Emerging Safety Science: Workshop Summary

Sally Robinson, Robert Pool, and Robert Giffin, Forum on Drug Discovery, Development, and Translation

ISBN: 0-309-11013-0, 150 pages, 6 x 9, (2008)

EMERGING SAFETY SCIENCE

WORKSHOP SUMMARY

Sally Robinson, Robert Pool, and Robert Giffin

Forum on Drug Discovery, Development, and Translation

Board on Health Sciences Policy

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

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

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. 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.

Support for this project was provided by the Department of Health and Human Services (Contract Nos. N01-OD-4-2139 and 223-01-2460); American Diabetes Association; American Society for Microbiology; Amgen, Inc.; Association of American Medical Colleges; AstraZeneca Pharmaceuticals; Blue Cross Blue Shield Association; Burroughs Wellcome Fund; Doris Duke Charitable Foundation; Eli Lilly and Company; Entelos, Inc.; GlaxoSmithKline; March of Dimes Foundation; Merck & Co., Inc.; Pfizer, Inc.; and UnitedHealth Group. 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: 978-0-309-11012-9
International Standard Book Number-10: 0-309-11012-2

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu.**

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Suggested citation: IOM (Institute of Medicine). 2008. *Emerging safety science: Workshop summary.* Washington, DC: The National Academies Press.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—Goethe

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

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

PLANNING COMMITTEE FOR THE WORKSHOP ON EMERGING SAFETY SCIENCE

Naomi Aronson, Blue Cross Blue Shield Association, Illinois
C. Thomas Caskey, University of Texas-Houston Health Science Center
Gail H. Cassell, Eli Lilly and Company, Indiana
Mikhail Gishizky, Entelos, Inc., California
Edward W. Holmes, A*Star Biomedical Research Council, Singapore
William Keane, Merck & Co., Inc., Pennsylvania
Ronald L. Krall, GlaxoSmithKline, Pennsylvania
Janet Woodcock, U.S. Food and Drug Administration, Maryland

IOM Staff

Robert B. Giffin, Director
Sally Robinson, Program Officer
Andrea Knutsen, Senior Program Assistant
Rona Briere, Consulting Editor

FORUM ON DRUG DISCOVERY, DEVELOPMENT, AND TRANSLATION

Gail H. Cassell (*Co-Chair*), Eli Lilly and Company, Indiana
Edward W. Holmes (*Co-Chair*), A*Star Biomedical Research Council, Singapore
Naomi Aronson, Blue Cross Blue Shield Association, Illinois
Leslie Z. Benet, University of California-San Francisco
Catherine Bonuccelli, AstraZeneca Pharmaceuticals, Delaware
Linda Brady, National Institute of Mental Health, Maryland
Robert M. Califf, Duke University Medical Center, North Carolina
Scott Campbell, American Diabetes Association, Virginia
C. Thomas Caskey, University of Texas-Houston Health Science Center
Francis D. Chesley, Jr., Agency for Healthcare Research and Quality, Maryland
Peter B. Corr, Pfizer Inc. (retired), New York
James H. Doroshow, National Cancer Institute, Maryland
Jeffrey M. Drazen, *New England Journal of Medicine*, Massachusetts
Joseph M. Feczko, Pfizer, Inc., New York
Garret A. FitzGerald, University of Pennsylvania School of Medicine
Elaine K. Gallin, Doris Duke Charitable Foundation, New York
Steven K. Galson, U.S. Food and Drug Administration, Maryland
Alan M. Garber, Stanford University, California
Mikhail Gishizky, Entelos, Inc., California
Stephen Groft, National Institutes of Health, Maryland
Peter K. Honig, Merck & Co., Inc., Pennsylvania
Richard Justman, UnitedHealth Group, Minnesota
Michael Katz, March of Dimes Foundation, New York
Ronald L. Krall, GlaxoSmithKline, Pennsylvania
Jeffrey M. Leiden, Clarus Ventures, Massachusetts
John R. Marler, National Institute of Neurological Disorders and Stroke, Maryland
Musa Mayer, AdvancedBC.org, New York
Mark B. McClellan, AEI-Brookings Joint Center for Regulatory Studies, District of Columbia
Joshua J. Ofman, Amgen, Inc., California
Suzanne Pattee, Cystic Fibrosis Foundation, Maryland
Joanne L. Rhoads, National Institute of Allergy and Infectious Diseases (retired), Maryland
B. A. Schwetz, U.S. Department of Health and Human Services, Maryland

Janet Shoemaker, American Society for Microbiology, District of Columbia
Lana Skirboll, National Institutes of Health, Maryland
Nancy S. Sung, Burroughs Wellcome Fund, North Carolina
Jorge A. Tavel, National Institute of Allergy and Infectious Diseases, Maryland
Janet Woodcock, U.S. Food and Drug Administration, Maryland

Reviewers

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 wish to thank the following individuals for their review of this report:

C. Thomas Caskey, University of Texas Health Science Center at Houston
Joseph M. Feczko, Pfizer, Inc.
Stephen Groft, Office of Rare Disease Research, National Institutes of Health
Patrice M. Milos, Helicos BioSciences Corporation
Joanne L. Rhoads, National Institute of Allergy and Infectious Diseases (retired)

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the final draft of the report before its release. The review of this report was

overseen by **Hellen Gelband,** Scholar-in-Residence, Institute of Medicine. Appointed by the Institute of Medicine, she was 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 authors and the institution.

SPECIAL ACKNOWLEDGMENTS

In addition to our reviewers, we would like to especially thank members of the U.S. Food and Drug Administration (FDA) staff who provided assistance in the development of the workshop concept and agenda. They include Daryl Allis, Rachel Behrman, Howard Chazin, Lois Chester, Charles Cooper, Gerald Dal Pan, Felix Frueh, Steven Galson, Federico Goodsaid, David Jacobson-Kram, John Jenkins, Jan Johannessen, Elizabeth Mansfield, Robert Meyer, Shirley Murphy, Judith A. Racoosin, George Rochester, Wendy Sanhai, Paul Seligman, Judy Staffa, Ana Szarfman, Ellis Unger, Keith Webber, and Janet Woodcock.

Preface

The genomic age of medicine and advancements in molecular medicine, bioinformatics, and information technology have equipped scientists with powerful new technologies that can be used to develop safer drugs and to monitor drugs more proficiently once they are on the market. These technologies have the potential to identify safety issues much earlier in the development process, reducing the number of expensive clinical trials, leading to more promising research avenues, and decreasing the exposure of human subjects and patients to products with safety problems. Furthermore, incorporating knowledge of these technologies in the U.S. Food and Drug Administration's (FDA's) review process can lead to more effective drug safety assessments and accelerate the drug approval process.

To explore the application of these innovative technologies to the assessment of drug safety in both the pre- and postmarket environments, the Institute of Medicine's Forum on Drug Discovery, Development, and Translation convened a workshop on Emerging Safety Science. A diverse group of experts from academia, industry, and government examined two broad approaches to improving drug safety: basic scientific approaches (genomics, metabolomics, pharmacogenomics, cell-based signaling, and standard toxicology) to better identify safety issues during development; and innovative techniques for collecting and analyzing postmarket data to identify safety signals more rapidly than is possible with traditional methodologies. The workshop presentations and discussions shed new light on the potential of these technologies to enhance the assessment of

safety, and provided important insights into the future challenges and opportunities in the promising field of safety science.

It is our hope that this workshop summary will serve as a resource enabling scientists to survey cutting-edge technologies being applied in the field of safety science and to consider ways of applying these technologies in their own work. The Forum remains committed to fostering an environment in which diverse groups of stakeholders can come together in a neutral setting to share their experiences, with the hope of furthering the advancement of drug discovery, development, and translation.

Edward W. Holmes, *Co-Chair*
Janet Woodcock, *Member*
Forum on Drug Discovery,
Development, and Translation

Contents

Tables and Figures

TABLES

FIGURES

1

Introduction

In recent years, the costs of new drug development have skyrocketed. The average cost of developing a new approved drug is now estimated to be $1.3 billion (DiMasi and Grabowski, 2007). At the same time, each year fewer new molecular entities (NMEs) are approved. DiMasi and Grabowski report that only 21.5 percent of the candidate drugs that enter phase I clinical testing actually make it to market. In 2007, just 17 novel drugs and 2 novel biologics were approved. In addition to the slowing rate of drug development and approval, recent years have seen a number of drugs withdrawn from the market for safety reasons. According to the Government Accountability Office (GAO), 10 drugs were withdrawn because of safety concerns between 2000 and March 2006 (GAO, 2006). Finding ways to select successful drug candidates earlier in development could save millions or even billions of dollars, reduce the costs of drugs on the market, and increase the number of new drugs with improved safety profiles that are available to patients.

Emerging scientific knowledge and technologies hold the potential to enhance correct decision making for the advancement of candidate drugs. Identification of safety problems is a key reason that new drug development is stalled. Traditional methods for assessing a drug's safety prior to

---

The planning committee's role was limited to planning the workshop, and the workshop summary has been prepared by the workshop rapporteurs as a factual summary of what occurred at the workshop.

approval are limited in their ability to detect rare safety problems. Prior to receiving U.S. Food and Drug Administration (FDA) approval, a drug will have been tested in hundreds to thousands of patients. Generally, drugs cannot confidently be linked to safety problems until they have been tested in tens of thousands to hundreds of thousands of people. With current methods, it is unlikely that rare safety problems will be identified prior to approval.

There is, however, an emerging safety science that seeks to change this paradigm by attempting to understand a drug's safety or toxicity earlier in its development. This emerging science is focused in two areas. One is the use of various basic sciences, including genomics, metabolomics, pharmacogenomics, and others, to understand the mechanisms underlying toxicity and to predict when a particular compound will have safety issues. The other is the use of new analytical tools for mining large data sets to identify signals that indicate safety problems (e.g., those associated with a class of drugs, those associated with particular molecular entities, or those associated with particular genetic profiles) and even to derive insights regarding a drug's mechanism of toxicity.

The application of emerging science to drug safety is one of the goals of the FDA's Critical Path Initiative. A 2004 FDA white paper, *Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*, describes this evolution as follows:

> Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new products can be demonstrated, in faster time frames, with more certainty, and at lower costs. In many cases, developers have no choice but to use the tools and concepts of the last century to assess this century's candidates. As a result, the vast majority of investigational products that enter clinical trials fail. . . . A new product development toolkit—containing powerful new scientific and technical methods such as animal or computer-based predictive models, biomarkers for safety and effectiveness, and new clinical evaluation techniques—is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product. (FDA, 2004:5)

Since the publication of that report, significant progress has been made in the development of just such techniques. But the diffusion of these innovations in drug development and drug review has been limited. To address this concern, the Institute of Medicine's Forum on Drug Discovery, Development, and Translation sponsored a public workshop—Emerging Safety Science—with the goal of surveying new technologies that can be used to better understand and predict the safety and toxicity of new drugs. The workshop was held April 23–24, 2007, at the FDA's White Oak Conference Center.

The workshop addressed two general approaches to safety science. Speakers on the first day discussed the use of basic-science approaches to understand the effects of various compounds on the body, with the ultimate goal of being able to predict which compounds will exhibit which safety problems in humans. During these sessions, speakers also considered the current and foreseeable difficulties/challenges involved in developing these approaches. These included

- the current limitations of using animal models to predict human toxicity;
- the likely complexity of underlying toxicity mechanisms and predisposing human factors, as well as challenges in defining and modeling their interaction;
- the need for (and difficulty of) validating biomarkers, and the necessity of confirming potential toxicity biomarkers with human data;
- the inherent difficulty of dealing with idiosyncratic (and rare) events; and
- the need to maintain a balanced perspective so that drug candidates are not discarded prematurely based on the potential for toxicity alone.

Speakers on the second day focused on new ways of obtaining and analyzing postmarket data to identify safety problems more rapidly once drugs are marketed. Discussion during these sessions focused on how deficiencies in the current systems available for detecting and evaluating adverse events could be improved and on the development of new methods for monitoring postmarket drug safety. Proposals for fundamental changes in how adverse event data are collected, shared, and analyzed were presented during this part of the workshop.

Throughout the workshop, participants emphasized that the ultimate goal of applying these new technologies in safety science is to create a continual, iterative process in which basic scientific data can help inform and predict clinical outcomes, and clinical outcomes can be used to inform and corroborate the basic science.

This report summarizes presentations and discussions at the workshop, which should serve as a useful survey of current and emerging tools in the drug safety armamentarium:

- Chapter 2 sets the stage by describing the current state of the art in investigative toxicology, including innovative ways of using traditional methods.
- Chapter 3 is the first of four chapters devoted to emerging screening technologies. It describes cell-based screening methods (*in vitro* exper-

iments conducted using human cells) and their uses in lead identification and optimization (the process used by companies to identify and select the candidate[s] most likely to succeed throughout the development process), safety evaluation, and off-target activities, as well as in clinical prediction and exploration of putative biomarkers.

- Chapter 4 reviews various uses of and methods for toxicogenomics (the conduct of gene expression analyses to help predict the toxic effects of compounds and provide insights into the mechanisms of toxicity).
- Chapter 5 describes how metabolomics (the detection and quantification of small molecules, or metabolites) is being used to gather information on drug toxicities and their underlying mechanisms.
- Chapter 6 considers drugs that are toxic in only a subset of patients. Using the case of the anti-HIV drug Abacavir, it describes how pharmacogenetics (the study of genetic variations that affect an individual's response to a drug) can be used to identify these patients so as to prevent or at least anticipate toxicity.
- Chapter 7 presents a case study involving the experiences of the Predictive Safety Testing Consortium, formed by the C-Path Institute to bring industry, academia, and the FDA together to investigate qualifying nephrotoxicity biomarkers (quantifiable biological responses that can provide information on disease states or drug responses) for use in safety testing.
- Chapter 8 describes new approaches to pharmacovigilance (the process of collecting, monitoring, and evaluating adverse event data from patients and health care providers to identify drug safety issues). These approaches include an online signal management program, new methods for analyzing data from the FDA's Adverse Event Reporting System, and proposals for a large-scale active surveillance network.
- Chapter 9 considers how to integrate the various approaches to safety science and create feedback loops that will allow information to be shared throughout the system. Means of achieving such integration include building interdisciplinary knowledge; creating databases that allow easier identification of associations between compounds and adverse events; understanding the relevance of animal models; and developing "bridging" biomarkers that can bridge, or translate, early preclinical findings to clinical findings.
- Finally, Chapter 10 addresses areas in which further work is needed and outlines possible next steps.

2

Investigative Toxicology: The State of the Art[1]

As context for the discussion of emerging safety science, it is useful to review the current state of the art in investigative toxicology. To this end, Dr. Frazier described the work of his group at GlaxoSmithKline (GSK) on a program no longer being pursued, aimed at identifying a TGF-beta (transforming growth factor) receptor kinase inhibitor with specific activity against ALK5 (activin receptor-like kinase 5) but not other ALK (activin receptor-like kinase) receptors. According to Frazier, this work illustrates how leveraging existing tools and techniques with recent advances can make it possible to achieve the potential of cutting-edge safety science. Frazier also explained that, instead of asking the classic mechanistic questions that many people try to answer using toxicology, his group first tries to decide whether a particular toxicologic liability is a class-wide pharmacological phenomenon or is due to some individual, off-target liability. This is information can help in making decisions about lead optimization, and about whether to take a program forward or revert to a backup program and start over.

ALK5 is a transmembrane TGF-beta receptor that signals through the Smad pathways and results in nuclear translocation and activation of TGF-beta-responsive genes. Research has shown that overexpression of TGF can lead to renal fibrosis and that using antagonists to ALK5 can stop and in some cases reverse the effects of the fibrosis. During the ALK5 program,

[1]This chapter is based on the presentation of Kendall Frazier, Director of Cellular and Molecular Pathology, GlaxoSmithKline.

three major toxicities were encountered at various stages of development: pulmonary hemorrhage (exhibited during early-stage development), bone physeal abnormalities (exhibited during first-time-in-human studies), and heart valve lesions (exhibited during 10-day dose range–finding studies prior to first-time-in-human studies). The investigation of each posed different challenges and demanded different techniques.

PULMONARY HEMORRHAGE

The first problem encountered in the ALK5 program was pulmonary hemorrhage. Histopathology of lung tissue in treated rats showed diffuse alveolar damage characterized by fibrin exudation, alveolar septal necrosis, and inflammatory cells; the damage was present with a number of test compounds. Because the researchers were looking at several compounds from a series, they wanted to determine whether they would encounter this problem with every compound.

Frazier's group hypothesized that the lung damage was being caused by free radical production and reactive oxygen species. Therefore, they decided to look at the different compounds, see which ones caused free radical production, and then determine whether this superoxide production correlated with the alveolar damage. Using an *in vitro* model that employed an A549 lung adenocarcinoma cell line, they incubated the cells with either the compound or a control for 4 hours, exposed the cells to a hydroethidine dye, and then ran them through a flow cytometer with a 488 nm argon laser. When reactive oxygen species are present, hydroethidine dye turns into ethidium bromide, which fluoresces when exposed to 488 nm light. Therefore, this approach made it possible to determine quickly whether free oxygen radicals were present in a given set of lung cells.

The group looked at 150 compounds originating from three separate programs at GSK, all of which had encountered the same kind of pulmonary hemorrhage. A clear dose–response relationship was found, with higher doses leading to greater superoxide production. Some compounds led to much greater superoxide production than others, while some showed few reactive oxygen species at all (see Figure 2-1). Furthermore, the superoxide production correlated with the histopathology results: the compounds that showed increased superoxide production were the same as those that showed increased lung damage. Finally, after determining which compounds were causing superoxide production, the group reexamined those compounds' biochemical structures and found that most had a similar side chain. Frequently, one chemical series tends to be highly prone to reactive oxygen, and in this case the effects of the reactive oxygen had nothing to do with the fact that the ALK5 signal was occurring. Thus the group concluded that the lung damage was not due

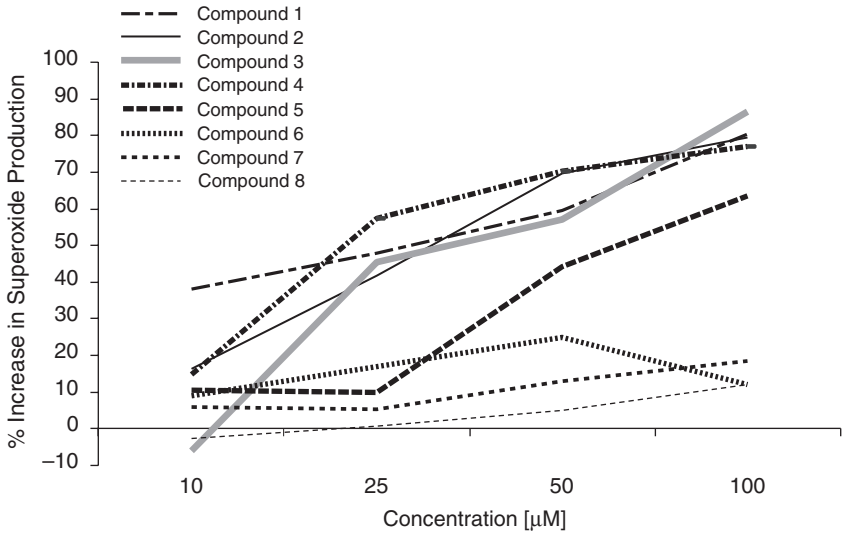


FIGURE 2-1 ALK5 (activin receptor-like kinase 5) inhibitors: the relationship between dose and intracellular superoxide production. The figure depicts the relationship between increasing doses of eight candidate compounds and the percentage of superoxide production. A clear dose–response relationship was found, with higher doses leading to greater superoxide production. It was further found that the increased superoxide production correlated with histopathology results: the compounds that induced superoxide production also exhibited increased lung damage.
SOURCE: Frazier, 2007.

to a class-wide pharmacological phenomenon, but was explained by a particular structure–activity relationship.

Frazier's group ranked 10 candidate compounds according to their superoxide generation; concurrently, the group also ran 10-day toxicology tests of the compounds. The results correlated remarkably well: the candidates that generated large amounts of superoxide showed lung lesions after 10 days, whereas those that induced limited superoxide production did not show lesions. The candidate with the least superoxide production, which was selected for moving forward, failed to elicit pulmonary hemorrhaging even after a 28-day toxicology study.

BONE PHYSEAL ABNORMALITIES

When the GSK researchers examined the femorotibial joint in animals treated with various ALK5 inhibitors, they discovered a second problem

with the ALK5 program: they repeatedly observed a particular type of bone lesion involving hypertrophy of the physes, or growth plates. The plates were much wider; the chondrocytes (cartilage cells) had a completely different appearance; and there was a large number of cells in the zone of proliferation, the part of the growth plate where the chondrocytes divide rapidly.

To understand the drug safety implications of this phenomenon, the group had to determine whether the lesions had a pharmacological basis—which would mean that every compound that inhibited ALK5 would have the same effect—and determine the clinical implications of the lesions. However, they hypothesized that because the target population for these drugs is adults, and adults have closed growth plates, the presence of the lesions in test animals might not be problematic.

Ten-day toxicological studies in rats showed a clear dose–response relationship between the various compounds and hypertrophy in the growth plates. Furthermore, the compounds that were most potent in inhibiting ALK5 had a greater effect on the physes, which implied that formation of the lesions was due to a pharmacological mechanism. When 10-week-old and 9-month-old rats were compared, the former were found to be more susceptible to the effect, implying that, as hypothesized, the clinical target population might not be affected even if there was a pharmacological basis for the effect. In a normal growth plate, there are a number of zones: a resting zone, a zone of proliferation, a zone of prehypertrophy, a zone of hypertrophy, and finally, a mineralization front. Studies conducted over the past few decades have revealed that each of these zones has a different population of chondrocytes, the cells that produce and maintain the cartilaginous matrix. The different populations have completely different cytokine profiles, gene expression, and protein expression. In short, the cartilage cells in the growth plate make up a highly heterogeneous population. It is important, then, to be able to examine the cells in each of these populations individually. Thus while recognizing the essential role of gene arrays and metabolomics, one must be sure to look at the correct cell population. Frazier's group therefore attempts to isolate individual cell populations on which to perform either transcriptomics or other profiling.

