nature Journal). 2002. Microarray standards at last [editorial]. Nature 419(6905):323.

NCT (National Center for Toxicogenomics). 2006. Using Global Genomic Expression Technology to Create a Knowledge Base for Protecting Human Health. National Center for Toxicogenomics, National Institute of Environmental Health Sciences [online]. Available: http://www.niehs.nih.gov/nct/ [accessed April 26, 2007].

Copyright © National Academy of Sciences. All rights reserved.

Appendix C

Overview of Risk Assessment

The objective of chemical risk assessment methodologies is to facilitate both scientific and data-informed decision making and also is increasingly expected to provide more accurate predictions of actual risk. The validity of a prediction of risk derived from a risk assessment depends largely on the quality and accuracy of data. Where data do not exist or are contradictory, regulatory agencies are forced to rely on default values, uncertainty factors, and modeling approaches to fill in the blanks. These defaults and extrapolations introduce uncertainty into the risk estimates. New methodologies and testing methods could fill key data gaps, clarify data inconsistencies, or otherwise reduce uncertainty. If applied appropriately, these approaches have the potential to improve the accuracy and scientific credibility of regulatory decision making.

An Overview of Current Risk Assessment Practice

Government agencies charged with protecting public and worker health are required to review, quantify, and ultimately regulate chemicals, physical agents, and pharmaceuticals in a manner that will protect and enhance the public health and the environment. One of these regulatory responsibilities is to assess the risk to human health from chemical exposures. This section provides a brief overview of the aspects of regulatory risk assessment practices most relevant to toxicogenomic technologies. Further details about the risk-assessment process can be found in several references (EPA 1986, 2004, 2005; PCCRARM 1997).

Human health risk assessment is the process of analyzing information to determine whether an environmental hazard might cause harm to exposed persons (EPA 2004). The risk-assessment process integrates many disciplines of toxicology. It has both qualitative and quantitative components and consists of

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

four general steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization (NRC 1983, 1994).

Step 1: **Hazard identification** entails identifying the contaminants that are suspected to pose health hazards, quantifying the concentrations at which they are present in the environment, describing the specific forms of toxicity (neurotoxicity, carcinogenicity, etc.) that the contaminants of concern can cause, and evaluating the conditions under which these forms of toxicity might be expressed in exposed humans.

Step 2: **Dose-response assessment** entails further evaluating the conditions under which the toxic properties of a chemical might be manifested in exposed people, with particular emphasis on the quantitative relation between the dose and the toxic response. The development of this relationship may involve the use of mathematical models. This step may include an assessment of variations in response—for example, differences in susceptibility between young and old people.

Step 3: **Exposure assessment** involves specifying the population that might be exposed to the agent of concern, identifying the routes through which exposure can occur, and estimating the magnitude, duration, and timing of the doses that people might receive as a result of their exposure.

Step 4: **Risk characterization** involves integrating information from the first three steps to develop a qualitative or quantitative estimate of the likelihood that any of the hazards associated with the agent of concern will be realized in exposed people. This is the step in which risk-assessment results are expressed. Risk characterization should also include a full discussion of the uncertainties associated with the estimates of risk (Adapted from NRC 1994).

Toxicogenomic information has a potential role in all aspects of the risk-assessment process. For example, in hazard identification, toxicogenomic data could inform the types of hazard a chemical presents (for example, whether it poses cancer or noncancer risks) and the modes and mechanisms of toxic action[1]

---

[1]The EPA (EPA 2005) provides the following definitions: "The term 'mode of action' is defined as a sequence of key events and processes, starting with interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in cancer formation. A 'key event' is an empirically observable precursor step that is itself a necessary element of the mode of action or is a biologically based marker for such an element. Mode of action is contrasted with 'mechanism of action,' which implies a more detailed understanding and description of events, often at the molecular level, than is meant by mode of action. The toxicokinetic processes that lead to formation or distribution of the active agent to the target tissue are considered in estimating dose but are not part of the mode of action as the term is used here. There are many examples of possible

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

through which it acts. Information on the mode of action is also a component in deciding the appropriate approach to dose-response assessment (described further below). Toxicogenomic approaches could support exposure assessment by indicating cellular exposure to toxicants. Toxicogenomic data may also be used to better understand areas of uncertainty, including variability in the human population, extrapolation of data from one species to another, identification of susceptible subpopulations, and provision of quantitative data to improve risk assessments.