To isolate cells from the various zones, Frazier's group used a battery of special stains and immunohistochemical approaches. Using these advanced techniques, they were able to isolate and gain additional information from these different cell populations:

- Confocal microscopy allowed them to obtain a reliable count of the number of cells in various sections of the physis.

- Immunohistochemistry showed increased physeal proliferation and decreased physeal apoptosis.
- Movat staining procedures revealed increased physeal proteoglycan deposition in the hypertrophic zone.
- Studies on Von Kossa–stained frozen whole-leg preparations yielded mineralization information indicating that the bone changes were limited to the area right at the growth plate, with very minimal changes in the subphyseal area. This finding had great clinical relevance because it implied that the only changes caused by the compounds were associated with an actively growing growth plate, which would not be expected in individuals much older than about age 16–17.
- *In situ* zymography showed that there was a loss of MMP9 and MMP13 activity in the growth plate, which indicated a decreased matrix turnover and altered chondrocyte proliferation. These outcomes are associated with TGF-beta signaling; thus an ALK5 inhibitor, which disrupts TGF-beta signaling, would be expected to cause changes in MMP activity, as well as the other changes.

All of this evidence was indirect, but it led the group to believe that the physeal effect was probably pharmacological. To confirm the mechanism involved, they performed laser-capture microdissection of the growth plate. While this procedure is difficult, Frazier's group has developed techniques that make it possible to extract individual cell populations, which has opened up many new opportunities. The group also has techniques for amplifying and successfully isolating RNA from less than 1 ng and fewer than 150 cells. This capability makes it possible as well to conduct proteomic and transcriptomic analyses of small groups of cells from archives of paraffin-embedded, formalin-fixed tissue and to determine retrospectively what was occurring in some earlier toxicological studies.

In the ALK5 studies, Frazier's group used this capability to isolate groups of cells from the various zones in the physes for gene expression analyses. After 3 days of treatment with an ALK5 inhibitor, they observed marked changes in the gene expression profile, particularly in those genes already known to be associated with TGF-beta. Tapping their knowledge of ALK5's impact on each of the downstream mediators, the group was able to correlate the genes with what would be expected if ALK5 were inhibited.

To summarize, Frazier's group found that ALK5 inhibitors—at high doses—result in

- the dysregulation of a number of ALK5-related cytokines involved in chondrocyte maturation at the growth plate;

- increased proliferation and decreased apoptosis of chondrocytes;
- decreased MMP activity; and
- alteration of the proteoglycan.

Furthermore, the physeal lesions caused by the ALK5 inhibitors are similar to those seen in previous studies when some of the same ALK5-related cytokines were inhibited or knocked out in experimental animals: there is a very limited effect in the surrounding bone. The researchers concluded that the physeal hypertrophy observed in the ALK5 studies is a suprapharmacologic effect of the inhibition of TGF-beta receptors at high doses. Even though suprapharmacologic doses were being used, this conclusion indicated a target liability that would need to be addressed if the program moved forward.

Expanding on the reasons for identifying the pharmacology so thoroughly, Frazier noted that TGF-beta inhibition would be expected to cause such physeal lesions given what was already known about the system, but that his group also wanted to understand why disruption of PTHRP (parathyroid hormone-related protein), VEGF (vascular endothelial growth factor), FGF (fibroblast growth factor), and several other disparate cytokines causes similar physeal lesions. They discovered that at the zone level—the cellular level of the chondrocyte—it was not one knockout but the synergy and interaction of many factors at once that caused the physeal dystrophy. Thus it is not FGF that is causing the problem or TGF-beta or PTHRP, because in each zone they are much different; it is their interaction. If they are not turned on at the right time in the right amounts, the chondrocytes do not know when to expand, when to divide, and when to undergo apoptosis. If these things do not occur in exactly the right order, the result is the physeal hypertrophy the researchers were observing. In short, the detailed investigation was done not so much to understand what was happening with ALK5, but to understand the details for application to other drug programs that might see a similar effect.

Finally, the effect of the ALK5 compounds was found to be dose- and time-dependent; it was also dependent on the age of the experimental animals, as the older rats were somewhat resistant to it. Thus the group was able to conclude that the risk to the target population should be fairly limited, since those who would be given the drug would be old enough that they would have closed physes.

HEART VALVE LESIONS

The third pathology associated with the ALK5 inhibitors was heart valve lesions. The lead ALK5 inhibitor demonstrated an incidence of hem-

orrhagic, degenerative, and inflammatory lesions in heart valves, which occurred during 10-day dose range–finding studies as soon as 2 days and no later than 10 days after dosing. The lesions appeared at doses that were far above clinically relevant levels: clinically relevant doses for most of the ALK5 inhibitors are in the range of 3–10 mg/kg, but the laboratory animals were given doses in the range of 100–1,000 mg/kg.

A few previous ALK5 inhibitors had produced similar lesions, so the researchers wanted to determine whether these lesions represented a class-wide pathology that would be a liability for the entire ALK5 program. To answer this question, the group performed 10-day toxicological studies of another six ALK5 inhibitors at high doses, and they found heart valve lesions in virtually every study. Indeed, they failed to find lesions only when there was inadequate exposure or when they used an ALK5 compound that had no pharmacological activity. Thus they concluded that this was a class-wide pharmacological effect. The mechanism appeared to be related to effects of the compounds on the endothelium and the basal lamina. This class of compounds exhibited a variety of vascular effects; because of the stress associated with the turbulence of blood flow, however, the heart valves would likely be the first place lesions would manifest.

The heart valve lesions were novel. They had very rapid onset and caused potentially irreversible functional damage, and even though they appeared only at doses much higher than clinical levels, they were considered problematic. In addition, there is currently no toxicological biomarker for such a heart valve lesion, especially one that would be observable in 2–3 days, and the researchers had no reason to believe that the lesions were rodent-specific. The group performed laser-capture microdissection on valves from rats, dogs, and monkeys, and found that ALK5 was expressed in the heart valves of all three species. While the researchers did not see the lesions in dogs, they had no basis for assuming that the lesions would fail to appear in longer-term studies or at higher doses.

Given the rapid onset, lethality, potential irreversibility, and lack of a biomarker for the heart valve lesions, Frazier's group recommended termination of the ALK5 program. This decision was validated by another company's recent findings of similar heart valve lesions with an ALK5 inhibitor and observation of heart valve lesions in dogs.

SUMMARY

The investigative studies in the ALK5 program had three distinct purposes. First, in the case of pulmonary hemorrhage, they were used to identify a potential lead that was lacking this specific toxicity. Second,

with the physeal abnormalities, they were used to examine a mechanism that would put a particular finding into proper clinical perspective. And third, with both the physeal abnormalities and the heart valve lesions, they were used to determine whether a particular finding was structurally based or a class-wide pharmacological effect. Such a determination is not always straightforward, since finding that multiple compounds in a class cause the same problem is not the same as showing that the problem is class-wide. To elaborate, a number of compounds caused pulmonary hemorrhage, and it would have been easy to conclude that the effect must be class-wide. Yet it turned out that this was probably not a pharmacological effect, but was associated with a structure–activity relationship. Thus it is important to explore the mechanism behind an effect, particularly if one is trying to answer a class-wide pharmacological question. Furthermore, it is important to sample the target populations cleanly, as there are multiple cell populations within every organ, and confocal imagery or laser capture microdissection (LCM) can be used to identify and isolate the individual cell populations of interest.

3

Screening Technologies I: Human Cell–Based Approaches

Much of the emerging science described at the workshop centers on ways to screen drugs for potential safety problems as early as possible in the development process. In introducing the session on human cell–based approaches, David Jacobson-Kram, Associate Director for Pharmacology and Toxicology, Office of New Drugs in the U.S. Food and Drug Administration's Center for Drug Evaluation and Research, noted that existing preclinical models and paradigms often fail to predict toxicology that appears later in development. Developing the tools to select better candidates is a challenging task. Therefore, much of the workshop was focused on how to find new and more effective ways to screen or predict toxicological outcomes of drugs. Speakers described a number of approaches to improve screening for candidates that could be applied at various points in the process, from basic research and discovery through clinical development. The first topic in this series was emerging screening technologies that are based on human cells.

THE IDEAL SCREEN[1]

Dr. Butcher outlined the characteristics of an ideal screen for drug evaluation:

[1]This section is based on the presentation of Eugene Butcher, cofounder and Chair of the Scientific Advisory Board, BioSeek, and Professor, Department of Pathology, Stanford University School of Medicine.

- *Quantitative, reproducible, robust, and high-throughput*—These characteristics would make it possible to carry out informatics correlations with clinical data.
- *Highly standardized*—Highly standardized assays would be well suited to database generation and archiving. Standardized screens would make it possible to perform multiple comparisons among test groups, as well as to make comparisons over time. This is a particularly important feature, one that is missing in current efforts to model human biology.
- *Based on human biology*—As it is not possible to use human beings, the next best option is to use human primary cells.
- *Of broad interest to many people*—Assays should cover a wide range of biology and a large number of mechanisms of toxicity, including various targets, pathways, and diseases.
- *Integrative*—Integrative assays would attract the interest of scientists from multiple disciplines, including, for example, biologists, chemists, clinicians, and safety scientists.
- *Predictive*—Assays should predict safety, toxicology, efficacy, and clinical indications.

Although no such ideal screens exist today, researchers should keep this vision in mind as they work to develop new screens. The value of a screen will depend in large part on how closely it approaches this ideal.

THE BIOMAP SYSTEM[2]

Elaborating on BioSeek's own efforts to develop an ideal screen, Butcher described the long-term goal as developing *in vitro* models that can predict *in vivo* biology. By developing a database that connects drug biology to clinical responses, BioSeek's BioMAP system, based on human cells, can be used to provide an early prediction of which drug candidates are most likely to be developed as safe and effective therapeutics.

System Overview

Although BioMAP uses an artificial cell culture system, it is based on human cells placed in complex environments designed to reflect key aspects of the natural environments the cells would experience in the human body. Using information from the literature, the company's scientists strive to create environments that mirror real situations in which multiple pathways are active at the same time—pathways similar to those believed to work together in different disease states. One cannot model *in vitro* biology by

[2]This section is based on the presentation of Dr. Butcher.

assaying a large number of individual targets because one cannot predict network biology from target and pathway biology. The BioSeek approach involves taking advantage of systems biology principles and engineering complexity into the system so as to model as closely as possible the biology that occurs *in vivo* in different disease physiological settings.

To create these complex environments, the BioMAP system combines a number of cells, sometimes of a single type and sometimes a defined mixture, such as peripheral blood mononuclear and endothelial cells. The purpose of using multiple cells is to capture cell–cell interactions and begin to model at a simple level what is going on at the tissue level. These systems are then stimulated by agents such as cytokines and growth factors to create an environment where a number of disease-relevant pathways are simultaneously active.

The development of these models is informed by *in vivo* data from the literature. The goal is to develop a model system that responds in a particular way to particular drugs because that is how the tissue responds *in vivo*. The process is an iterative one, with lessons learned at each step being applied to modify the system in useful ways.

In particular, the focus of the BioMAP system is on factors that reflect and control biology *in vivo* and that mediate disease, such as small receptors, cytokines, chemokines, enzymes, and growth factors. Because it is more cost effective, the system uses standard enzyme-linked immunosorbent assay (ELISA)–based or morphologic readouts rather than microarray data.

The experimental process follows a standard pattern. The assays are set up and stimulated, and the researchers then read off the various parameters in which they are interested. Next, the system is perturbed with a drug to generate a profile that is entered into a database. Through numerous iterations of this process, BioSeek has accumulated a database containing thousands of BioMAP profiles that catalogue the effects of thousands of different drugs in a variety of disease model systems.

BioSeek has developed at least 30 of these complex cell systems designed to model different aspects of disease and is actively working to develop additional systems. The types of cells used include, for example, primary endothelial cells, monocytes, lymphocytes, macrophages, mast cells, smooth muscle cells, keratinocytes, bronchial epithelial cells, and smooth muscle cells.

Table 3-1 illustrates the kinds of cell systems BioMAP employs. The first two systems contain primary endothelial cells. In the first, the cells are stimulated in a TH-1 environment with three cytokines added, creating an environment similar to one that might be seen in psoriasis or rheumatoid arthritis. In the second system, there is a TH-2 environment with IL-4 (interleukin-4) and histamine, resulting in an environment more

TABLE 3-1 Examples of BioMAP Systems

System	Environment	Cell Types	Readouts
3C	IL-1β + TNF-α + IFN-γ	Endothelial cells	E-selectin, VCAM, ICAM, uPAR, MCP-1, MIG, IL-8, HLA-DR, CD142
4H	IL-4 + Histamine	Endothelial cells	VEGFRII, P-selectin, VCAM, uPAR, Eotaxin-3, MCP-1, viability, morphologic score
LPS	LPS (TLR4)	Endothelial cells + lymphocytes/ monocytes	CD14, CD141, CD142, CD40, CD69, MCP-1, E-selectin, IL-1α, IL-8, M-CSF, VCAM, PGE
SAg	Superantigens (TCR)	Endothelial cells + lymphocytes/ monocytes	CD38, CD40, CD69, E-selectin, IL-8, MCP-1, MIG, prolifn, viability

SOURCE: Butcher, 2007.

relevant to asthma. The two other systems combine peripheral blood mononuclear and endothelial cells, stimulated through either (1) a selector that would selectively activate through a monocyte cascade that would activate many pathways, or (2) the T cell receptor.

These four systems alone, because they encompass most of the targets and mechanisms and pathways involved in inflammation, allow the BioSeek researchers to detect and discriminate compounds of basically every important immunomodulatory agent, as well as many molecules that are important in other biological and therapeutic areas, including cardiovascular disease, metabolism, and cancer. This broad range of applications is not surprising because even though these four cell systems were created to model inflammatory and cardiovascular states, the cells they contain express the receptors that cancer cells adapt and use for their own purposes, as well as many of the receptors involved in controlling metabolism and lipid biology. The resulting breadth of coverage of targets and pathways provides BioSeek with a unique opportunity to assess effects across a broad array of human biology in a common format.

BioMAP Profiles

Butcher provided an overview of how the BioMAP system is used in the development process. Once a compound has been received, it is run through a set of cell-based assays, all of which are performed using robotics and micro titer plates. The resulting profiles are reviewed for quality control and then archived. BioSeek researchers have developed visualization and knowledge management tools that allow them to correlate the profile data of a compound with any other information of interest. Chemical informatics is used for structure–activity relationship (SAR) analysis, as well as for literature mining.

An example is a profile of a p38 inhibitor that was tested in multiple cell systems (see Figure 3-1). All of the systems were normalized to a control level without the presence of the drug, and by stimulating each system with the test compound, the researchers established dose–response curves. Because these assays are performed by ELISA using robotics, it is possible to test hundreds of compounds in multiple assays each week, to repeat the tests a number of times, and therefore to collect enough data that the statistical analysis can be quite sophisticated. These statistics allow the researchers to calculate a 99 percent significance envelope, which is indicated in the figure by a gray background; anything outside of this envelope is a highly significant response.

With thousands of profiles, it is possible to look for similarities among them and thus identify compounds with similar responses in the cell systems. In particular, the BioSeek scientists analyze their profiles using a form of clustering known as multidimensional scaling. An example of such a clustering analysis is shown in Figure 3-2 (see p. 20). Each dot represents a profile; some profiles end up close together, indicating similarity, while others end up far apart. Since the graph is formed by collapsing 40-dimensional space into 2 dimensions, two dots being positioned close together on the graph does not always imply that their profiles are similar. To make the similarities clear, lines are drawn between compounds whose profiles are statistically similar within the data set, so that only compounds connected by those lines have similar profiles.

The analysis makes it possible to cluster compounds with similar mechanisms of action rapidly and to identify secondary off-target activities quickly. For example, Figure 3-2 shows two MEK inhibitors, PD098059 and UO126, that have the same primary activity but do not cluster together because of their strong secondary activities. In a similar way, the clustering analysis has separated two PPARγ inhibitors, Rezulin and avandia, which have very different biological activities. The same approach can also be used for pathway analysis.

When a new molecule is submitted for analysis, the BioSeek researchers run its profile and then compare this against the thousands of pro-

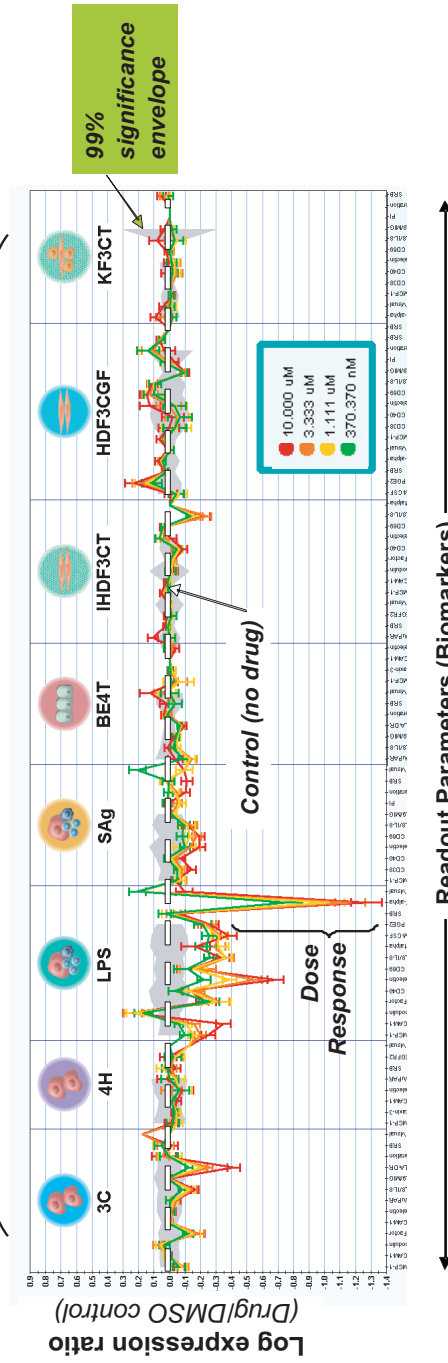


FIGURE 3-1 Example of a BioMAP profile developed using a reference p38 inhibitor. The figure displays data from multiple experiments in which a p38 inhibitor was tested in several BioMAP cell systems. Readout parameters are listed at the bottom of the figure. Once the systems were normalized to a control level without drug, they were treated with a range of doses of the p38 inhibitor to develop a dose-response curve. The gray background represents a 99 percent significance envelope; therefore, anything outside of that envelope is considered a highly significant response. These data demonstrate that BioMAP activity profiles are robust and reproducible, and that the profiles retain their shape across multiple drug concentrations. The shape of a drug's BioMAP profile thus provides a fingerprint or signature for its biological function and target. SOURCE: Butcher, 2007.

files in the database, looking for homology, similarity, and function. The database returns a list of compounds that are most similar to the test compound, ranked in order of similarity and with accompanying statistics. An example is an analysis of a MAP kinase inhibitor. When screened against the database, the top two hits returned previous runs performed with the same compound. However, the next dozen hits were other p38 MAP kinase inhibitors. This example demonstrates that analysis of the BioMAP profiles makes it possible to rapidly link chemistry to biology and identify mechanisms of action.

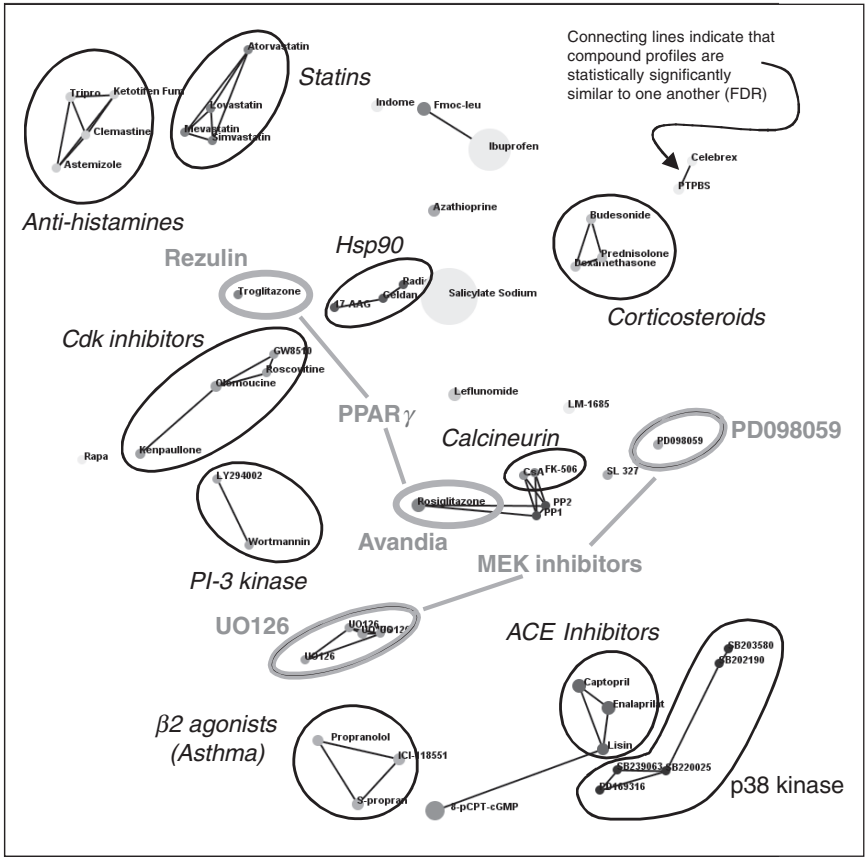
Applications to Safety Science

Although BioSeek is focusing on techniques for application to drug development, lead identification, and optimization, techniques based on the BioMAP system are equally applicable to safety evaluation. For example, BioSeek was examining a candidate drug for which there were compelling animal data in models of inflammatory bowel disease, but the development had stalled because the mechanism of the drug's action was unknown. When the researchers compared the compound's profile against the database, they identified a potential match, and this match suggested a mechanism of action. After confirming the target with a biochemical assay, the researchers were able to reject the program because that particular target had known target-specific toxicities.

In another case, a BioSeek partner was screening compounds that had been identified in a simple inflammatory assay. The BioMAP profiles were performed and run against the database, and the database comparison identified a cancer target as the mechanism in one of the compounds. The researchers at the partner company were at first skeptical of the results because the chemistry was incorrect. However, the target was confirmed, and the toxicity expected from the compound made it unacceptable for the desired indication.

BioSeek has also found that it can look at compounds at higher doses, induce toxicity, and then classify those compounds by the mechanisms by which the toxicity is induced. Many toxic compounds, for example, have a common final pathway for inducing apoptosis but different mechanisms of action leading up to the apoptotic event. BioMAP clustering analysis can separate compounds according to these varying mechanisms. This ability is of interest to partners in a variety of areas. For example, BioSeek recently undertook a collaboration with the Environmental Protection Agency to characterize biologically the mechanisms of toxicity of a wide array of both drugs and environmental chemicals.

Yet another application of the BioMAP system is to identify off-target activities. For instance, ibuprofen was recently found to cross-react with



Mechanism of Action

Compounds with the same target (mechanism of action) cluster together

Secondary Activities

Compounds with different targets or strong secondary activities do not cluster together

nuclear hormone receptors of the PPAR family, a result that would have been immediately apparent from BioMAP analyses. In comparing the ibuprofen profile against the BioSeek database, a number of nonsteroidal anti-inflammatory drugs (NSAIDs) closely related to ibuprofen would appear, but one would also pick up various PPAR agonists. Thus ibuprofen's tendency to activate PPARs would be apparent. This ability to identify off-target activities can also be used in comparing and differentiating

FIGURE 3-2 *(facing page)* Example of a computational analysis of BioMAP profiles of various compounds. This example of a clustering analysis illustrates the power of BioMAP to classify compounds by their mechanism of action. On the figure, each dot represents a drug profile. Clustering together indicates that the compounds have a similar biology, implying a common target or mechanism of action. Compounds with different targets or strong secondary activities do not cluster together. Compounds whose targets are highly sensitive to conformational effects (e.g., nuclear hormone receptors) may also display divergent biology. Because this graph was formed by collapsing 40-dimensional space into two 2 dimensions (by multidimensional scaling), statistical relationships were added to improve clarity: the lines drawn between compounds indicate statistically significant similarities between compounds within the data set. Only compounds connected by those lines have similar profiles.
SOURCE: Butcher, 2007.

among similar drugs. Butcher showed a figure with profiles of several p38 inhibitors. As expected, the profiles were similar, but there were noticeable differences, and in many cases these differences could be associated with specific secondary targets. Even if an off-target activity cannot be identified, the fact that such activity exists in a new compound should lead researchers to think carefully about what the compound might be doing differently in biological terms.

Finally, BioMAP profiles can be applied to clinical prediction and establishment of biomarkers. Indeed, the BioMAP system was purposely developed with clinically relevant readouts. The system focuses on molecules, selected for their information content, that mediate disease, are sensitive to many targets, have high predictive value, and have potential suitability as clinical biomarkers.

Butcher described BioSeek's goal of developing a comprehensive BioMAP database connecting drug biology to clinical responses. Such a database could help identify drug candidates with safe and effective therapeutic profiles. Accomplishing this goal will demand accumulating a great deal of clinical data about what drugs are doing in people in addition to BioMAP profiles, conventional toxicological data, and models of human biology and disease. It will be an iterative process in which biological information and statistical and informatics correlations will be used to develop predictors; the predictions will be made; and they will be improved over time, with clinical outcomes being fed back to inform the further development of the BioMAP system and interpretations.

CONTEXTUAL DRUG ANALYSIS[3]

Dr. Westwick described a second analytical technique based on human cells. The ultimate goal is conceptually similar to that of the BioMAP process: to develop profiles of responses to various compounds that can then be used to analyze new drugs, pinpointing their mechanisms of action and predicting potential toxicities. However, the technique used to develop the profiles is very different.

Biologists often draw pathways as linear or circuit diagrams, but that is not how signal transduction happens or how drugs act on their targets. Instead, drug effects occur on protein complexes containing dozens of protein components—not on isolated proteins in a test tube, and not even on isolated proteins in the cell. Furthermore, in real biological pathways, the proteins and protein complexes move around. To gain a better understanding of what is happening on the cellular level, it is necessary to look in the cell because the effect of a drug will be dependent on the localization of the drug target. To better understand these mechanisms, Odyssey Thera developed a method for observing the actions of drugs in the context of living human cells.

High-Content Chemical Biology

The profiling method developed by Odyssey Thera relies on two main techniques. The first is a high-content cellular analysis that employs automated, high-throughput confocal microscopy. Excellent instrumentation is available in this area, and automated image analysis has also improved dramatically and can be combined effectively with the automated microscopy. The second technique is a proprietary process based on protein-fragment complementation assay (PCA). With this technique, two fragments of a rationally dissected reporter protein are attached to proteins of interest, for example, a kinase and a substrate. When those two proteins come into close proximity, this allows the spontaneous refolding of the reporter protein and the generation of a signal, which can be enzymatic, fluorescent, or luminescent. With a microscope, one can observe these protein complexes in live cells—not just their existence, but also their positions within the cells.

Westwick showed a real-time film of the types of interactions that can be observed by this method. Reporter protein fragments had been attached to two protein kinases, AKT and PDK-1, so that when they came together, a fluorescent signal was generated, and a greenish glow marked

[3]This section is based on the presentation of John Westwick, President and Chief Scientific Officer, Odyssey Thera, Inc.

their presence. At the beginning, the protein complexes could be seen in the cytoplasm of the cells. When the cells were stimulated with a growth factor, the complexes moved to the membrane. Finally, when an inhibitor of an upstream kinase was added, it inhibited the membrane localization, and the complexes moved back to the cytoplasm.

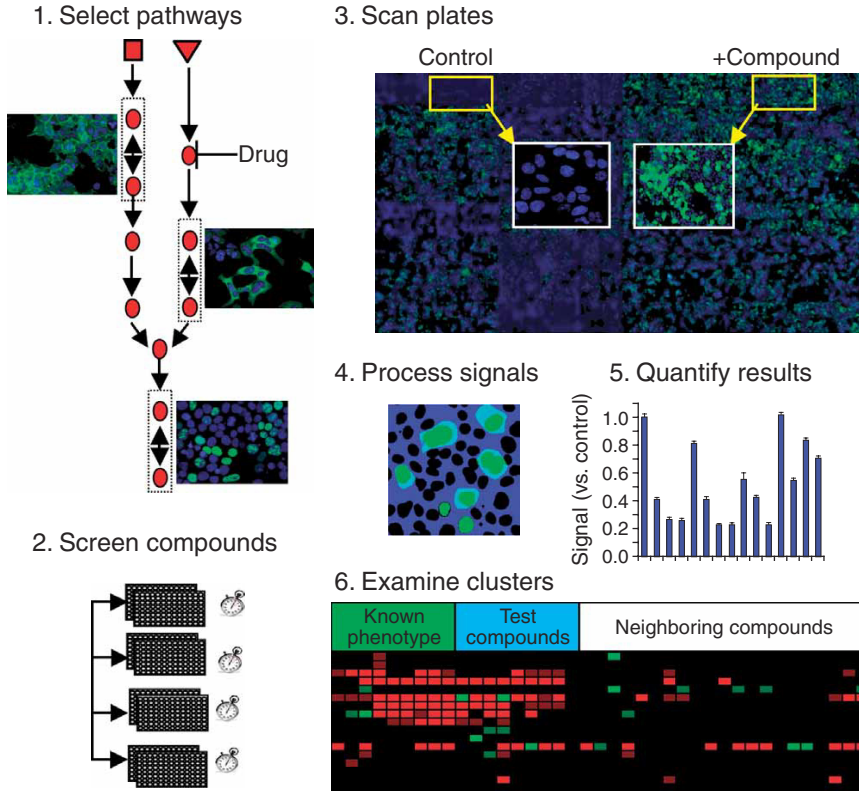
According to Westwick, the system is a cell biologist's dream as it offers the opportunity to generate a tremendous amount of data about what is going on inside a cell, and where. At the same time, it also poses a number of major challenges. Westwick focused on two challenges in particular: the informatics challenge, and the challenge of developing diverse biological assays for use in identifying the numerous pathways in a cell.

Work on the informatics challenge is the more advanced of the two. The company has five automated microscopes capturing more than half a million images per day, 6 days a week. With hundreds of compounds being analyzed every few weeks with several hundred assays, the company's researchers are generating several terabytes of data per week. Only in the past year did the company finally solve the information technology infrastructure challenge by developing a novel strategy, and Odyssey Thera is now able to handle this avalanche of data effectively.

The challenge of developing diverse biological assays is much greater. Good instruments are available, as are the engineered platforms, but there are relatively few assays for use on these platforms. Moreover, the sort of profiling done by Odyssey demands a wide range of assays. Westwick stressed the importance of being "agnostic as to target class and pathway"; one must cast a broad net when looking for off-target effects.

To overcome these challenges, Odyssey devoted much of the past few years to expanding its assay panel, trying to cover as many pathways and as much cellular space as possible. The company's assays now encompass a wide variety of targets in the cell: GPCRs, kinases, cytoskeletal proteins, GTPases, G proteins, transcription factors such as p53, and nuclear receptors such as PPARγ, as well as some less common targets such as protein ubiquitination and the proteasome. The assays can also be used to look at apoptotic machinery, heat shock proteins, ion channels, and protein complexes involved in chromatin remodeling. In short, the company's assays cover a wide range of target classes and processes.

To create profiles, a large and varied panel of assays is chosen—at least 100 are essential to encompass diverse pathways and target classes. These assays are then used to screen the compounds of interest at multiple time points. The researchers choose doses that are efficacious for the compound on its target in cell-based models. After the assays have been performed, the plates are scanned, and the signals are processed at the subcellular level to generate quantitative data. Various data classification strategies are then applied for data summary and display (see Figure 3-3).



The resulting data are then expressed as a heat map that looks something like a gene expression heat map. From this heat map it is possible to identify structure–activity relationships. The statins cluster together, for instance. Beyond this clustering, the other important thing about the heat maps is that each of these of these compounds has a unique fingerprint; no two look the same. It is simple to use the underlying data to understand the effect of a specific chemical substitution on a compound's activity within the cellular networks.

Application to Safety Science

Application of this technology for safety and toxicology analysis of new compounds requires the generation of fingerprints or signatures for toxicants as well as for efficacious drugs. Following the generation of a

FIGURE 3-3 *(facing page)* Strategy for pharmacological profiling of compounds with high-content PCAs. (1) Pathways of interest are selected and high-content PCA assays are created (pathway represented as red spheres connected by arrows). Assays measure dynamics of specific pathway activation or inhibition by quantifying changes in abundance or location of protein complexes coupled to that pathway that are elicited in response to activator (red square and triangle) or inhibitor drugs (>). Inset are images of three such assays that report on dynamic complexes coupled to the individual pathways (dotted-line boxes) localized to membrane, cytosol and nucleus. PCA signal is in green; nuclear (Hoechst) staining is in blue. (2) Cells expressing PCAs arrayed in 96-well plates are treated with compounds or vehicle controls, fixed after specified times and treated with cell compartment–specific counterstains. (3) Multiple images are captured from control and compound-treated wells. Pixel intensities from PCA signals are extracted from one or several cell compartments on the basis of colocalization with counterstain (4) and tabulated for individual compound treatments (5) (Methods and Supplementary Methods). Data for each compound versus PCA response at different times are represented as an array. Changes in signal intensity or location for compound versus vehicle control are represented by a color code, where green represents an increase and red a decrease in PCA signal versus control in units of coefficient of variation of each assay. Data are clustered by compounds and assays to identify on-pathway or off-pathway effects of compounds on specific pathways. The matrix also allows for the identification of test compounds that cluster with drugs of a known phenotype and could be expected to share the same phenotype.
SOURCE: MacDonald et al., 2006. Figure and legend reproduced as published from *Identifying Off-Target Effects and Hidden Phenotypes of Drugs in Human Cells*. Reprinted by permission of Macmillan Publishers, Ltd: Nature Chemical Biology, copyright 2006.

signature (a subset of assays that appear to be characteristic for a particular compound), automated algorithms are used to search through all the other thousands of compounds in the company's database for those that have a signature similar to that of the compound of interest. Comparison of compound signatures within a target class can lead to interesting discoveries. To illustrate, Figure 3-4 shows signatures from a number of different statins. From these data, the following types of information can be derived:

- Looking at the statins, one can see that there are characteristic activities within cellular networks. Each of the points on a graph represents the results of a single assay, and there are several hundred for each compound (the figure does not show all of them).
- The signatures of the statins are similar, which would be expected, but there are also differences. Throughout one indicated region, for

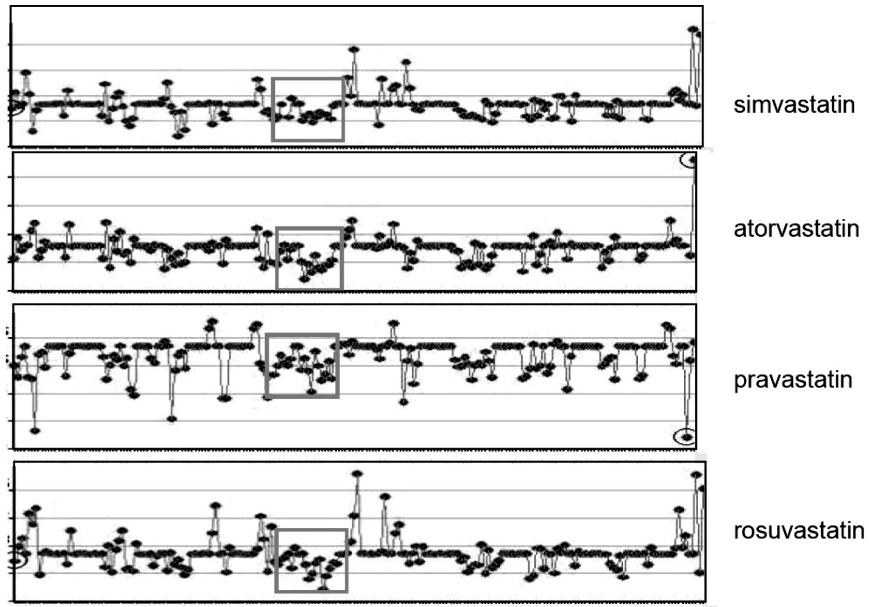


FIGURE 3-4 Exploring global mechanistic differences within multiple statin compounds. Each of the points on this graph represents the results of a single assay, and there are several hundred for each compound (the figure does not show all of them). The signatures of the statins are similar, as expected. Using this technology, an average profile for all statins across the whole assay panel can be created, and then various drugs in the database can be compared with this average profile. While most of the statins tested resemble this average profile, pravastatin deviates notably from the average.
SOURCE: Westwick, 2007.

instance, pravastatin looks distinctly different from simvastatin. One thing that can be done is to generate an average profile for the statins across the whole assay panel and then compare the various drugs in the database with this average profile. Not surprisingly, most of the statins tend to resemble this average profile, but pravastatin is an exception that deviates notably from the average.

- One can also ask whether other drugs have similarities to the statins. Some do, of course, and the analysis identifies them. It is then interesting to note exactly what similarities they have. By examining the statins more closely, as well as the differences among them and their similarities to other drugs, one can draw conclusions about how the various statins are functioning. For example, cerivastatin is moderately

similar to the other statins, but it also has similarities to rotenone and other compounds that disrupt the mitochondrial energy chain, as well as to some antibiotics and toxicants, such as n-nitrosodiethylamine. In that same vein, the comparisons show that rosuvastatin is most similar to the "core" statin signature. In some respects it is the average statin, but it also has profile matches to microtubule modulators and DNA binding compounds, which could explain some of its off-target activities.

- Atorvastatin, or Lipitor, is very similar to the other statins, but it also matches closely both FK506 and rapamycin, both of which are immunosuppressive drugs used to prevent rejection of transplanted organs. In particular, atorvastatin has activity on the S6 kinase pathway, which is why it matches up with the two immunosuppressive compounds, suggesting that it, too, will have immunomodulatory activity. This is important information because the anti-inflammatory activity of statins is essential to how they work.
- Finally, pravastatin's global profile is notably different from the profiles of all the other statins. The part of its profile relating to its activity on HMG-CoA reductase is similar to that of the other statins, since all statins are inhibitors of this enzyme; over the rest of its profile, however, pravastatin is similar to cyclooxygenase inhibitors, indicating that it may have unique anti-inflammatory properties. This is a testable hypothesis, one that is supported by the literature.

A global profile of cholesterol was also run to identify drugs with similar profiles. A number of hits were returned, including rotenone, β-laphachone, Ketek, and nefazodone. All of these compounds exhibit some toxicity, and all can be toxic to hepatocytes. Because high levels of cholesterol are also toxic to hepatocytes, these results make sense. Although this profile would not necessarily disqualify Ketek from consideration, it would prompt researchers to investigate further its effects on hepatocytes *in vitro* and *in vivo*.

SUMMARY

The profiles generated by this technology can offer a number of insights into the potential toxicity of compounds, as well as into desirable drug mechanisms. BioSeek is working to develop a comprehensive BioMAP database connecting drug biology to clinical responses. Odyssey Thera's current strategy is to rigorously define training sets based on toxicants as well as desirable drug classes and then to match test compounds to these profiles. In this way, the researchers hope to be able to enable a deeper understanding of cellular networks and drug targets and to facilitate more informed discovery and development decisions.

4

Screening Technologies II: Toxicogenomics

In recent years, toxicogenomics has started to become fully integrated into drug safety assessment and into the efforts of the U.S. Food and Drug Administration (FDA) to build more safety into drugs, noted Federico Goodsaid, Senior Staff Scientist in Genomics at the FDA's Center for Drug Evaluation and Research. Toxicogenomics has also increasingly become a major tool in the development of new biomarkers for drug safety assessment. Four speakers from three different companies—Iconix, Bristol-Myers Squibb, and Abbott Pharmaceuticals—explained how their firms are developing and applying toxicogenomic tools for drug safety. The goal was to describe the range of drug safety data now being supplied by toxicogenomics—from preclinical safety assessments to the clinic. The following summaries address data derived from studies conducted in rats. It is important to note that these results apply only to rats and have not yet been shown to be clinically relevant. Future steps for these technologies and these types of databases include determining whether they are clinically relevant.

MODERNIZING PREDICTIVE TOXICOLOGY[1]

Dr. Halbert discussed how current work in toxicogenomics is an important part of the FDA's Critical Path Initiative. Specifically, Item 20

[1]This section is based on the presentation of Don Halbert, Executive Vice President for Research and Development, Iconix Pharmaceuticals.

on the Critical Path Opportunities List calls for "modernizing predictive toxicology," described as follows:

> Identifying preclinical biomarkers that predict human liver or kidney toxicity would speed innovation for many different types of therapeutics. Activities to develop genomic biomarkers for the mechanistic interpretation of toxicological observations—complementary to but independent of these classic toxicological observations—could begin to create the data foundation for qualification of new safety biomarkers. Collaborations among sponsors to share what is known about existing safety assays could be a first step toward the goal of safer medical products. (FDA, 2006:9)

Halbert explained that the fundamental underlying principle of toxicogenomics is that compounds with similar mechanisms of toxicity and efficacy will have similar gene expression profiles. Thus information about how various compounds affect gene expression—in the context of other knowledge about those compounds—can lead to a better understanding of both the compounds' mechanisms of action and their toxicity. One of the goals of toxicogenomics is to identify biomarkers—generally sets of genes or RNA—from data collected on known drugs and toxicants, and use these biomarkers to predict mechanisms of action or toxicity in new compounds.

To be effective, toxicogenomics requires the collection and analysis of large amounts of data. These data must be highly diverse, in terms of not only the types of drugs and compounds that should be represented in the database, but also the types of data collected. For example, gene expression data should be collected in addition to traditional toxicology end points such as clinical chemistry and histopathology. The data should be organized in a well-curated database, and their interpretation requires novel methods of analyzing patterns and predicting outcomes.

Toxicogenomics is currently being used in a variety of ways in drug discovery and development. It is being applied

- to rescue at-risk programs at the preclinical or early clinical stages by gaining additional insight into a compound's mechanism of action and how it is causing toxicity;
- to screen and evaluate leads at different stages proactively by predicting toxicities and mechanisms of action so that candidate compounds can be eliminated from the development pipeline as early as possible; and
- to develop preclinical biomarkers of drug response and toxicity.

Toxicogenomics offers a number of advantages. Gene expression can be predictive and can be more sensitive than traditional approaches. It

is high-content: many things are measured at the same time, and in particular, biomarkers from multiple end points can be measured in a single experiment if one understands what those biomarkers are. Toxicogenomics also supports an understanding of both toxicity and safety. Moreover, it can quickly provide a great deal of additional mechanistic understanding when a problem with a compound arises. The hope is that all of these capabilities will lead to better decision making, removal of candidate compounds from the development pipeline earlier in development, and increased confidence in moving compounds forward.

TOXICOGENOMICS AT ICONIX[2]

Halbert described how his company, Iconix, uses toxicogenomics in drug discovery, in biomarker identification and validation, and in preclinical safety assessment.

The DrugMatrix Reference Database

At Iconix, toxicogenomics is grounded in a large database called the DrugMatrix Reference Database. It allows researchers to identify the mechanisms of toxicity of novel compounds through comparison with the database's reference set of compounds, to benchmark the effects of unknown compounds against these reference compounds, and to identify potential biomarkers that can be used to predict both toxicological and pharmacological end points in rats.

The DrugMatrix database was assembled by accumulating standardized information on more than 640 compounds across nine different tissues in male Sprague-Dawley rats. In choosing the compounds to include in the database, researchers ensured that there were at least three molecules in the database for each structure–activity class of compounds.