Quantification of Risk

Different analytical techniques are used in cancer and noncancer risk assessments to quantify risk. The end result of a noncancer risk assessment can be the determination of a quantitative human reference dose (RfD) for oral exposures or a reference concentration (RfC) for inhalation exposures (both are generically referred to as a "reference value").[2]

As summarized by the U.S. Environmental Protection Agency (EPA) (EPA 2004): When developing a noncancer reference value (a RfD or RfC) for a chemical substance, EPA surveys the scientific literature and selects a critical study and a critical effect. The critical effect is defined as the adverse effect, or its known precursor, that occurs at the lowest dose identified in the most sensitive species as the dose rate of an agent increases. When a no-observed-adverse-effect level (NOAEL)[3] can be identified in a critical study, it becomes the basis for the reference value derivation. If NOAEL cannot be identified, then a lowest-observed-adverse-effect level (LOAEL)[4] is identified instead. Recently, benchmark doses (BMDs) (EPA 2000) from the modeling of dose-response data

---

modes of carcinogenic action, such as mutagenicity, mitogenesis, inhibition of cell death, cytotoxicity with reparative cell proliferation, and immune suppression."

[2]Reference concentration (RfC): An estimate (with uncertainty spanning perhaps an order of magnitude) of a continuous inhalation exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime. Reference dose (RfD): An estimate (with uncertainty spanning perhaps an order of magnitude) of a daily oral exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime. (Source: EPA 2007)

[3]The highest exposure level at which there are no statistically or biologically significant increases in the frequency or severity of adverse effects between the exposed population and its appropriate control is called the no-observed-adverse-effect level (NOAEL) (EPA 2004).

[4]"A LOAEL is the lowest exposure level at which there are biologically significant increases (with or without statistical significance) in frequency or severity of adverse effects between the exposed population and its appropriate control group. The NOAEL is generally presumed to lie between zero and the LOAEL, so an UF [uncertainty factor] (generally 10 but sometimes 3 or 1) is applied to the LOAEL to derive a nominal NOAEL" (EPA 2004).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

have been used instead of the traditional NOAEL/LOAEL approach; however, most RfDs are based on NOAELs (EPA 2004). The NOAEL, LOAEL, or BMD is divided by appropriate uncertainty factor to derive the final reference value. The uncertainty factors are generally 10-fold (but can be higher or lower if informed by data) and are intended to account for uncertainty in the available data from (1) variation in the human population, (2) extrapolating animal data to humans, (3) extrapolating from less-than-lifetime exposures to lifetime exposure, (4) extrapolating from a LOAEL rather than from a NOAEL, and (5) using incomplete databases.[5] For carcinogenic compounds, data from human epidemiologic studies are preferred, but, in the absence of human epidemiologic data, animal data are used.[6] Dose-response curves are constructed from these studies; however, the range of doses is frequently above the levels of environmental interest. To estimate the risks below the levels tested, the observed data are used to derive a point of departure[7] followed by extrapolation to lower exposures (EPA 2005). Linear or nonlinear approaches can be used to extrapolate to low doses, and the choice of methods is critical because the derived risk estimates vary by technique. In general, linear approaches produce more conservative risk estimates than nonlinear approaches (NRC 2006).

The selection of the various models used to extrapolate to low doses is informed by a compound's mode of action. The EPA cancer guidance (EPA 2005) states that "when available data are insufficient to establish the mode of action for a tumor site and when scientifically plausible based on the available data, linear extrapolation is used as a default approach."[8] Further, "A nonlinear approach should be selected when there are sufficient data to ascertain the mode of action and conclude that it is not linear at low doses and the agent does not demonstrate mutagenic or other activity consistent with linearity at low doses." For

[5]In some cases, the largest divisor the EPA will use is 3,000 because of the uncertainty when so many uncertainty factors are applied (for example, see the risk assessment for trichloroethylene (EPA 2001).

[6]As described by the EPA (2005): "When animal studies are the basis of the analysis, the estimation of a human-equivalent dose should utilize toxicokinetic data to inform cross-species dose scaling if appropriate and if adequate data are available. Otherwise, default procedures [described the EPA 2005] should be applied."