The maximum tolerated dose (MTD) and fully effective dose (FED) were estimated from the literature, and a preliminary range-finding study of dose versus toxicity was performed to determine the dose levels to use for each compound. Full studies were then carried out on at least 24 rats for each compound: two dose levels (MTD and FED), at least four or five time points (¼, 1, 3, 5, and 14 days), and three rats per group.

When the rats were sacrificed, all of the tissues and blood were harvested and stored in freezers so they would be available later for gene expression or histopathological studies. Gene expression studies were carried out for the more than 640 compounds, and a full set of histopathological information was generated for each, including histology, clinical

[2]This section is based on the presentation of Dr. Halbert.

chemistry, hematology, and body and organ weights. Furthermore, full pharmacological profiling was carried out on 870 compounds, including the 640 that were chosen for gene expression profiling. This pharmacological profiling consisted of 127 assays, including receptor binding, enzyme, and drug-metabolizing enzyme (DME) assays. The result was a highly comprehensive set of information on how these compounds affect rats.

Biomarkers

Once the information described above had been accumulated and organized into a database, it was used to identify "RNA-based biomarkers." The purpose of identifying biomarkers is to be able to predict various end points from the gene expression data—that is, to know in advance what outcomes can be expected by detecting certain patterns in how a compound affects gene expression.

Iconix attempted to relate four types of phenotypic end points to patterns of gene expression—histopathology, pharmacology, clinical chemistry, and hematology measurements—and sought literature annotations as to how the compounds affect laboratory animals. To develop a biomarker, researchers select a phenotypic end point of interest and create a training set. The training set consists of a positive set of treatments that lead to the desired end point and a negative set of treatments that do not. This training set, along with the gene expression data for each treatment, serves as the input to a Support Vector Machine (SVM) classifier algorithm, which in turn identifies a biomarker—a pattern of gene expression correlated with the end point of interest. The biomarker can then be validated internally within the data set, and possibly in a forward-validated way.

To use these biomarkers, a gene expression profile is generated for a test compound. If the test compound matches any biomarker for a particular end point, this indicates that the test compound causes gene expression changes in that tissue, changes similar to those caused by compounds in the class used to build that biomarker. This in turn means that the test compound is similar to that particular class of compound and can be expected to generate a similar end point.

In addition to helping to identify biomarkers, the information in the DrugMatrix database provides rich insight into what is occurring at the transcription level when animal organs are perturbed with these compounds at a variety of different doses and time points. This information can help achieve an understanding of the mechanisms of activity for compounds at a much deeper level.

Example: Developing a Kidney Biomarker

One of the fundamental questions regarding gene expression is whether it can be used to identify changes in an animal that are predictive of something that has not yet occurred. Halbert described the development of a kidney biomarker that Iconix hoped would be indicative of the development of kidney injury. Researchers hypothesized that the biomarker would be seen in advance of the detection of any pathological changes in the animal, thus ultimately predicting injury prior to its actual incidence.

Researchers designed an experiment to produce latent renal tubular injury in rats. As is well known, the renal tubules are a major site of toxicity and can be damaged by a variety of drugs. Working with a set of 119 compounds that caused the kind of delayed kidney damage of interest—that is, no histopathological injury at day 5 but measurable injury at day 28—the researchers identified a multigene biomarker that could predict the kidney damage from gene expression patterns that were apparent on day 5. The biomarker was validated by being tested on 32 compounds that had not been used in the original training set.

The researchers then checked the literature for individual gene biomarkers identified as predicting this sort of damage and compared them with the new multigene biomarker. The highest-performing individual gene was Tsc-22. It had a sensitivity of about 63 percent but had a very high false positive rate, so that its specificity was only about 44 percent. By contrast, the multigene biomarker had a sensitivity of 83 percent, a specificity of 79 percent, and an overall accuracy of about 75 percent.

An additional benefit of this biomarker was that it contained a number of genes that could be related to the types of injury that were occurring in the kidney. Thus the gene expression changes served to highlight the various early mechanisms and pathways that contributed to the eventual nephrotoxicity (see Figure 4-1). Despite this biomarker's success, however, a number of questions remain, such as whether it can be validated in other laboratories; whether its level of accuracy is sufficient to make it useful for drug discovery and development and allow it to replace more costly and time-consuming assays; and how relevant it will be for humans, given that it was derived in rats.

The Future of Toxicogenomics

Generally speaking, then, genomic biomarkers have great potential, holding promise for increased predictivity, sensitivity, and specificity, although in every case it is necessary to do an independent forward validation. Furthermore, these biomarkers are relatively easy to apply. Gene expression can be measured easily in target organs by using microarrays

FIGURE 4-1 Example of gene expression changes highlighting early mechanisms and pathways that contribute to nephrotoxicity. The above signature was developed to predict nephrotoxicity. In this experiment, gene expression levels were measured after five days, prior to any indication of histopathological injury to this tissue. Many of the individual genes that had increased expression are relevant to the types of injury that would be expected in nephrotoxicty.
SOURCE: Halbert, 2007.

and RT-PCR (reverse transcription polymerase chain reaction), and special models or treatment conditions are unnecessary—the gene expression studies can be piggybacked on other studies already under way. As a result, expression profiling is increasingly being incorporated into standard lead optimization and preclinical studies, and is being used as part of the general evaluation performed when one is deciding whether to take a particular compound forward.

More specifically, toxicogenomics is being applied in a variety of ways to drug discovery and the development interface. It is being used for the prospective prediction of toxicology and the retrospective understanding of why a compound is causing a particular problem. At the drug discovery stage, it is being used to help rank and select compounds or chemical platforms *in vivo*, and researchers are beginning to develop the technology to the point where it can be used *in vitro*. Later in the process, toxicogenomics is proving useful in the benchmarking of compounds relative to competitor molecules. This technology is also being used to verify safety at the gene expression level. Researchers establish confidence in a compound by studying it in a variety of tissues and at various doses and time points, and determining that it does not cause the same sorts of changes in gene expression that are signals of problems in the reference compounds that have been studied.

Halbert predicted that the utility and impact of toxicogenomics will drive its acceptance. Already the FDA is learning how to apply this technology. The agency has had a copy of the DrugMatrix database for 2–3 years and has been applying the technology and trying to understand how gene expression can be useful in assessing compounds. Validation of and improvements in the technology will make it possible to move it upstream in the drug discovery process. Working with Abbott Laboratories, Iconix has done a great deal of work with *in vitro* primary rat hepatocytes that can be used to predict particular end points. Moving from *in vivo* to *in vitro* applications will make it possible to increase the sample throughput and begin to look at molecules at a much earlier stage, gaining some understanding of the safety of the molecules at the transcriptional profiling level very early in the process. This capability will lead in turn to reduced costs, and it will also be possible to automate much of this work.

The question remaining is how to transition from prediction of toxicity in rats to prediction of toxicity in humans. Data from rats are very important for addressing regulatory questions, and there have been some efforts to link drug responses in rats with clinical outcomes (toxicological responses in humans). However, a reliable method for doing so has not been developed, and a correlation between rats and humans has yet to be established.

TOXICOGENOMICS AT BRISTOL-MYERS SQUIBB[3]

As a second example of the use of toxicogenomics in safety science, Dr. Cockett described how Bristol-Myers Squibb (BMS) uses gene chips and gene expression analysis in the late discovery stage, as well as in the early candidate assessment stage. The experimental design for this technology includes treatment of animals or cell lines with a compound at varied doses and time points. Following treatment, RNA is extracted from the animal tissues or cells and analyzed with a gene chip, such as the Affymetrix Human Genome U133, which contains 44,000 probe sets corresponding to about 32,000 genes. The resulting data are displayed with a heat map that indicates those genes whose expression levels have changed significantly as a result of exposure to the compound, as well as the degree to which the expression level has changed.

A Simple Example

To provide an idea of how this technology might be applied, Cockett described an experiment aimed at creating a disease-like state and then identifying an optimal treatment to cure that state. THP-2 monocyte cells were exposed to tumor necrosis factor (TNF) to create a "surrogate disease phenotype," and gene chips were used to measure the cells' changes in gene expression. The TNF-exposed cells were then treated with a variety of drugs, and the response was measured again. The goal was to find drugs that would reverse this phenotype, thereby "curing" the surrogate disease. The gene expression results were displayed in the form of a principal component analysis (PCA) of 515 response markers (see Figure 4-2).

In the graph of that PCA, control treatments can be seen clustered near the bottom, all in the same color; the TNF-treated cells can be seen clustered near the top of the graph, again in the same color; and scattered around the graph are clusters of other colored dots representing the outcomes of various treatments. The multiple dots of each color that cluster together are experimental replicates and serve to demonstrate the reproducibility of these measurements. As can be seen in the graph, some of the drugs—represented by dots that lie near the controls—completely reversed the TNF response. These were the drugs with the desired response against the surrogate disease. Other drugs not only reversed the TNF stimulation, but also created other effects in the cells as well, as measured by increases

[3]This section is based on the presentation of Mark Cockett, Vice President, Applied Genomics, Bristol-Myers Squibb.

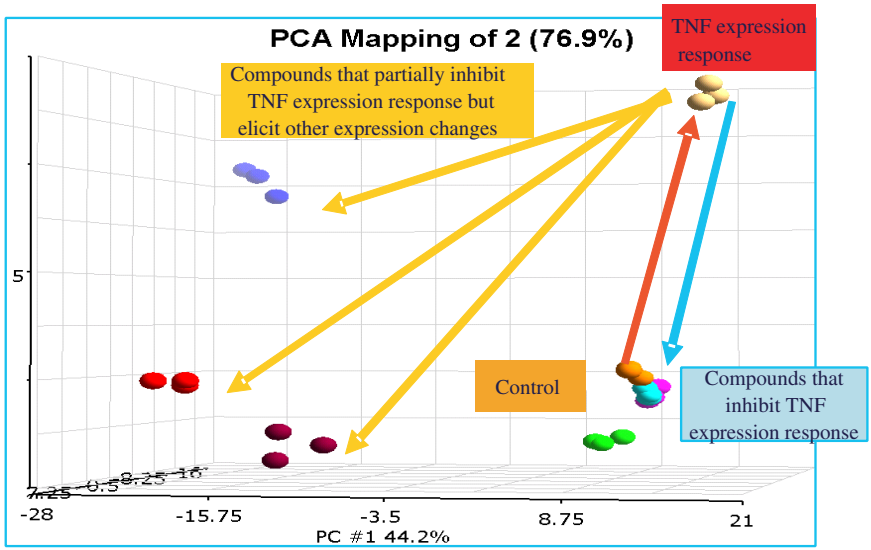


FIGURE 4-2 Use of a principal component analysis (PCA) to identify compounds that reverse surrogate disease phenotypes. In this experiment, a surrogate disease phenotype was created by exposing THP-2 monocyte cells to tumor necrosis factor (TNF). After the disease phenotype had been created, the cells were treated with a variety of drugs to see whether those drugs could inhibit the TNF response. The control treatments (non-TNF-treated cells) are clustered near the bottom, in orange; the TNF-treated cells are clustered near the top of the graph, in yellow; and scattered around the graph are clusters of other colored dots representing the outcomes of various drug treatments. Multiple dots of the same color are experimental replicates. The drug treatments represented by blue, pink, and green completely reversed the TNF response. These were the drugs with the desired response against the surrogate disease. The other drugs represented by purple, red, and maroon not only reversed the TNF stimulation, but also created other effects in the cells as well, as measured by changes in the expression of various different genes.
SOURCE: Cockett, 2007.

in the expression of various genes. The dots representing those drugs are scattered around the graph.

Using this technique, it is also possible to examine what happened in the cells by looking at the effects on individual genes. The TNF treatment caused the expression of MCP-1 to increase sharply, for instance, as would be expected. A number of drugs reversed this effect, bringing the expression of MCP-1 down to control levels. But some of these drugs caused an

increase in the expression of other genes, representing an off-target activity. These are the drugs seen to the left in the PCA plot.

In sum, this type of analysis allows one to look holistically at how various drugs are acting. This capability can provide insights into the mechanisms behind different off-target activities and help in deciding whether these activities are desirable or not, information that in turn can be used to help guide drug selection.

How BMS Uses Toxicogenomics

One key to revolutionizing the current drug development paradigm is for organizations to commit to providing the training and technology upgrades necessary to enable applications of toxicogenomics. Toxicogenomics has been integrated into much of the drug discovery and development work being done at BMS. The company has trained toxicologists and pathologists in how to understand, analyze, integrate, and communicate transcriptional profiling data. Furthermore, the company enhanced its informatics infrastructure to enable its scientists to use the technology. Scientists at BMS use a number of tools, including the Rosetta Resolver, an analysis system for gene expression data from Rosetta Inpharmatics, and the Iconix DrugMatrix database, as well as a number of tools developed in house. Having a variety of tools with which to analyze the data is useful because it enables scientists to look at the experimental system from multiple angles. The various tools also have different abilities to distinguish the signals from the noise.

With the goal of learning how toxicogenomic data compare with results achieved through standard toxicology assessments, BMS now includes transcriptional profiling as part of routine toxicology assessments and prior to conducting GLP (good laboratory practices) safety studies. Roughly 40 percent of the nonclinical toxicogenomic studies at BMS employ this additional assessment tool, and approximately 60 percent of those toxicogenomic studies have been aimed at investigating the mechanisms of toxicity in molecules in which toxicity has already been observed.

To date, toxicogenomics has proved valuable to BMS in a number of ways. It has been useful in identifying pharmacological markers, and in studying on-target versus off-target activities and the tissues in which these activities occur. Sometimes pharmacological events involving tissues in rodents are very different from what occurs in humans; these differences can be investigated by looking at gene expression data from human tissues. Toxicogenomic studies have aided in the understanding of mechanisms of toxicity, particularly in those cases in which a compound can be classified as similar to a control compound class in the Iconix Drug-

Matrix database. When such a match occurs, researchers can expect that the pharmacological and toxicological activity of the new compound will be similar to that of the compound class in the database, making it possible to decide how to proceed without the need for more experiments.

Potential pharmacological biomarkers are identified in about 30 percent of the studies. These biomarkers are in a target pathway, they can be modulated across the toxicity target tissues and other tissues, and they are often known markers of effect. Furthermore, the response correlates with expected efficacious concentrations, and often the story one can tell is biologically compelling. In short, about a third of the time, BMS researchers find that they can gain a biological understanding related to the known literature, understand the biology of a toxicogenetic experiment in a rat, and proceed rapidly to the next stage of development.

Global Transcription as a Marker for Effect

Cockett described a technique for looking at global effects on the transcriptome that involves graphing all the genes in an experiment on a single plot showing how much they have been changed, either repressed or induced (see Figure 4-3). In the resulting diagram, the marks at the bottom in green represent genes that were repressed, while those on the top in red represent genes that were induced.

In this particular experiment, the researchers found that 221 genes, or 1.4 percent of the transcriptome, had changed. This is actually not a large percentage, since with significance set at $p < 0.01$, 1 percent of the genes will meet this cutoff as the result of random chance. To determine the relevance of this percentage that is just slightly greater than random chance, a second experiment was conducted, involving two different dose levels—a low dose of 10 mg/kg and a high dose of 50 mg/kg. The low dose yielded a 1.3 percent change in gene expression (slightly greater than random), while the high dose yielded an 8.7 percent change—that is, 8.7 percent of all the genes that showed up on the chip had been induced or repressed as a result of the experiment. The low dose in this experiment correlated very closely with the NOAEL (no observed adverse effect level) for that drug. The 10 mg/kg dose showed no adverse effects in the rat, while the 50 mg/kg dose clearly did. Thus the researchers concluded that analyzing gene expression at a global level in this way can provide types of information similar to those gleaned during standard toxicology assessments.

Different compounds can display very different patterns when viewed from this global perspective (see Figure 4-4), and this information can be used to make decisions on whether to continue developing a drug. The researchers tested four compounds in an attempt to draw further conclu-

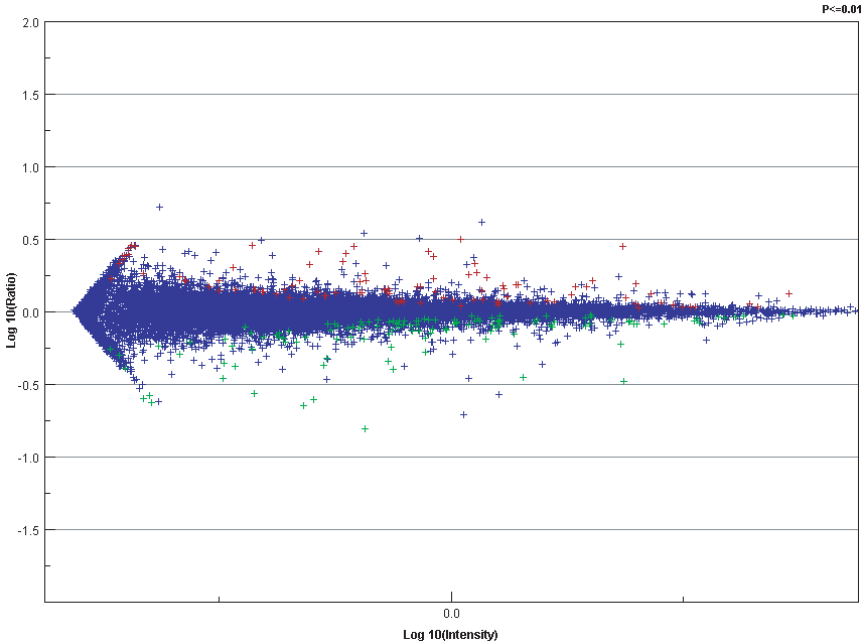


FIGURE 4-3 Observation of global changes within the transcriptome. By graphing all the genes in an experiment on a single plot, as shown above, one is able to visualize globally the extent to which genes were repressed or induced. The marks at the bottom in green represent genes that were repressed, while those at the top in red represent genes that were induced. Within the liver, 222 or 1.4 percent of transcripts changed at $p < 0.01$ level.
SOURCE: Cockett, 2007.

sions from this technology. The first compound resulted in expression changes in 11.4 percent of the measured genes. Such a high percentage of change was indicative of a nonspecific compound that was hitting multiple targets and causing a great deal of transcriptome change.

The second compound resulted in a 2.9 percent gene expression change, which was associated with potent pharmacology as well as myopathy. Although this was not a large percentage change, certain specific effects of the compound needed to be explored.

The third compound led to only a 1.4 percent change. This was a highly selective compound with no obvious off-target effects—there was no toxicology in rodents dosed with the compound, and very little in the transcriptome. The molecule subsequently failed, however, because it had a cardiac liability with an ion channel. The lesson here is that toxicoge-

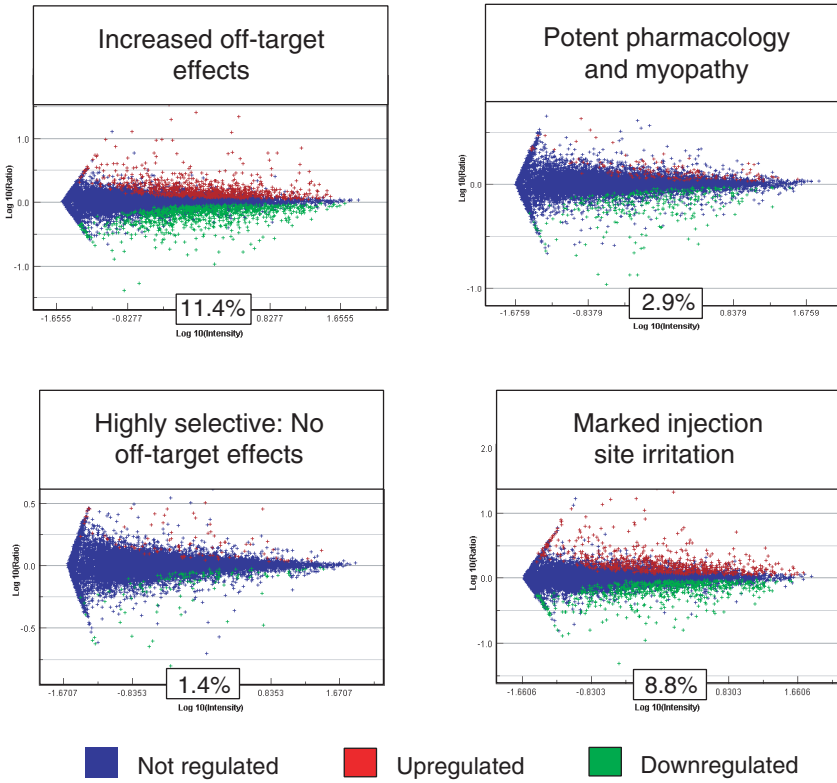


FIGURE 4-4 A global transcriptional profile as a biomarker for NOAEL (no observed adverse effect level). In attempt to draw further conclusions from this technology, four compounds were tested. The first resulted in expression changes in 11.4 percent of the measured genes and was indicative of a nonspecific compound that was hitting multiple targets; the second resulted in a 2.9 percent expression change, which was associated with potent pharmacology as well as myopathy; the third led to a 1.4 percent expression change and was determined to be a highly selective compound with no obvious off-target effects; and although the fourth resulted in an 8.8 percent expression change, it was determined that these changes were associated with an acute phase response at the site of injection, which was largely irrelevant to the pharmacology of the drug and absent when the drug was delivered via a different route.
SOURCE: Cockett, 2007.

nomics cannot detect everything. In this case, for instance, the assay may not have been looking at the right type of tissue to discern the ion-channel effect.

Finally, the fourth compound resulted in an 8.8 percent gene expression change. A large number of genes were changing, but in this case it

turned out that they were associated with an acute phase response at the site of injection, which was largely irrelevant to the pharmacology of the drug and absent when the drug was delivered via a different route. Thus it is possible to be misled in other ways beside missing a toxicity that exists, and it is important to be careful in interpreting these sorts of experiments.

After combining all of its toxicology experiments, BMS found that whenever there was a greater than 3 percent transcriptome change, there was also a clear pathology present; furthermore, most of the compounds with no pathology caused much less than a 3 percent change. Therefore, a broad rule of thumb emerged that a 3 percent transcriptome change represented a crude cutoff point for where one could expect to observe pathology.

Thus BMS has learned to use global transcriptional profiling in its drug safety work. Generally speaking, increases in transcriptional change correlate with increasing pathology and increasing dose, and a level of transcriptional change greater than 3 percent suggests drug-related pathology. Profound transcriptional change—at a level of 7 percent or greater—is usually associated with multiple toxicities, and it is often problematic to interpret these data because it is difficult to disentangle the many phenomena involved. On the other hand, minimal global change—less than 3 percent—is not an assurance of drug safety, but it is suggestive of at least a pharmacological specificity, and one must look at the specific genes and the pathways that are modulated to understand the response in greater detail. Finally, transcriptional changes are distinct from histopathology. They may be less sensitive, or they may arise from pathology elsewhere, as in the case of a liver transcriptional readout of an acute phase reaction in skin. In such situations, one can be misled by a set of gene changes in one tissue responding to changes in another.

TOXICOGENOMICS AT ABBOTT LABORATORIES

At Abbott Laboratories, toxicogenomics is increasingly being integrated into the drug discovery and development process. Brian Spear, the company's Director of Genomic and Proteomic Technologies, explained that information gained from studying changes in gene expression can be of value at four different levels of discovery and development:

- Identifying toxicological issues early prior to large financial investments
- Selecting compounds least likely to fail because of toxicological issues
- Understanding mechanisms of toxicity

- Bridging the preclinical and clinical data by understanding the mechanisms and commonalities in responses

Selecting Compounds[4]

Abbott uses gene expression to try to predict whether a compound is likely to be a hepatotoxicant. Gene expression assays are used to determine not whether a compound is safe, but whether it has a high enough chance of being hepatotoxic that discontinuing its development would be advisable. In short, Abbott uses this technology for lead optimization rather than for safety assessments.

Before the researchers could begin using this technology, they had to develop a hepatotoxicity reference set by administering various doses of multiple known hepatotoxicants and multiple known nonhepatotoxicants to rats, and then observing gene expression changes in a set of 40 genes. The researchers established that different patterns of gene expression are elicited by acute hepatotoxicants, moderate hepatotoxicants, and nonhepatotoxicants (see Figure 4-5). Therefore, when a new compound is administered to rats, the resulting gene expression profile can be compared with that resulting from the reference compounds to determine whether the profile matches that of a severe hepatotoxicant, a mild hepatotoxicant, or a nonhepatotoxicant. Specifically, the similarity between the new compound's gene expression pattern and that of the test set is analyzed with a pattern-recognition algorithm based on a neural network, and the degree of similarity is reduced to a numerical score. In the example described by Spear, the scores were on a scale of 1 to 4, where 1 indicated severe hepatotoxicity, 2 moderate hepatotoxicity, 3 mild hepatotoxicity, and 4 no evidence of liver injury.

After establishing this predictive reference set, the researchers tested it to see whether it was actually predictive for rats. Using 278 different expression profiles with multiple drugs, multiple times, and multiple doses, they compared the predictions for these various treatments with what was already known about the compounds in the literature. (All the test compounds in this case were ones for which information existed in the literature.) Among the 278 expression profiles, the predictive assay yielded the same numerical score as the literature in 246 cases. In another 21 cases, the assay's score was within 1 of the score in the literature—for example, 3 instead of 2. Thus for 267 of the 278 expression profiles, the assay agreed closely—and often exactly—with the results reported in the literature on the degree of hepatotoxicity to be expected.

The next step was to test new compounds with the assay to see

[4]This section is based on the presentation of Dr. Spear.

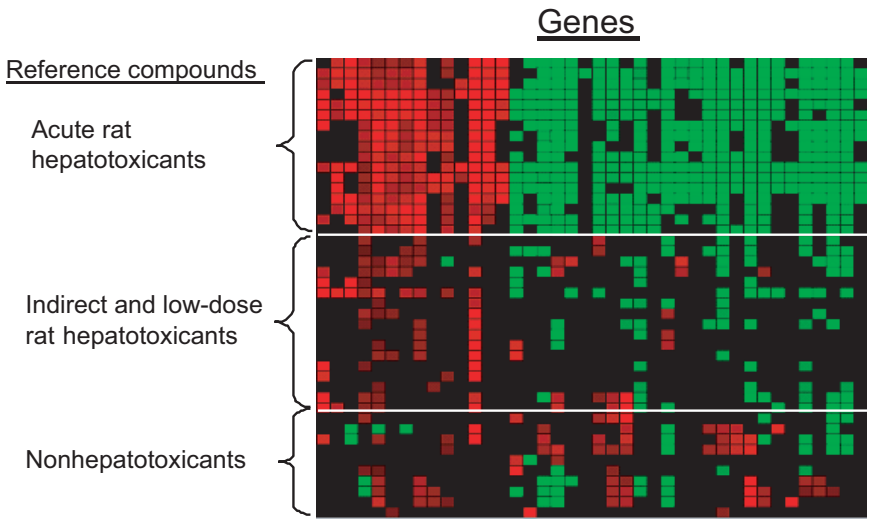


FIGURE 4-5 Gene expression profile of a hepatotoxicity reference set. This figure exhibits gene expression patterns elicited by acute hepatotoxicants, moderate hepatotoxicants, and nonhepatotoxicants. When a new compound is administered to rats, the resulting gene expression profile can be compared with that resulting from the reference compounds to determine whether the profile matches that of a severe hepatotoxicant, a mild hepatotoxicant, or a nonhepatotoxicant
SOURCE: Spear, 2007.

whether it could be used in a predictive way. The first trials compared how well the gene expression patterns at 2 weeks correlated with the results of histopathology and clinical chemistry at 2 weeks. For these trials, the researchers used a cutoff score of 2.5 to rate the gene expression patterns, so that any compound scoring below 2.5 was said to be positive for hepatotoxicity and any compound scoring above that level was negative—a simple yes/no score. The assay was 100 percent accurate: it correctly predicted six of six negatives and two of two positives. But since the gene expression patterns, like the histopathology and clinical chemistry studies, were from the 2-week time point, the results were a bit like predicting the present. The real question was whether gene expression patterns observed earlier in the process could predict hepatotoxicity prior to its physical manifestation.

To answer this question, the researchers compared the results of short-term gene expression assays—performed 3 or 5 days after exposure—with the results of 2-week toxicology studies. Again using yes/no scoring, the short-term assays correctly predicted 50 of 52 hepatotoxicity results, or 8

of 9 positive outcomes and 42 of 43 negative outcomes. This added up to a specificity of 97.7 percent, a sensitivity of 88.9 percent, and an overall accuracy of 96.2 percent, which meets Abbott's needs. Spear explained that the accuracy need not be 100 percent, just high enough to provide sufficient confidence that work on a compound should be discontinued based on 3- or 5-day exposure data, without the need for expensive and time-consuming animal studies.

Abbott now uses this assay regularly for screening new compounds. In one case, for instance, three compounds were examined for a project looking at kinase inhibitors. One of the three compounds scored very low on the assay, implying severe hepatotoxicity, and work on that compound was discontinued. The other two had scores in the mild or nonhepatotoxic range, and work on them moved forward.

Spear offered several conclusions and lessons learned from Abbott's experience with these gene expression assays:

- The accuracy of the assay must be established, and while it need not be 100 percent, it must be good enough to establish sufficient confidence for decision-making purposes. When multiple compounds are being considered, about 96 percent accuracy is sufficient.
- The validity of Abbott's gene assay applies only to rats; the assay has not been validated in humans.
- Conventional toxicology and pathology remain the gold standard, and it is necessary to compare what happens in the animals with what happens in the assay. If the two sets of results conflict, the toxicology is considered correct, and the assay must be reworked.
- The value of such an assay is greatest during lead optimization, prior to candidate selection. The assay is most useful when there are multiple compounds involved and the project team needs help in making decisions about which ones to pursue.

Understanding Mechanisms of Toxicity[5]

A second application of toxicogenomics is to help elucidate the mechanism of toxicity once a compound has shown toxicity in rats or in some other preclinical model. It is important to understand the mechanism involved and to know whether it can be screened for. Toxicogenomics can be useful for this purpose because various gene expression patterns have been associated with specific mechanisms. An important caveat is that toxicogenomics should be relied upon not as a way of identifying mechanisms of toxicity, but as a way of generating a hypothesis that

[5]This section is based on the presentation of Dr. Spear.

can be tested with more conventional approaches. Thus the technology's value lies in its ability to help obtain an answer more quickly.

An example is a drug that showed cardiotoxicity in rats in a 2-week study at a high dose—200 mg/kg—with myocardial degeneration and necrosis. At lower doses of 30 or 80 mg/kg, however, there was no evidence of cardiotoxicity. Since the researchers had no serum protein or other biomarker with which to monitor the toxicity, they turned to gene expression patterns in an attempt to understand the mechanism of toxicity. Among rats in the low-dose treatment groups, there were very few gene expression changes in the heart, while rats given 200 mg/kg for 5 days showed striking gene expression changes. What was most interesting was that in rats given 200 mg/kg for 1 day, there were no physical signs or symptoms of cardiotoxicity, but the gene expression pattern was quite similar to that seen in the rats given the high dose for 5 days. Thus the gene expression pattern was an early indicator of gene expression changes related to cardiotoxicity.

To explore these results further, the researchers looked at the particular genes that were up- and down-regulated, and were able to determine that a number of the genes were related to mitochondrial impairment. Some were mitochondrial function genes; these appeared to be down-regulated. Others were genes related to oxidative stress; these were up-regulated. Accordingly, the researchers hypothesized that the compound was inhibiting mitochondrial function, and designed experiments to test this. The first test was to treat the animals with the compound for 4 days and then remove their mitochondria and determine the mitochondria's oxygen consumption. The mitochondrial oxygen consumption in the treated rats had been reduced to a degree that was comparable to that seen with doxorubicin, another cardiac toxin. In an *in vitro* experiment in which mitochondria were isolated from cardiac tissue and then treated with different compounds, the test compound was also seen to result in mitochondrial inhibition. Thus the researchers concluded that the compound was likely to be a mitochondrial toxin.

Spear pointed out that the gene expression assay was used to generate a hypothesis—that the mechanism of toxicity was inhibition of mitochondrial function—but other tests were then used to test this hypothesis. The test for toxicology is still clinical chemistry and histopathology, and gene expression studies are not going to replace *in vivo* toxicological studies. While gene expression studies may shorten the path to an answer, conventional studies will still be necessary.

Gene Expression Profiling in Early Discovery Studies[6]

Dr. Blomme elaborated on the toxicogenomics work being done at Abbott, describing the added value of rat exploratory toxicology studies.

During the lead optimization process, molecules are characterized with a battery of *in vitro* and *in vivo* assays to evaluate various physical, chemical, pharmacological, metabolic, pharmacokinetic, and toxicological properties. If assays are to be used to help make go/no go decisions early in the development process, they must have two important characteristics: they must utilize limited quantities of compound (milligram to gram range) since at this stage compound availability is a major limitation, and results need to be delivered rapidly.

Generally speaking, efforts to reliably evaluate physical, chemical, pharmacological, metabolic, and pharmacokinetic properties during the lead optimization process have been successful, but the same cannot be said for toxicology. Consequently, Abbott came up with the concept of using short-term rat studies to study toxicology during the lead optimization process.

Ideally, for these studies to be useful in the lead optimization process, they should last no more than a few days, use limited numbers of animals, and be performed with 2–4 grams of compound, a quantity Abbott researches have found to be sufficient. Traditional dose range–finding studies involve five animals per group, dosing for 7 or more days, and more than 10 grams of compound. Requiring less compound can translate into studies being completed earlier in the development process.

Traditional toxicology end points include clinical pathology and histopathology; after a short period of dosing, however, these analyses cannot predict toxic events consistently. Furthermore, when only a small number of animals are used, the pathological results are often difficult to interpret. Predictive toxicogenomics is a valuable technology because it has greater sensitivity than traditional methods. As explained earlier in this chapter, gene expression changes typically occur before the functional and morphological changes that are detected by histopathology or clinical pathology.

The increased sensitivity of predictive toxicogenomics implies that gene expression studies should theoretically make it possible to dose animals for shorter durations, and the literature suggests that in some cases, gene expression changes can be observed within hours or perhaps 1 day of administering a toxicant. However, this is the exception rather than the rule. For many compounds, a steady state in tissue kinetics

[6]This section is based on the presentation of Eric Blomme, Project Leader in Cellular, Molecular, and Exploratory Technology, Abbott Laboratories.

must be achieved before gene expression changes can be measured reliably. In particular, after only 1 day of exposure, gene expression changes can vary greatly among individuals, making interpretation challenging. Further dosing generally leads to less variability and more reliable interpretation.

To illustrate, Blomme described work on developing a signature for bile duct hyperplasia, done in collaboration with Iconix Pharmaceuticals. The goal was to predict bile duct hyperplasia in the liver of rats using liver gene expression profiles from either 1 or 5 days of dosing. Typically it takes several days for bile duct hyperplasia to occur and to be visible morphologically to pathologists.

The signature was derived from a training set of DrugMatrix reference profiles with the same procedures described earlier by Dr. Halbert. The researchers then evaluated the ability of that signature to detect bile duct hyperplasia in rats by exposing rats to a set of 10 compounds not included in the training set. Rats were treated for 1, 5, and 28 days with daily doses of one of the 10 compounds. Gene expression profiles were created from the livers of the rats treated for 1 or 5 days, while the livers were examined histopathologically for the rats treated for 28 days to determine whether treatment with a particular compound had led to bile duct hyperplasia. Then the researchers looked at how often the signature correctly predicted the presence or absence of bile duct hyperplasia after 28 days of dosing.

Using the expression profiles generated after 5 days of dosing, the signature correctly predicted the occurrence of bile duct hyperplasia after 28 days in all cases—seven of seven positive compounds and three of three negative compounds. Using the 1-day gene expression profiles, however, the signature was much less successful, predicting only three of seven positive and three of three negative compounds. Thus it can be concluded that prediction in a single-dose study is not reliable, and that 3 to 5 days of dosing will generally be necessary to enable gene expression profiles to predict outcome reliably.

Another advantage of gene expression profiling is that, in general, fewer animals are necessary. Abbott researchers have found that gene expression profiling of compounds in animals is not as variable as many of the traditional biomarkers, such as histopathology or clinical pathology. Thus for their short-term toxicology studies, the researchers are confident in making decisions using only three animals per group. A significant reduction in numbers of animals corresponds to a significant reduction in the amount of compound required for the studies.

A third advantage of gene expression profiling in the context of short-term studies is the ability to generate mechanistic data that are useful in understanding changes detected by other means. Since short-term explor-

atory studies use limited numbers of rats per group, limited numbers of groups, and doses that are not always optimized, the data are often quite challenging to interpret. By gaining a better understanding of changes, it becomes possible to make better predictions about their progression and significance.

An example is a compound that was given at four different doses up to 300 mg/kg/day, which after 5 days of dosing led to a dose-dependent increase in liver weight. Because many drugs on the market cause an increase in liver weight, particularly at high doses, the researchers sought to determine the toxicological significance of this finding.

Using a gene expression–based artificial neural network algorithm to predict the hepatotoxicity potential, the researchers found that after prolonged dosing at 200 or 300 mg/kg per day, the compound would likely become toxic in the rats. Therefore, to perform 2- or 4-week studies, they would have to use doses lower than 200 mg/kg. They then used the DrugMatrix database to evaluate the gene expression profiles at the various doses. At doses greater than 200/mg/kg/day, there was a significant correlation with several gene expression profiles in the database that were induced by hepatic toxicants, such as dipyrone and econazole. Next the researchers tried to determine which pathways were being affected by the compound. The data indicated that when doses greater than 200 mg/kg/day were administered, several toxicologically relevant pathways were affected, including oxidative stress, cholesterol biosynthesis, and aryl hydrocarbon receptor signaling. This finding provided additional evidence that at doses greater than 200 mg/kg/day, the compound would result in rat hepatotoxicity. Thus a dose of 100 mg/kg/day was selected for the subsequent 2-week rat toxicology study.

According to Blomme, these examples demonstrate that gene expression profiling can be a valuable addition to early discovery studies. It is a sensitive and specific indicator of toxicity. It is associated with less inter-individual variability, making it possible to use fewer animals and thus to conduct studies with less compound. And it adds a level of mechanistic information that is quite useful in improving the interpretation of findings of short-term studies. Abbott researchers are using this technology to assess the toxicity of compounds and make go/no go decisions about their advancement.

SUMMARY

There are various uses of and methods for conducting gene expression analyses to help predict the toxic effects of compounds and provide insights into the mechanisms of toxicity. Gene expression can be predictive and, in particular, can be more sensitive than traditional approaches;

it was suggested that a level of transcriptional change greater than 3 percent indicates drug-related pathology. Information gained from studying gene expression changes can be used to

- identify toxicological issues early prior to large financial investments;
- select the compounds that are least likely to fail because of toxicological issues;
- understand mechanisms of toxicity; and
- bridge preclinical and clinical data by understanding the mechanisms and commonalities in responses.

These gene expression assays, however, apply only to rat models, and the next challenge is to transition from prediction of toxicity in rats to prediction of toxicity in humans.

5

Screening Technologies III: Metabolomics

Metabolomics, like genomics and proteomics, can be used to assess drug safety. Rather than patterns of gene expression or protein–protein interactions, metabolomics (the characterization of small molecule metabolites produced in response to particular stimuli) is used to study the effects of drugs on various biochemical pathways. According to Klaus Weinberger, Chief Scientific Officer at Biocrates, one advantage of metabolomics over other approaches is that scientists currently have a stronger qualitative understanding of underlying biochemical pathways than of protein–protein interactions or interactions at the transcription level. The presentations addressing metabolomics illustrated how the technology is being used to gather information on toxicities and their underlying mechanisms. They highlighted four categories of metabolites that can provide insights at varying levels of complexity:

- Markers for the activities of single enzymes
- Direct multiparametric markers, which can indicate lipid elevation or lowering, metabolic control, insulin sensitivity, or inflammation
- Multiparametric surrogate markers, which offer details about questions that are difficult to analyze directly, such as gluconeogenesis/glycolysis, oxidative stress, and tissue damage and apoptosis
- Mode-of-action markers, which indicate the presence of such responses as lipid signaling and regulatory metabolites

METABOLOMICS AT METABOLON[1]

Dr. Milburn discussed some of the advantages of studying metabolites as opposed to proteins or gene expression. Biochemical molecules are the end result of many biological processes, and they can reflect the impact of a number of factors, such as the environment, a patient's overall health, and any drugs a patient might be taking. While the technology used to analyze the human metabolome is complex, the metabolome is smaller than the genome or the proteome. According to the most recent estimates, there are only about 2,400 metabolites in the human body—significantly fewer than the approximately 25,000 genes, 100,000 transcripts, and millions of proteins with which other fields must work.

In a sense, metabolomic analysis can be thought of as an expansion of the traditional diagnostic tests performed on blood or urine and used to measure the levels of, for example, blood urea nitrogen, creatinine, and glucose. While these molecules represent a small portion of the total biochemistry of the body, the aim of metabolomics is to look at all, or at least a large proportion, of the body's small molecules.

The Metabolon Process

At Metabolon, the goal is to be able to identify and quantitatively measure all of the small molecules in any sample type—urine, blood, tissue, or cell extract. The process used is illustrated in Figure 5-1. Sample preparation begins with four different fractionation steps to extract all polar and nonpolar molecules with a mass of 50–1,500 daltons. Once these small molecules have been separated out through these four extraction steps, they are pooled back together, and that sample is then split for analysis by two different platforms—a liquid chromatography–mass spectrometry system (LC-MS) and a gas chromatography–mass spectrometry system (GC-MS). Company scientists use both of these platforms because small molecules can be very polar as well as very nonpolar; the two chromatography methods work well together for profiling of most of the small molecules in the samples.

Metabolon has developed proprietary software that makes it possible to identify automatically all the ions that are scanned by the spectrometers. Using automated processing techniques based on the biological variation of the compounds within samples, the researchers are able to reconstruct the original molecules to which the ions belonged before going through the system. With the help of a standard chemical library,

[1]This section is based on the presentation of Michael Milburn, Chief Scientific Officer, Matabolon.

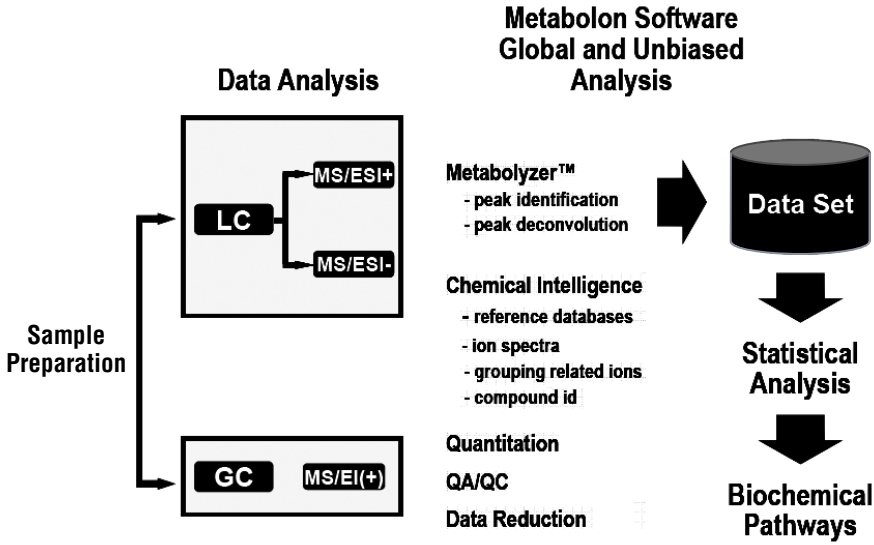


FIGURE 5-1 Overview of the Metabolon process. Once the sample has been prepared, it is analyzed by a liquid chromatography–mass spectrometry (MS) method (with and without electrospray ionization [ESI]) and a gas chromatography–mass spectrometry method (with electron ionization). Proprietary software is used to analyze the data, and identify the molecules present in the sample and quantify their amounts.
NOTE: QA = quality assurance; QC = quality control.
SOURCE: Milburn, 2007.

the molecules are identified, and their amounts are quantified. The end result is a data set that identifies all the small molecules seen in the sample and their relative amounts.