[7]As described by the EPA (2005): "A 'point of departure' (POD) marks the beginning of extrapolation to lower doses. The POD is an estimated dose (usually expressed in human-equivalent terms) near the lower end of the observed range, without significant extrapolation to lower doses."

[8]The 2005 Cancer Guidance also states: "Linear extrapolation should be used when there are MOA [mode of action] data to indicate that the dose-response curve is expected to have a linear component below the POD [point of departure]. Agents that are generally considered to be linear in this region include: agents that are DNA-reactive and have direct mutagenic activity, or agents for which human exposures or body burdens are high and near doses associated with key precursor events in the carcinogenic process, so that background exposures to this and other agents operating through a common mode of action are in the increasing, approximately linear, portion of the dose-response curve."

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

cancer risk assessments, extrapolations from the point of departure can be used to calculate a cancer slope factor[9] (for linear extrapolation) and a RfD or RfC among other outputs (for nonlinear extrapolation) (EPA 2005).

The values derived from the cancer and noncancer risk assessments are used to protect the public from unacceptable chemical exposures and can be the basis of regulatory decision making (for example, in establishing standards for water and air quality and requirements for environmental cleanup). These values have a range of implications to stakeholders and, because of the uncertainty inherent in the risk assessment process, their derivation can be quite controversial (e.g., NRC 1999, 2001, 2005, 2006). As a result, tools such as toxicogenomics that can be used in the risk-assessment process to increase the certainty of risk estimates are of great importance for protecting public health.

REFERENCES

EPA (U.S. Environmental Protection Agency). 1986. Guidelines for Carcinogen Risk Assessment. EPA/630/R-00/004. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.

EPA (U.S. Environmental Protection Agency). 2000. Benchmark Dose Technical Guidance Document. External Review Draft. EPA/630/R-00/001. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. October 2000 [online]. Available: http://www.epa.gov/nceawww1/pdfs/bmds/BMD-External_10_13_2000.pdf [accessed April 13, 2007].

EPA (U.S. Environmental Protection Agency). 2001. Trichloroethylene Health Risk Assessment: Synthesis and Characterization. External Review Draft. EPA/600/P-01/002A. Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC. August 2001.

EPA (U.S. Environmental Protection Agency). 2004. An Examination of EPA Risk Assessment Principles and Practices. Staff Paper. Prepared for the U.S. Environmental Protection Agency by Members of the Risk Assessment Task Force. EPA/100/B-04/001. Office of the Science Advisor, U.S. Environmental Protection Agency, Washington, DC [online]. Available: http://www.epa.gov/OSA/pdfs/ratf-final.pdf [accessed Jan.13, 2006].

EPA (U.S. Environmental Protection Agency). 2005. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. March 2005 [online]. Available: http://www.epa.gov/iriswebp/iris/cancer032505.pdf .[accessed April 13, 2007].

EPA (U.S. Environmental Protection Agency). 2007. Glossary of IRIS Terms. Integrated Risk Information System, U.S. Environmental Protection Agency [online]. Available: http://www.epa.gov/iris/gloss8.htm [accessed April 13, 2007].

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press.

[9]An upper bound, approximating a 95% confidence limit, on the increased cancer risk from a lifetime exposure to an agent. This estimate is usually expressed in units of proportion (of a population) affected per mg/kg-day (Adapted from EPA 2007).

Copyright © National Academy of Sciences. All rights reserved.

Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment
http://www.nap.edu/catalog/12037.html

NRC (National Research Council). 1994. Science and Judgment in Risk Assessment. Washington, DC: National Academy Press.

NRC (National Research Council). 1999. Arsenic in Drinking Water. Washington, DC: National Academy Press.

NRC (National Research Council). 2001. Arsenic in Drinking Water: 2001 Update. Washington, DC: National Academy Press.

NRC (National Research Council). 2005. Health Implications of Perchlorate Ingestion. Washington, DC: National Academy Press.

NRC (National Research Council). 2006. Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues. Washington, DC: The National Academies Press.

PCCRARM (Presidential/Congressional Commission on Risk Assessment and Risk Management). 1997. Framework for Environmental Health Risk Management. Vol. 1. Washington, DC: The Commission.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.