Using these techniques, Metabolon is able to detect and analyze metabolites that capture the vast majority of the biochemistry that occurs in the human body. Company scientists predict that after refining their methods, they will be able to identify even more molecules.

Examples

Several examples illustrate how the technology is used. The first is a study carried out with Bristol-Myers Squibb researchers who were interested in different HIV protease inhibitors, many of which have a side effect of lipidystrophy, a degenerative condition in the body's adipose tis-

sue. Metabolon researchers exposed cultured liver and fat cells to the different HIV protease inhibitors and conducted a metabolomic analysis.

The results were displayed in scatter plots (see Figure 5-2). The dots on each horizontal line represent the readout of a particular molecule. Since the experiment was run a number of times, there are many readings for each metabolite, and these are plotted along the horizontal axis in terms of their Z scores, where the zero point is the mean value of the control group for that particular molecule. Thus the amount of scattering in the scatter plot—how tight or how loose it is—offers a rough visual measure of the extent to which the total biochemistry of the cells was perturbed by the compound under examination.

The scatter plots of the different HIV protease inhibitors display very different levels of perturbation. The earlier protease inhibitors, such as lopinavir and nelfinavir, tend to show much more overall biochemical perturbation, while the newer ones show less. Atazanavir caused the least amount of perturbation and was most similar to the vehicle control group. This result agreed with the clinical data, which indicated that some of the earlier protease inhibitors had a greater lipidystrophic side effect; the newer compounds, which had been specifically developed to have less of that effect, did indeed have fewer lipidystrophic side effects.

Following this global analysis, the researchers tried to identify which compounds were being affected by the drugs and which pathways were being altered. In the liver cells, they found an increase in metabolized biochemicals produced in fatty acids—fatty-acid triglyceride metabolites. In the fat cells, on the other hand, they found an impairment of the Krebs cycle intermediates. Thus the analysis implied that the drugs were causing a large change in the energy metabolism of the fat cells.

A second example illustrates how this metabolomics technology can be used to examine a drug's mechanism of action. The study was performed using an oncology drug in a myeloma cell line. This drug was known to induce apoptosis in about 1 day, but the mechanism of apoptosis was unclear. A simple study was conducted to look for significantly altered compounds in cell cultures treated with the drug or with a control at four time points spread over 27 hours and with six cultures per group.

The number of biochemicals that were significantly changed (with $p \leq 0.1$, $q \leq 0.2$) rose over time and then leveled off after about 20 hours, at 65 compounds. The amount of change was more than twice as great as was commonly seen in such experiments—an indication that there had been very large perturbations of biochemistry because the cells had gone into an apoptotic state.

By examining the individual biochemicals whose levels were altered, the Metabolon researchers were able to identify those metabolites that were most strongly affected. The level of sorbitol, for example, rose

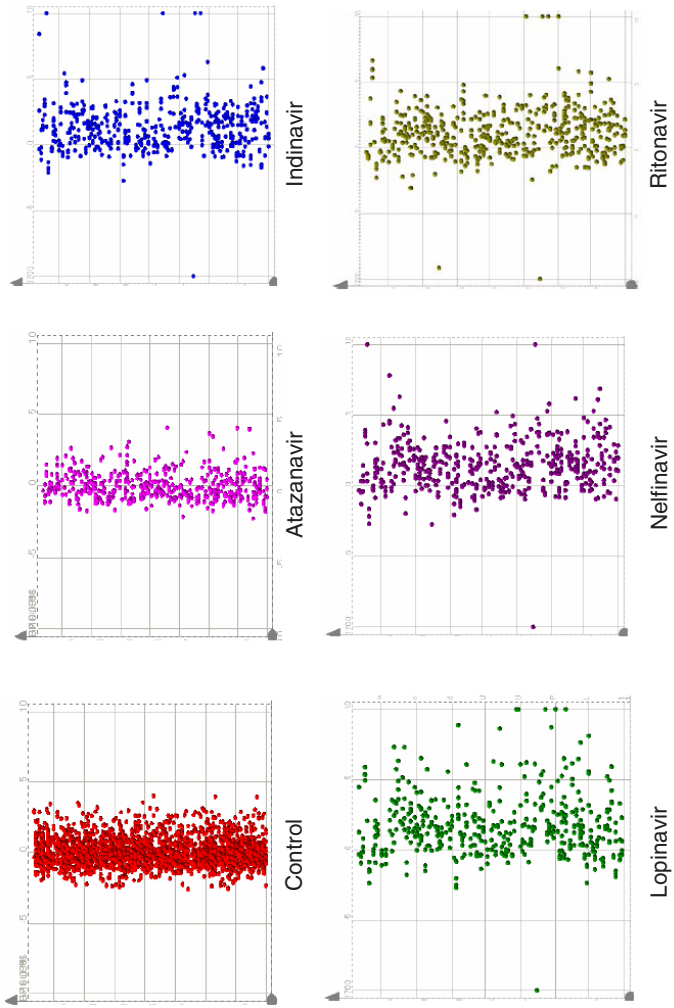


FIGURE 5-2 Global metabolomic analysis of commercially available HIV protease inhibitors. The scatter plots displayed above, depending upon how tight or how loose the scatter is, indicate different levels of cellular perturbation. Earlier protease inhibitors, such as lopinavir and nelfinavir, show more overall biochemical perturbation, while the newer protease inhibitors show less. Atazanavir caused the least amount of perturbation and was most similar to the vehicle control group.
SOURCE: Milburn, 2007.

steadily throughout the experiment, increasing by 3-fold at 13 hours and by 23-fold at 27 hours. Another compound, fructose-1-phosphate, performed similarly, increasing steadily throughout the course of the experiment until it was up 30-fold at the end of the 27 hours.

From changes seen in the individual biochemicals, the researchers could also identify biochemical pathways that were being affected—both the subpathways and the superpathways. Among the superpathways, for instance, carbohydrate metabolism and lipid metabolism were strongly altered. Because the study drug clearly up-regulated the level of sorbitol, a molecule known from a number of previously published studies to induce apoptosis, the researchers attributed the mechanism of apoptosis to an increase in sorbitol.

Summary

As indicated by the above examples, metabolomic analysis is an efficient and valuable technology. With a turnaround time of 3–4 weeks, these studies can be performed relatively quickly. Further, this technology makes it possible to obtain highly specific data by analyzing biochemicals individually. These data can help in evaluating the side effects of test compounds, as well as in understanding mechanisms of action and of toxicity.

METABOLOMICS AT BIOCRATES[2]

Weinberger's presentation paralleled that of Milburn in a number of ways, addressing various means by metabolomics can be used to help ensure drug safety.

The Biocrates Process

The technology platforms used by Biocrates and Metabolon are similar. The Biocrates platform provides fully automated sample preparation, mass spectrometric identification, and quantitation; bioinformatics is used for technical validations, visualization of statistics, and biochemical interpretation; and the entire process is based on an in-house bio bank or on samples collected from the partners with which Biocrates works.

The analytical portfolio includes more than 1,000 annotated metabolites encompassing the main areas of intermediary metabolism. It contains primary and secondary amines, such as proteinogenic and nonproteinogenic amino acids, acylcarnitines and free carnitine, reducing monosac-

[2]This section is based on the presentation of Dr. Weinberger.

charides and oligosaccharides, phospholipids, glycolipids, prostaglandins, bile acids, and many other metabolites.

The technology can be applied to basic research, the agriculture and nutrition industries, clinical diagnostics and theranostics, and pharmaceutical research and development. Pharmaceutical applications include studies of drug metabolism and pharmacokinetics, safety and toxicology, and pharmacodynamics and efficacy.

Four categories of metabolites have been established, all of which offer insights at differing levels of complexity:

- Biomarkers for the activities of single enzymes, which are relatively straightforward, as a simple ratio of product concentration to substrate concentration will generally provide an idea of the quantitative activity of a single enzyme
- Direct multiparametric markers, or groups of markers that can indicate lipid elevation or lowering, metabolic control, insulin sensitivity, or inflammation
- Multiparametric surrogate markers, or groups of markers that offer details about questions that are difficult to analyze directly, such as gluconeogenesis/glycolysis, oxidative stress, and tissue damage and apoptosis
- Mode-of-action markers, or markers that indicate the presence of such responses as lipid signaling and regulatory metabolites

An Example

To illustrate the utility of metabolomics, Weinerger described a study of puromycin-induced toxicity. The study was conducted using four groups of six Sprague-Dawley rats: a vehicle control, and a low-dose (10 mg/kg), a medium-dose (20 mg/kg), and a high-dose (40 mg/kg) group. The researchers were blinded to the test compound given to the rats, so they did not know it was puromycin. Histopathology at 3 weeks revealed no damage in the control or the low-dose group, and only moderate nephrosis in the medium-dose group. The high-dose group, by contrast, developed end-stage renal disease after only 2 weeks and had to be sacrificed at that point. Plasma and urine samples were taken on days 3, 7, 14, and 22 in the first three groups and on days 3, 7, and 14 in the high-dose group. These samples underwent a metabolomics analysis that was correlated with histopathology, pathophysiology, expression profiling, and proteomics.

The analysis revealed a marker for general tissue damage. Acylcarnitines are compounds that are produced in the mitochondria of energy-metabolizing cells, and in healthy tissue they remain inside these cells, so

that there are very low levels of circulating acylcarnitine in plasma. The metabolomic analysis showed that in the low-dose group, acylcarnitine plasma levels were very similar to those in the control group, but in the medium-dose and high-dose groups, there was a clear time-dependent increase in the circulating acylcarnitine levels, which implied tissue leakage and, in particular, mitochondrial damage.

The observance of general tissue damage prompted the researchers to seek to identify where the damage had occurred, and they checked for known markers of organ-specific toxicity. In testing for hepatotoxicity, they could not show a significant change in the bioassays for the bile acids, so they concluded that the liver was not the main site of the tissue damage.

The analysis did, however, identify a number of markers for a wide variety of kidney-specific outcomes and mechanisms. It identified markers for

- moderate polyuria and, in at least half a dozen compound classes, tubular dysfunction;
- general inflammation and oxidative stress, such as dose-dependent activation of COX, 12-LOX, and 15-LOX, although there was no sign of systemic oxidative stress; and
- time-dependent moderate ketosis and the de-repression of NO synthase.

In contrast with previously published, studies, in which tryptophan depletion was found to be due to increased synthesis of kynurenine, the markers showed that in this case, the tryptophan depletion was due mainly to conversion to serotonin, which implied that there was an additional vasoconstrictor in this model. Using the information gained from all of these markers, the researchers were able to form a biochemically functional and plausible model of what was taking place in the study.

Summary

As evidenced by the above example, metabolomics can be useful in attempting to determine causes or sites of drug toxicity. Knowing as much as possible about how a drug might affect a specific pathway helps researchers see a more complete picture as they try to formulate answers. An important factor to consider in using metabolomics is heterogeneity of responses. In the above example, the animals used were genetically identical; in clinical settings, there will be widespread genetic variability. Researchers will need to determine whether the observed effects of

metabolomic analysis are great enough to be significantly higher than the biological variability among a population.

Although metabolomics can make a valuable contribution to understanding disease, researchers continue to characterize disease from the perspective of different disciplines (e.g., pathology, physiology, and clinical chemistry). Weinberger asserted that the community must aim to unite these different disciplines in the assessment of molecular pathology. Further, he suggested that throughout the pharmaceutical industry, pharmacology, preclinical, clinical, and toxicology departments should focus on the same question of drug reaction utilizing all available perspectives. Such unification of disciplines could help reinforce evidence-based drug development.

6

Screening Technologies IV: Pharmacogenetics[1]

The screening technologies discussed in the previous three chapters are used mainly to address the issue of whether a particular compound is toxic and if so, why. Dr. Lai raised a different issue: Given a useful drug that is toxic in only a subset of patients, how can those patients be identified so the toxicity can be prevented or at least anticipated? Lai described how pharmacogenetics provided an answer to that question in the case of the anti-AIDS drug abacavir.

ABACAVIR AND THE HYPERSENSITIVITY REACTION

Abacavir is a reverse transcriptase inhibitor used against HIV. It is the sole ingredient in Ziagen, an anti-AIDS drug marketed by GlaxoSmithKline (GSK), and it is also used in combination drugs such as Trizivir, which contains abacavir, zidovudine, and lamivudine. Abacavir is a highly effective medication and is well tolerated in most patients, but a small percentage of people who take it experience hypersensitivity reaction (HSR). HSR is a multiorgan syndrome whose most common symptoms are fever, rash, nausea and vomiting, and malaise or fatigue. The overall rate of HSR among abacavir users is about 5 percent, and most of these are nonserious episodes that are resolved by discontinuing use of the drug. However, the discontinuation must be permanent. If a

[1]This chapter is based on the presentation of Eric Lai, Vice President, PGx Experimental Project Coordination and Analysis, GlaxoSmithKline.

patient who has once experienced HSR starts taking abacavir again, the HSR returns very quickly—in a matter of hours to a day or so—and this time it is lethal.

The HSR phenotype is complex. About 78 percent of HSR patients have fever, about 65 percent have rash, and about 96 percent exhibit fever or rash or both. There is a long list of other symptoms that appear in at least 10 percent of HSR cases: nausea/vomiting, malaise/fatigue, muscle or joint pain, headache, diarrhea, itching, abdominal pain, dyspnea, and cough. Most patients have three or more of these symptoms in varying combinations.

The time of onset is also variable. A number of patients experience HSR within the first week of taking abacavir, sometimes within 24 hours, but for others it takes longer, and the median time to onset is about 11 days. About 93 percent of reported cases occur within 6 weeks of starting abacavir, so one of the exclusion criteria in the GSK studies is that a patient must experience HSR within the first 6 weeks.

In 1999, at the time of abacavir's approval, a two-part postmarket risk management program was established. The first part was aimed at educating health care providers; this included updating labeling information on a regular basis and monitoring the occurrence of HSR among abacavir users. Monitoring data have revealed that although the number of people taking abacavir has increased steadily over the past 8 years—to more than 1 million in 2006—the number of deaths caused by the drugs has remained relatively stable since 2002 (see Figure 6-1). Thus one can infer that physicians now know that once any kind of HSR-related symptoms appear in a patient taking abacavir, the patient must be taken off the drug and never given it again. Despite physician awareness, however, the rate of spontaneous HSR has not decreased, as it is not possible to predict whether a patient will exhibit HSR until abacavir is taken. The second part of the postmarket risk management program included a pharmacogenetics study designed to look for genetic factors associated with abacavir-related HSR.

THE ABACAVIR PHARMACOGENETICS PROGRAM

The goal of GSK's pharmacogenetics program was to identify genetic markers that could predict patients at risk of developing HSR from abacavir and prevent them from taking the drug, thereby improving its benefit-risk balance. The study would involve gathering patients who had developed HSR; matching them with patients who had not; and then performing association studies, first with candidate genes and then later—as it became possible—with whole-genome analysis.

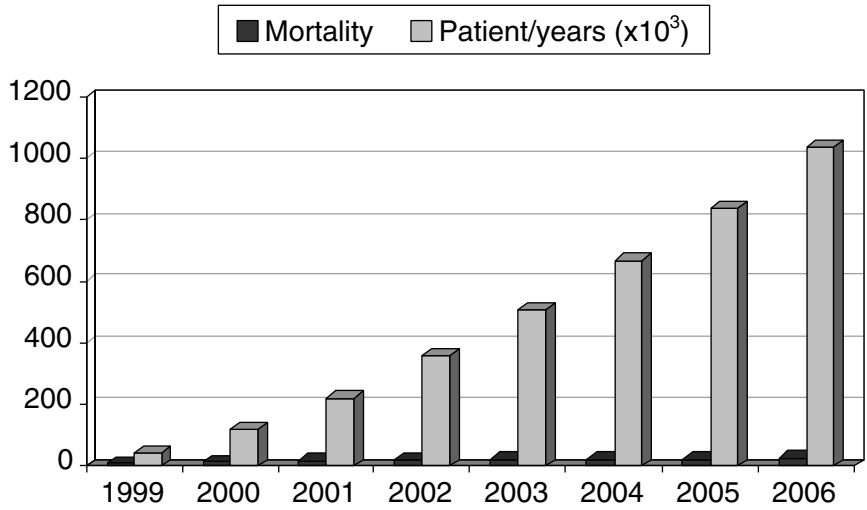


FIGURE 6-1 Cumulative patient/years of exposure to abacavir products and spontaneous reports of HSR-associated mortality among those taking abacavir. These data show that while abacavir use greatly increased, HSR-related mortality remained low.
SOURCE: Lai, 2007.

The Identification of HLA-B57

Initial calculations, derived from such assumptions as the allele frequency of the causative locus and the effect of the locus's being variably penetrant, implied that it would take 500 case-control pairs to power the study adequately. Since the GSK trial included only 44 cases and 78 controls, the researchers did not expect to identify any candidates. In July 2001, however, they discovered that one of the candidate genes, HLA-B, was playing a major role in abacavir-related HSR. Of the 44 cases in the study, 25 (57 percent) had the HLA-B57 variant, while of the 78 controls, only 3 (4 percent) had that same variant. After receiving confirmation of the results of the assays from two other laboratories, the GSK researchers were confident that they had identified a predictive biomarker for HSR.

To follow up on this conclusion, the group continued to accumulate data; they currently have data from 444 cases and 486 controls. They have further zeroed in on the marker—the HLA-B*5701 subtype of HLA-B57—and in their studies, this marker predicted HSR with a sensitivity of 50 percent and a specificity of 98 percent.

These results have been confirmed by a group of researchers headed by Simon Mallal at the Royal Perth Hospital in Perth, Australia, who performed a study with 18 cases and 230 controls and obtained the same results, except with a much higher sensitivity. In 14 of the 18 cases, they found the HLA-B*5701 allele, yielding a sensitivity of 94 percent. Lai hypothesized that this increased sensitivity was related to the fact that in the Australian study, one physician saw all of the patients. Therefore, the inclusion criteria were based on this one physician's diagnoses, and it was possible to follow up with the patients to determine whether they did indeed have HSR. In the GSK study, by contrast, the cases were scattered over several dozen centers, and the only source of information was the case report forms. Thus it was impossible to go back and ask the patients whether they had taken abacavir or whether, for instance, they had ever had a fever or some other symptom.

A second difference was that the GSK study included a number of ethnic groups, and the results differed among groups. The sensitivity among whites, for example, was 50 percent, while that among Hispanics was only 22 percent, and there was no significant association among blacks. The problem may lie in the fact that the frequency of HLA-B*5701 varies greatly among ethnic groups, and the rate of HSR in ethnic groups varies as well. In blacks, for example, the rate is about one-half or one-third the frequency in whites. GSK is now exploring the issue of abacavir-related HSR in different ethnic groups.

Applying the Biomarker

The HLA-B*5701 biomarker can be used to stratify patients into groups at high and low risk of HSR. Indeed, since the biomarker was published, a number of academic groups have been screening patients for HLA-B*5701 before treating them with abacavir. At the Royal Perth Hospital in 2000–2001, before screening for the biomarker was performed, 11 of 131 patients on abacavir developed HSR. After screening became a routine practice, however—from the beginning of 2004 through July 2005—only 1 of 49 patients exhibited HRS. That patient was known to be positive for the biomarker but chose to try abacavir anyway because there were no other options.

More recently, a French study involving 137 patients found a decrease in the HSR rate from 12 percent to zero after screening for the HLA-B*5701 biomarker was implemented. Lai explained that although prospective screening with the HLA-B*5701 marker shows great promise, academic groups that have been testing its use have been conducting small studies involving 50–100 cases. Therefore, it is still necessary to validate the

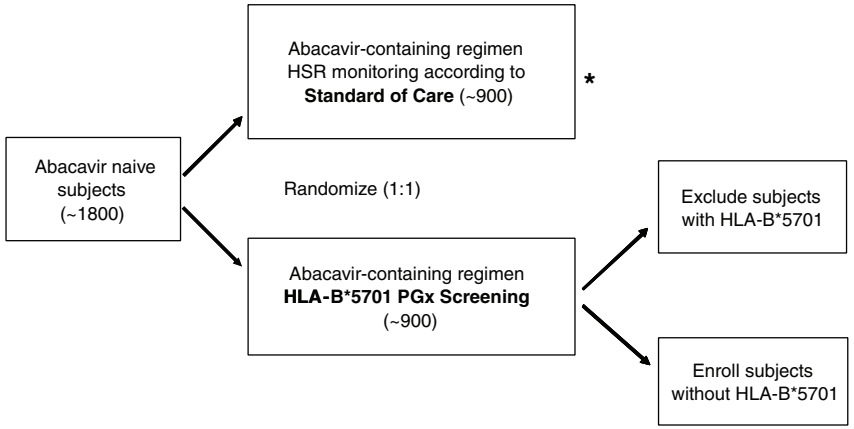


FIGURE 6-2 PREDICT-1 study design. The objectives of the trial are to compare HSR rates (± abacavir skin patch testing) in the two study arms marked with an asterisk (*), and evaluate the sensitivity of HLA-B*5701 in cases in the standard-of-care arm.
NOTE: PREDICT = Prospective Randomized Evaluation of DNA Screening in a Clinical Trial; PGx = pharmacogenomic.
SOURCE: Lai, 2007.

screening in an adequately powered prospective clinical trial. In 2006, GSK initiated two clinical trials—PREDICT-1 and SHAPE.[2]

PREDICT-1 (see Figure 6-2) is a highly powered prospective study that will examine the utility of HLA-B*5701 screening in a European HIV population. It will enroll 1,800 patients who have never been treated with abacavir. These patients will be randomly assigned to one of two groups, each with about 900 patients; one of the groups will be screened for HLA-B*5701, and the other will not. In the group that is screened, patients with HLA-B*5701 will be excluded from taking abacavir. By comparing the rates of HSR in the two randomized groups, GSK researchers will be able to measure the power of HLA-B*5701 screening to reduce the occurrence of HSR compared with the usual standard of care.

By contrast, SHAPE is a retrospective, matched-group, case-control study intended to estimate the sensitivity of HLA-B*5701 in both black and white patients, but using skin-patch testing to supplement the clinical diagnosis of HSR. Skin-patch testing provides a much better clinical indi-

[2]Note that following the workshop, the PREDICT-1 study results were published online in *Pharmaceutical Statistics* on May 29, 2007, and results from the SHAPE study were presented at the 4th International AIDS Society Conference on July 22–25, 2007, in Sydney, Australia.

cation of HSR than does standard diagnostics, so the study will provide a clearer measure of the usefulness of the biomarker in black and white populations.

IMPLICATIONS FOR THE FUTURE

Using this type of pharmacogenetics analysis to identify safety biomarkers prior to the approval of new drugs will demand the development of methodologies for prospectively managing drug-associated adverse events. One group at GSK, run by Clive Bowman, has begun developing a method for the real-time management of patients' adverse events. This method includes

- creating a collection of genetic markers—a thousand or more—to examine in people who present with an adverse drug event;
- creating a control set of genetic markers by genotyping people who have taken the drug without adverse effects; and
- as patients report with HSR or some other adverse reaction, genotyping them and comparing their genetic markers with those of the control group.

Calculations show that by the time 18–19 patients have reported with a particular adverse drug event, it should be possible to tell whether there is a genetic basis for the event and to identify potential markers in the genome. To test this methodology, GSK researchers designed a real-time retrospective whole genome scan study with abacavir data on 22 cases and 316 controls and worked with the data as though the cases were coming in prospectively one at a time. By the time they had 22 cases, they could identify 10 loci that correlated with HSR, and the fifth of those was the HLA-B locus. The implication is that by the time the 22nd case comes in, one will have identified that there is a problem, and one will have a number of loci that are potentially associated with a marker for HSR. Continuation of the simulation for the next 100 cases that presented allowed the researchers to eliminate the nine loci other than HLA-B as false positives and identify a clear marker for hypersensitivity—HLA-B*5701. Lai emphasized that the important difference between the simulation and how the marker was actually discovered is that using the simulation, it was possible to pinpoint the marker much sooner and potentially save hundreds of lives.

7

Qualifying Biomarkers[1]

Many of the biomarkers discussed at the workshop are observational or exploratory in nature. Such biomarkers are useful for screening compounds for toxicity but have not been qualified or validated for use in regulatory decision making. Dr. Vonderscher discussed what is involved in transforming observational or exploratory biomarkers into valid biomarkers that can be used in making regulatory decisions.

THE IDEAL BIOMARKER

Vonderscher outlined the characteristics of an ideal biomarker for kidney toxicity:

- It should be visible early, prior to histopathological changes, and should be indicative after active damage.
- It should be sensitive, but it should also correlate with the severity of damage.
- It should be accessible in the peripheral tissue; in the case of the kidney, for example, it should be measurable in either the blood or the urine.
- It should be analytically stable in tissue so it can be measured after

[1]This chapter is based on the presentation of Jacky Vonderscher, Vice President, Head of Exploratory Development in Europe, Novartis.

some time has passed, for example, after a biopsy has been taken or a necropsy performed.

- It should be translational; that is, it should bridge across species.
- It should be associated with a known mechanism. Many current biomarkers are identified through statistical analysis of gene expression, as discussed in Chapter 4, but one should be able to understand the biomarker and what is really going on in a biomolecular sense when it appears.
- A biomarker should be able to localize damage. For example, it should pinpoint the particular area of the kidney that has been damaged rather than just indicating kidney toxicity in general.

Given this extensive list of characteristics, a panel of biomarkers rather than any single ideal biomarker will likely be needed to characterize nephrotoxicity.

QUALIFICATION OF NEPHROTOXICITY BIOMARKERS

Before attempting to establish pathways for clinical qualification of biomarkers, Novartis qualified a set of nephrotoxicity biomarkers in animals. The qualification study was performed with 10 compounds: 8 nephrotoxicants plus 2 hepatotoxicants as negative controls. For each compound, the researchers used 96 rats: four dose levels, including the control, which was a zero dose; four termination time points; and six animals per group. The duration of each exposure was 2–3 weeks. In addition to the traditional toxicology analysis, the researchers performed gene expression analysis on kidney and liver tissue and also multiplex ELISA (enzyme-linked immunosorbent assay) on kidney, liver, urine, and plasma.

The nephrotoxicants were chosen to have a variety of modes of toxicity, including oxidative stress and damage to podocytes. The hepatotoxicants were known to cause cholangitis and liver cancer. The team chose 15 biomarkers, representing 85 percent of the markers being used by the Predictive Safety Testing Consortium (PSTC),[2] from various sources and publications, including some early gene expression work and some known proteomics work. The researchers attempted not to be selective about the source of the markers and to cover most of those that were interesting. Before running the studies, they performed a series of prestudies on the nephrotoxicants, in which they determined the correct doses to create lesions between grades 1 and 3.

[2]The PSTC public–private partnership, comprising members from industry, academia, and government, was established to identify and clinically qualify safety biomarkers. Novartis is a participant in the PSTC's efforts to identify and qualify nephrotoxicity biomarkers.

One of the key aspects of the process was settling on a lexicon of histopathology. After extensive discussion among the PSTC members, a list that included 12 primary kidney lesions and a larger number of secondary lesions was assembled. Tubular cell degeneration was one of the primary lesion types, for instance; it was subdivided into two secondary types, necrosis and apoptosis. Each of these lesion types was further classified according to where the lesion was localized in the kidney: the proximal convoluted tubule, the thick descending tubule, the loop of Henle, etc. One category was "no precise localization possible."

Example: Establishing Biomarkers to Predict Cisplatin Toxicity

After dosing the rats and examining them at various time points, the team identified 79 different types of localized lesions in the kidney. They then tried to determine how each biomarker correlated with the histopathological findings. Rats dosed with cisplatin evidenced tubular necrosis and apoptosis. The team tried to identify biomarkers that predicted the damage and, in particular, that showed a quantitative relationship between the level of biomarker and the amount of damage.

The researchers found that serum creatinine was not a particularly useful biomarker (see Figure 7-1) because, although some of the middle-dose animals had histopathology grades 1 and 2 (the highest), only the animals in the high-dose group had serum creatinine above the threshold. The results were similar for blood urea nitrogen (BUN): only the high-dose group showed BUN levels above the threshold, while a number of the animals in the middle-dose group had pathology grades of 1 and 2. In contrast, Kim-1 (kidney injury molecule-1) was a much more effective biomarker for tubular necrosis and apoptosis (see Figure 7-2). Unlike creatinine and BUN, it was elevated not only in the high-dose group but also in the middle-dose group—but only in those animals that showed histopathology grades of 1 or 2. Furthermore, there was a clear correspondence between Kim-1 levels and histopathology grades, with the higher Kim-1 levels correlating with the higher histopathology grades. Only a few animals deviated from that pattern: one with a histopathology grade of 1 with Kim-1 levels slightly below the threshold, and four with a grade of 0 that fell somewhat above the Kim-1 threshold. The marker urinary clusterin exhibited properties similar to those of Kim-1, but it had more false negatives—that is, animals with levels below the threshold but with histopathology grades of 1 or 2.

To obtain a quantitative measure of how well the various biomarkers predicted lesions, the team performed an ROC (receiver operating characteristic) analysis on the data. The animals were divided into two groups: control animals that had no lesions (histopathology grade of 0) and were

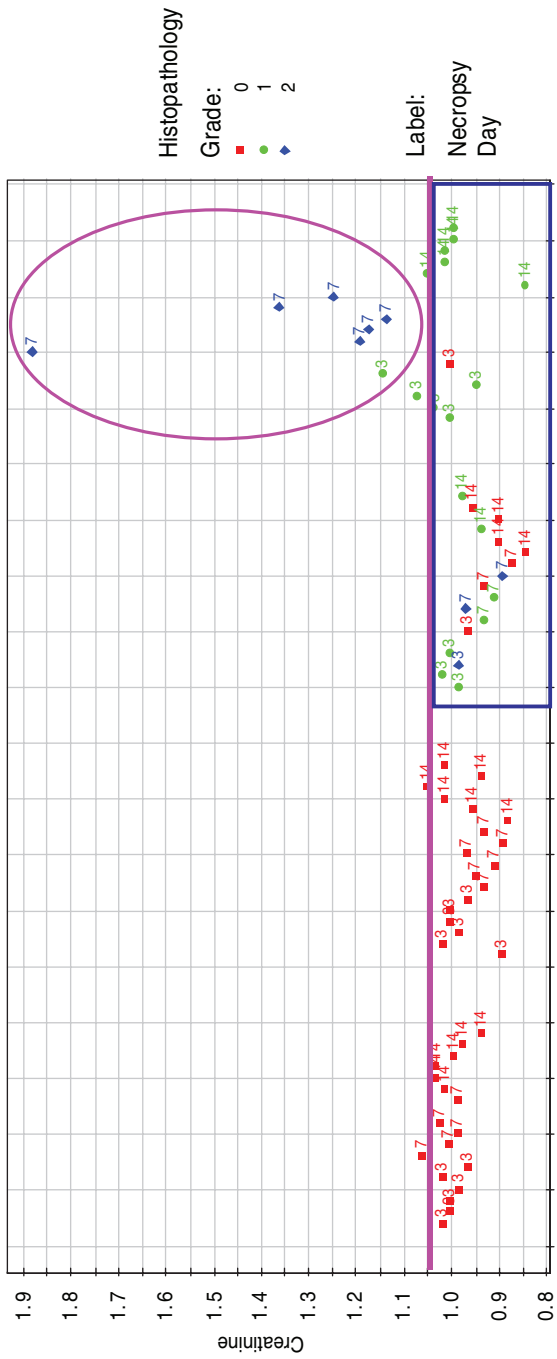


FIGURE 7-1 Serum creatinine as a biomarker for cisplatin-induced tubular necrosis/apoptosis. The data shown represent the fold change in serum creatinine at increasing dose levels. Results indicate that serum creatinine was not a particularly useful biomarker because, although some of the middle-dose animals had histopathology grades 1 and 2, they were comparable to control animals and only the animals in the high-dose group had serum creatinine above the control threshold.
SOURCE: Vonderscher, 2007.

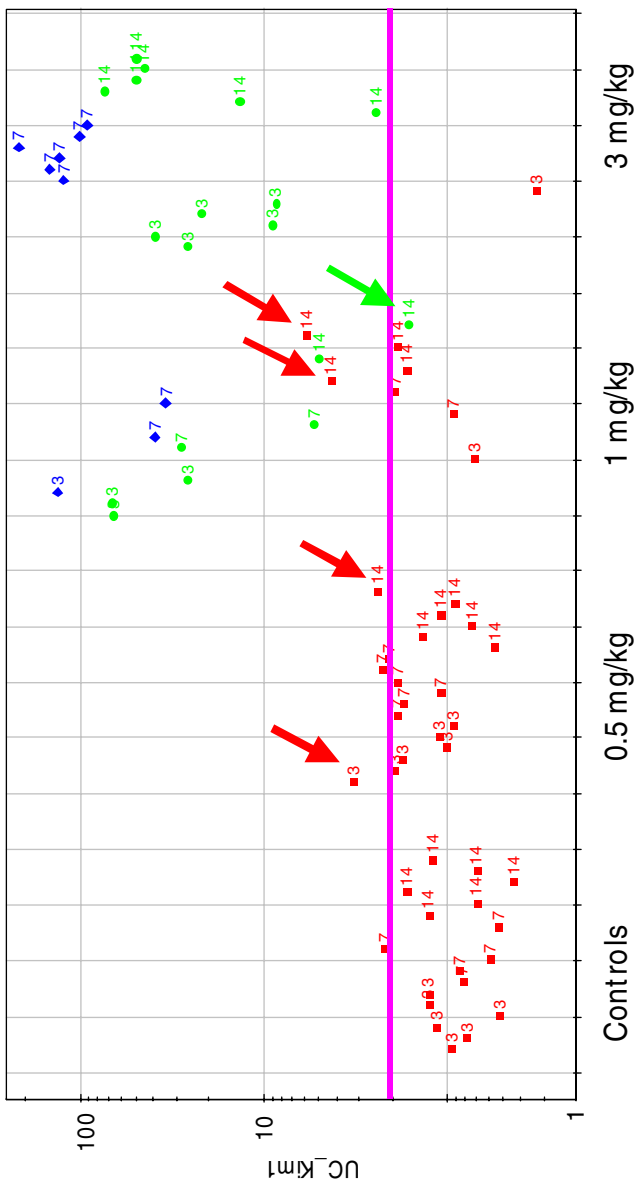


FIGURE 7-2 Kim-1 as a biomarker for cisplatin-induced tubular necrosis/apoptosis. The data shown represent the fold change in urinary Kim-1 at increasing dose levels. Unlike creatinine, Kim-1 may be a useful biomarker for cisplatin-induced tubular necrosis/apoptosis. Kim-1 was elevated not only in the high-dose group but also in the middle-dose group (in animals that showed histopathology grades of 1 or 2). There was a clear correspondence between Kim-1 levels and histopathology grades, with the higher Kim-1 levels correlating with the higher histopathology grades. (Note that the concentration of Kim-1 in this figure is exhibited on a logarithmic scale, while the concentration of serum creatinine in Figure 7-1 is shown on a linear scale.)
SOURCE: Vonderscher, 2007.

either nondosed or hepatotoxicant-dosed; and animals with lesions (histopathology grade of 1 or 2), regardless of their dose status. To generate an ROC curve, the true positive rate was graphed against the false positive rate as the threshold was varied continuously. The area under the ROC curve gives a quantitative measure of how good the predictions are: in the case of a perfect predictor, with a threshold that has all the positives above and all the negatives below, the area under the curve will be 1.0; in the case of a random predictor, the area under the curve will be 0.5.

Vonderscher displayed a graph with the ROC curves for the four markers mentioned above: serum creatinine, BUN, Kim-1, and clusterin (see Figure 7-3). In the case of creatinine, the area under the curve was 0.53—just better than random. BUN was somewhat better, with an area under the curve of 0.62. Clusterin yielded an extremely good result, with an area under the curve of 0.93. But Kim-1 was nearly perfect, with an area under the curve of 0.99.

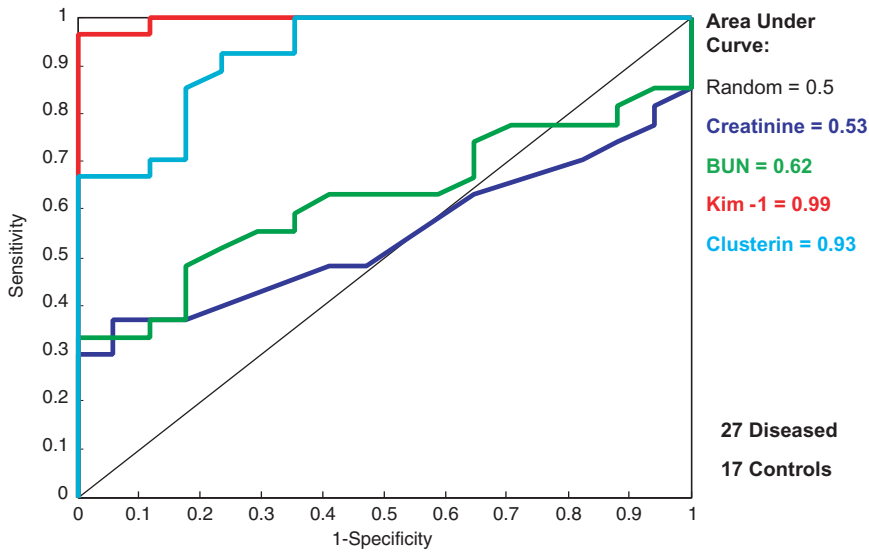


FIGURE 7-3 ROC (receiver operating characteristic) analysis to compare biomarkers for cisplatin-induced tubular necrosis/apoptosis. The area under the curve for a biomarker that perfectly predicts cisplatin-induced tubular necrosis/apoptosis would be 1.0. In this experiment, creatinine had an area under the curve of 0.53—just better than random; BUN was somewhat better, with an area under the curve of 0.62; clusterin yielded an extremely good result, with an area under the curve of 0.93; and Kim-1 was nearly perfect, with an area under the curve of 0.99.
SOURCE: Vonderscher, 2007.

Analysis of the Remaining Nephrotoxicants

Similar analyses were performed for the remaining nephrotoxicants and two hepatotoxicants. Using ROC curves, the team summarized how well the markers predicted various types of lesions caused by the compounds. In one analysis, for example, the team looked at proximal and nonlocalized tubular necrosis (see Figure 7-4). In this case, Kim-1 was still the best-performing biomarker, but its lead over clusterin was reduced, to 0.95 versus 0.93. The researchers found that creatinine and BUN performed much better when all of the compounds were included in the analysis rather than just cisplatin. The area under the ROC curve for creatinine was 0.83 and for BUN was 0.81. Part of the reason that the ROC

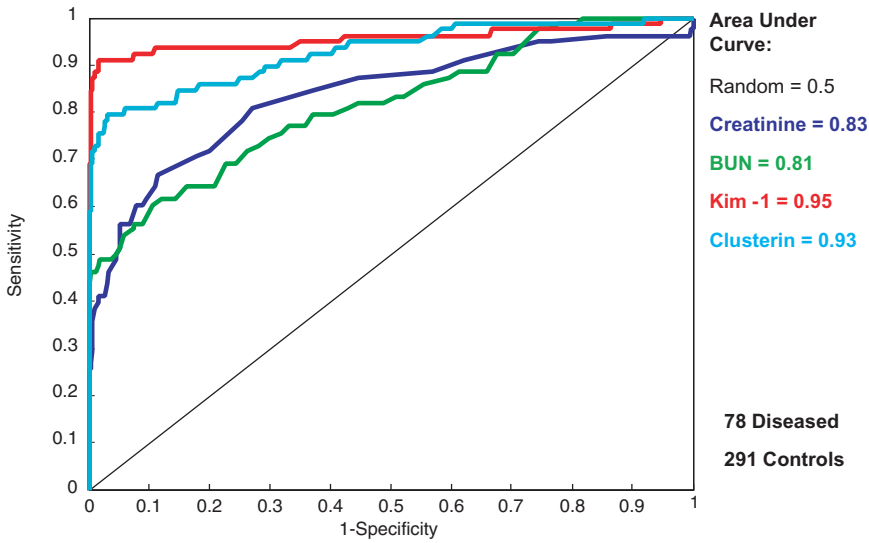


FIGURE 7-4 ROC (receiver operating characteristic) analysis to compare biomarkers for tubular necrosis mostly proximal (but sometimes not clearly localized) in 10 studies with different nephrotoxicants. The area under the curve (AUC) for a biomarker that perfectly predicts tubular necrosis would be 1.0. As in the experiment described above in Figure 7-3, Kim-1 was the best-performing biomarker (AUC = 0.95) followed closely by clusterin (AUC = 0.93). Creatinine and BUN performed much better when all of the compounds were included in the analysis rather than just cisplatin but were definitely not as good as Kim-1 and clusterin. The area under the curve for creatinine was 0.83 and for BUN was 0.81.
SOURCE: Vonderscher, 2007.

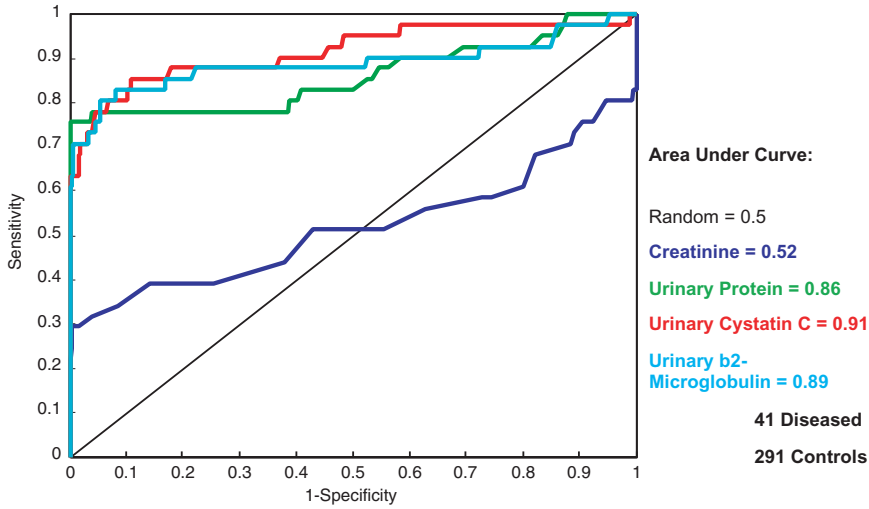


FIGURE 7-5 ROC (receiver operating characteristic) analysis to compare biomarkers for glomerular alteration/damage in 10 studies with different nephrotoxicants. The area under the curve for a biomarker that perfectly predicts glomerular alteration/damage would be 1.0. In this experiment, creatinine, which had an area under the curve of 0.52, was not a good marker for glomerular alteration. However, there were several other markers that were promising; the areas under the curve for urinary proteins, urinary β2-microglobulin, and urinary cystatin C were 0.86, 0.89, and 0.91, respectively.
SOURCE: Vonderscher, 2007.

score for Kim-1 dropped to 0.95 when all the different compounds and lesions were included was the inclusion of one compound that caused lesions in the tubular collecting ducts, where Kim-1 is not expressed and so cannot serve as an effective marker.

In a similar analysis for glomerular alteration and damage (see Figure 7-5), creatinine once again performed little better than random (area under the ROC curve of 0.52). Thus the researchers concluded that creatinine is not a good marker for glomerular alteration, but that several markers are very promising for this sort of damage. For example, urinary proteins have an area under the ROC curve of 0.86. For urinary β2-microglobulin, the area under the curve was 0.89 and for urinary cystatin C was 0.91.

SUMMARY

In a very narrowly defined context, there appear to be some markers that could potentially be viewed as known valid biomarkers. As noted above, however, a panel of biomarkers will likely be required to characterize nephrotoxicity rather than a single ideal biomarker. A panel will be necessary in particular to specify different localizations in the kidney and to differentiate among toxicity types. While Novartis and the PSTC have not yet achieved this capability, their ultimate goal is to assemble a collection of kidney toxicity markers that will be visible prior to histopathological changes and can serve as a panel covering most nephrotoxicity.

8

Pharmacovigilance

Following the discussion of techniques being developed for use in the discovery and preclinical stages to predict and understand drug toxicity, the workshop turned to the postmarket stage and ways to monitor adverse events and identify safety concerns as quickly as possible. As demonstrated by the experience with Vioxx and other drugs that had to be withdrawn from the market, drugs can make it through the development and approval processes without unanticipated serious adverse effects being recognized. In such cases, some people will inevitably take the drug and experience adverse effects; thus the goal must be to identify the problem quickly to minimize the number of people affected. To this end, three speakers described approaches to pharmacovigilance that can be applied to identify safety problems as early as possible after drugs have been put on the market.

PHARMACOVIGILANCE AT GLAXOSMITHKLINE[1]

Dr. Almenoff described an aspect of GlaxoSmithKline's (GSK's) pharmacovigilance program called online signal management. This program combines a number of technologies into one tool that can help safety evaluators review information on marketed drugs more efficiently and in much greater detail than previously was possible.

[1]This section is based on the presentation of June Almenoff, Vice President, Safety Evaluation and Risk Management, Global Clinical Safety and Pharmacovigilance, GlaxoSmithKline.

Safety Data Mining

GSK receives approximately 90,000–100,000 spontaneous adverse event reports each year. Recognizing that its researchers needed new tools to help them understand and prioritize the most important data, in 2002 GSK began using data mining to evaluate its safety data.

Because postmarket information is reported voluntarily, there is no control group (i.e., it is impossible to know how many people took a drug, how many people experienced an event, and how many people experienced that event after taking the drug). Therefore, it may be difficult to evaluate precisely how rare or common a particular adverse event is. For example, if there are 30 reports of strokes occurring in individuals taking drug X, is that too many? This is a difficult question to answer as it depends on how common strokes are in the normal population, how much exposure there has been to drug X, and other factors. Answering such questions therefore requires an objective, systematic approach.

GSK uses a statistical approach called disproportionality analysis (DPA) to identify rare events that occur at a greater frequency than would be expected by chance. The DPA calculation is derived from a two-by-two table such as that shown in Figure 8-1. If the ratio of A/(A + B) is greater than the ratio of C/(C + D), there is a potential association between the drug and the event of interest.

For example, to determine whether there was an association between drug X and stroke, one would look at the number of stroke cases reported for drug X as a proportion of all the adverse events reported for that drug. One would then compare that result with the number of strokes reported for all drugs as a proportion of all the adverse events reported for all

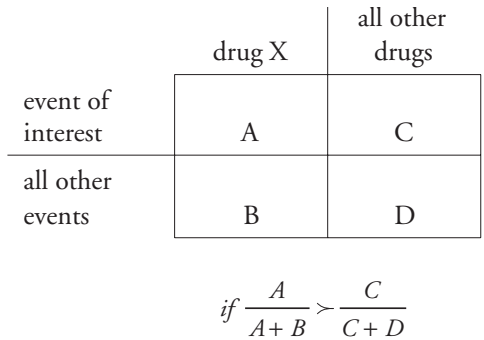


FIGURE 8-1 Disproportionality analysis calculation.
SOURCE: Almenoff, 2007.

drugs and ask whether the proportion of strokes for drug X was greater than the proportion for all drugs. If it was, drug X might be associated with stroke.

The statistical tool GSK uses for such analysis is called MGPS (Multi-item Gamma Poisson Shrinker), and it produces a statistical output called the EBGM (empirical Bayes geometric mean). The EBGM is a measure of association, and it can be thought of as a relative reporting ratio—if the number is greater than 1, there is a statistical association between drug and event.

Almenoff warned that such statistical analysis is not sufficient by itself, as there are always biases in the data. Thus GSK researchers medically verify all signals identified by such data mining. And while such data mining can be an important tool in the armamentarium of postmarket product surveillance, it is intended only to enhance current pharmacovigilance techniques, not replace them.

Online Signal Management

GSK currently uses data mining in conjunction with other pharmacovigilance techniques to enhance and streamline the surveillance process. Online signal management (OSM), a tool GSK developed in collaboration with Lincoln Technologies, integrates safety data mining with case-based screening algorithms and also provides an opportunity for traditional case review.

The system constantly monitors GSK's database, performing various analyses to look for patterns and changes and posting alerts when such events occur. The goal is to provide a filter that will allow GSK safety evaluators, without having to sort through every event, to focus on three things: new data, important safety signals, and fluctuations in the data.

When safety evaluators log on to the system, they are provided with a primary review that includes

- a listing of all serious adverse event reports for the drug they are responsible for monitoring in a particular time interval;
- all events with a rising trend; and
- all nonserious unlisted reports that have EBGM values above a defined threshold.

OSM combines this filtering capability with a number of other tools that enable safety evaluators to follow up on a signal to determine whether it represents a problem. For example, the system is equipped with data retrieval capabilities so that when an evaluator sees a signal, it is possible to click on the relevant medical issue and retrieve information on the

cases that underlie that signal. The system also incorporates trend analysis and visualization tools, such as heat maps on which the EBGM signal scores are shown in red if they are high and in green or black if they are lower. Within the heat map, it is possible to click on a particular spot and view the cases that are represented by that spot. This is useful because by looking directly at the data, reviewers can generally verify whether a particular signal is a false positive.

In addition, sophisticated knowledge management tools help evaluators prioritize their time. With the capability to attach annotations to previous analyses that have been run, evaluators can easily track prior work and thought processes and include these when new analyses are run. Another valuable aspect of OSM is the ability to look at subpopulations. It is possible, for example, to scan the entire database and find for a particular drug what side effects are occurring more frequently in pediatric patients than in adults or the elderly. It is also possible to see what sorts of events are reported more frequently in overdose versus nonoverdose situations. This is done with quantitative algorithms that flag the situations automatically. There are also more qualitative flags, such as for product complaints, that help identify manufacturing problems.

OSM has been a success at GSK, receiving positive user reviews and saving safety evaluators between 30 and 40 percent of their time. Almenoff asserted that this program has dramatically improved the quality and focus of postmarket safety reviews at GSK.

STATISTICAL ISSUES IN ANALYZING SPONTANEOUS REPORT DATABASES[2]

The U.S. Food and Drug Administration's (FDA's) primary method of collecting postmarket data and monitoring for adverse events is passive surveillance. Reports of unexpected outcomes are submitted voluntarily by patients and health care practitioners on the FDA's Medwatch form 3500, which includes a box by which suspected products can be linked to the outcome. The FDA receives more than 400,000 of these spontaneous reports each year (IOM, 2007). Dr. DuMouchel described some of the issues involved in the statistical analysis of spontaneously reported adverse event databases. In particular, the analytical capability at the heart of the OSM tool described by Almenoff was developed through a cooperative research and development agreement between Lincoln Technologies and the FDA, and DuMouchel offered further detail on this tool. In addition to analyzing the FDA's Adverse Event Reporting System (AERS), the tool

[2]This section is based on the presentation of William DuMouchel, Chief Statistical Scientist, Lincoln Technologies.

can be used to analyze a variety of other spontaneous report databases, including the Vaccine Adverse Event Reporting System run by the FDA and the Centers for Disease Control and Prevention, the World Health Organization's VigiBase, databases for medical devices, and others.

Data Cleaning

Before the data in these databases can be analyzed, they must be put in a standardized form, a process called "data cleaning." For the AERS database, there were two primary data cleaning issues: a lack of standardized nomenclature for drug names and duplicate submissions.

Because the AERS database does not use standardized nomenclature, it contains roughly 300,000 different names for drugs. These include both generic and trade names and many different misspellings, and the dosage is given with the name, so that "25 milligrams" or "25 mg" appears as part of the drug name. Since there are 3 million reports in the database and each entry typically includes several drugs, some 10 million drug names in the database had to be reviewed, one by one, and put in a standardized form. This was primarily a manual process, with little computer assistance, and took years to complete, but eventually Lincoln Technologies was able to reduce the 300,000 different names to about 3,000 ingredients in a standardized generic form. Because the FDA's Medwatch forms continue to be collected using nonstandardized nomenclature, this process must be repeated every quarter when Lincoln receives new data from the FDA.

The second data cleaning issue was detection of duplicate submissions. If a person is taking three drugs from three different manufacturers, for example, an adverse event will often filter back to the FDA in three separate reports from the three manufacturers. In addition, follow-up reports sometimes are not linked properly to the original report, and the process of sorting these reports can be tedious. Sorting is done by means of a computer science discipline called record linkage.

Of the 3 million reports in the AERS database, there are probably about 300,000 duplicate reports that need to be removed. While this may not sound like a large number—only about 10 percent of the database—it can make a big difference in the signal that is extracted from the data. For example, a rare drug–event combination that should have a count of one might have four duplicates, raising an unnecessary safety concern. This example shows the importance of flexible computer tools; if the safety analyst can bring up the five detailed reports with a mouse click, the duplicates are much more likely to be detected.

Statistical Underpinnings

The major problem with performing statistical analysis on a database of spontaneous reports is that there is no denominator (no way of knowing exactly how many people took a particular drug), and there is no numerator (no way of knowing exactly how many events occurred because of underreporting). Thus there is no way to calculate an adverse event rate directly. The solution is to compare the adverse event rates for one drug with those for all other drugs.

To this end, the analysis works from a two-by-two table. For every drug of interest, D, and every event of interest, E, one obtains the four entries in the table by counting all the reports that do or do not involve D and that do or do not involve E. The top left entry, for instance, is the number of reports, n, that involved drug D and adverse event E. This is the number one must examine to determine whether it is larger than would be expected, and this is done by using the other three numbers to calculate an expected value, e, which is then compared with the actual value. There are a variety of ways to calculate this expected value, but regardless of which method is used, the final step is to divide the actual number of events associated with a particular drug by the expected number to obtain a disproportionality ratio, n/e. If this ratio is much larger than 1, there may be a problem.

The idea of computing these ratios is a simple one, so one might ask why the use of such calculations has become widespread only in the past decade or so. Part of the answer is that recent computer and database advances have made it easier to perform this sort of analysis, but another part of the answer is that biostatisticians are sometimes hesitant to conduct formal statistical analyses on data collected outside of a controlled clinical trial environment. Only recently did scientists begin applying statistical models to spontaneously reported data.

Inherent limitations make it necessary to analyze these data carefully. For instance, if a particular drug is taken primarily by one age group or one sex, spurious associations can appear in the database. An example is sudden infant death syndrome (SIDS) and childhood vaccines. If one simply performed the calculations naively, one would find a large disproportionality ratio even though there is no causal relationship between SIDS and the vaccines. The Mantel-Haentzel adjustment—whereby the data are stratified by age, sex, and report year, and the expected values are computed separately for each group—can be applied to deal with this problem.

A trickier issue is the fact that with thousands of drugs and millions of ratios being calculated, large ratios will inevitably appear. If there is just one event of a particular type, but the expected value is only a small

fraction, say 10^{-5}, the disproportionality ratio will be quite large without necessarily signifying anything other than random chance. For example, if there were 1 million different drug–event combinations, each with an expected frequency of 10^{-5}, one would expect 10 of them to have an observed count of 1 by chance, with the remaining 999,990 having a count of 0. But the 10 that happened to show up would each have a p value of about 10^{-5}, which a naïve analysis might deem significant.

The question thus arises of how to take into account simultaneously the proportionality ratios, the p values, and the multiplicity of counts. When calculating millions of two-by-two tables, there will inevitably be many cases with large ratios, and researchers must determine how they should be sorted to identify those cases most likely to be associated with real problems. Suppose, for example, that there is one case in which there are 2,000 adverse events compared with only 1,000 expected events. Then n = 2000, e = 1,000, and the disproportionality ratio is only 2, but the p value is minuscule, implying a very clear signal. Researchers must determine how this case should be compared with one in which n = 20, e = 0.2, and n/e = 100, but the p value is much larger.

This problem can be addressed with a statistical tool called a Bayesian shrinkage model, first applied to the FDA database in 1999. Working with statistics from the entire database, this model allows one to combine the disproportionality ratio and the p value into a single value—the EBGM mentioned earlier. This number can be thought of as an a posteriori estimate of the ratio based on looking at the data as a whole.

The practical effect of performing this statistical analysis is to shrink the calculated ratios for cases in which the p value—and thus the uncertainty—is large. In calculating ratios from the database, there are many cases in which n = 1 and e is some very small number, so that n/e is, say, 3,000. In such cases, the model realizes that there is so much variance in the estimate of the ratio that a value of 2 or 3 is a better estimate than 3,000. When n is in the range of 10 to 20, by contrast, there is typically only a slight shrinkage, and for an n of several hundred, there is generally no shrinkage at all.

The bottom line is that this statistical analysis modifies the original calculated disproportionality ratio to take into account how variable that ratio estimate was and provide a better indication of how significant the event or events really are. The analysis is particularly useful because it provides a single number that can be graphed or plugged into other models.

As an example of the usefulness of having a single number, DuMouchel showed a heat map of adverse events for a single drug (see Figure 8-2). This heat map is divided into different spaces according to MedDRA

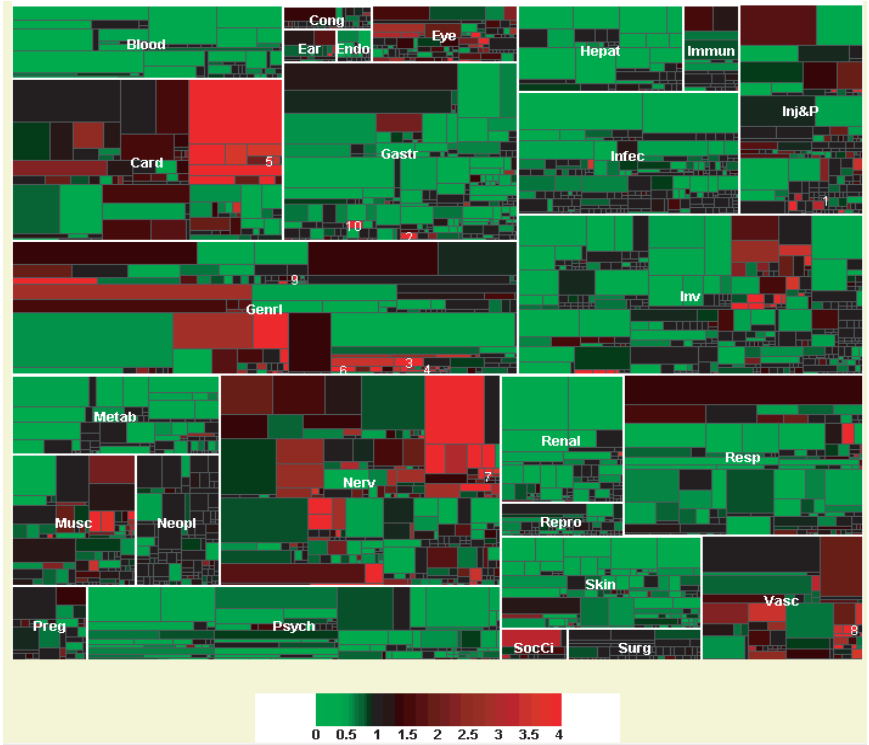


FIGURE 8-2 Heat map profiling of spontaneously reported adverse events for a single drug. This heat map is divided into different spaces according to MedDRA (the Medical Dictionary for Regulatory Activities) terms; the rectangles outlined in white and labeled are the system–organ classes. The smaller squares represent the 10,000 or so MedDRA preferred terms that are grouped with respect to the hierarchy of MedDRA. The heat map is interactive, and as a computer cursor is moved over these squares, information appears concerning where that square falls in the MedDRA grouping.
SOURCE: DuMouchel, 2007.

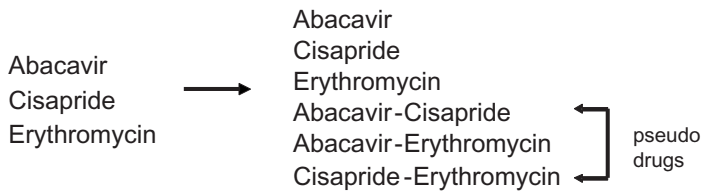
(the Medical Dictionary for Regulatory Activities) terms. The biggest rectangles are the system–organ classes—blood, cardiovascular, respiratory, renal, gastrointestinal, and so forth—but all of the 10,000 or so MedDRA preferred terms are grouped into very small squares where the grouping respects the hierarchy of MedDRA. One can explore this heat map by moving the computer cursor over these squares; as this happens, information appears concerning where that square falls in the MedDRA grouping.

Drug Interactions

It is also possible to use the above analysis to look for adverse events related to interactions between drugs. The more medications a person takes, the greater is the chance of a drug interaction. Therefore, as more people take more prescription drugs (DuMouchel reported that 12 percent of the elderly take at least 10 drugs a week), interaction effects are becoming more important.

The process of looking for adverse events due to drug interactions is straightforward. A pair of drugs is treated as an additional "pseudo-drug." If, for example, there is a report of a patient's taking three drugs and the three drugs are listed in the report, the analysis treats the case as though the patient were taking three drugs—A, B, and C—as well as three pseudo-drugs—A + B, A + C, and B + C. From this point, the analysis is the same, with the observed number of events being compared with the expected number of events, and an EBGM being calculated to express the modified disproportionality ratio.

As an example, DuMouchel used an analysis of the drugs cisapride and erythromycin, alone and in combination, and how often they were associated with torsades de pointes, an uncommon variant of ventricular tachycardia (see Figure 8-3). The EBGM for each drug alone relative to torsades was about 20, but the EBGM for the combination of the two drugs was nearly 230, a huge disproportion.



Drug	Event	N	EBGM	PRR
Cisapride	Torsade de pointes	92	19.525	69.919
Erythromycin	Torsade de pointes	58	20.425	13.227
Cisapride-Erythromycin	Torsade de pointes	18	**228.733**	**755.355**

FIGURE 8-3 Association between torsades de pointes and the drugs cisapride and erythromycin, alone and in combination. The EBGM (empirical Bayes geometric mean) for each drug alone relative to torsades was around 20; however, when the two drugs were taken in combination, the EBGM was much greater, as was the PPR (proportional reporting ratio).
SOURCE: DuMouchel, 2007.

Complications

A number of complications and subtleties must be taken into account in performing this sort of analysis. For one thing, the analysis is based on the assumption that all reports except those concerning the drug of interest can be considered "background noise." The problem is that the "control group"—all reports except those concerning the drug of interest—may include other drugs with very high signals for the event of interest, and in this case it is not a very good control group. The denominator will be inflated, and this will partially mask the effect that is the target of the analysis. Improved methods for dealing with this issue are needed.

Another issue that needs to be considered during analysis is confounding due to people taking more than one drug at a time. If a particular drug that causes an adverse event is often coprescribed with another drug, that second drug will inherit the association with the adverse event. This is called signal linkage or the innocent bystander effect, and it is particularly prominent in drugs used for certain chronic conditions, such as diabetes or HIV infection, for which a set of drugs is often prescribed together. If one of those drugs has a serious association with adverse reactions, that association will propagate to the others.

The standard way of dealing with such confounding is multiple regression analysis; however, there are complications that must be addressed. For such a regression analysis, the adverse event is taken as the response, or dependent variable, and the stratification variables and the presence or absence of various drugs are taken as the predictors, or independent variables. With this analysis, the background noise rate can be estimated automatically and can be extended to estimate drug interactions.

This is a time-consuming process as it is necessary to perform a multiple regression analysis for every adverse event; thus if 10,000 MedDRA terms are being considered, 10,000 regressions must be calculated. Furthermore, with the presence or absence of a drug as a predictor and with 3,000 drugs in the database, there is a very large number of predictors for the regression model. In addition, if a large number of coefficients are estimated simultaneously, it becomes necessary to add shrinkage methods to the regression analysis.

One final confounding factor that must be taken into account is that drugs taken for particular diseases can appear to be related to the symptoms of the disease. For example, if a disease causes nausea, nausea may emerge from the analysis as an adverse event related to a drug prescribed for that disease, even if it is an antinausea medication that is being evaluated. In such cases, it is important for medical expertise and judgment to be involved in the analysis to rule out such factors.

Summary

The issues that need to be considered when one is performing DPA on spontaneously reported events can be summarized as follows:

- Extensive data cleaning is necessary to sort and organize millions of records.
- There are many noncausal reasons for associations between drugs and events.
- In comparison with clinical trial or cohort data, where participants can be followed from start to finish, these studies are poorly designed.
- Interpretation of comparator groups is difficult.
- Multiple comparison and post hoc fallacies are endemic.

Despite the need to address these issues, systematic DPA can yield a number of beneficial results, including the following:

- This method is considered the only way to learn about very rare adverse drug events.
- It provides hypothesis generation and a second data source for comparisons.
- The Bayesian approach to multiple comparisons aids in assessment.
- Computer tools have improved productivity.
- The signal management approach enables institutional "memory."

One weakness of DPA was brought out in the discussion when John Senior, FDA, questioned whether quantitation of DPA can provide numbers in which one can be confident and how well those numbers relate to real risk. DuMouchel agreed that DPA is not as good as incidence rates or relative risks, but stressed that it is useful nonetheless. He explained that if DPA were viewed as estimates of an event that was overrepresented in the database, there would be no problem with comparing two drugs to determine whether one was more represented than another. Once risk has been assessed, case reports can be examined and medical judgments made.

ACTIVE SURVEILLANCE FOR ANTICIPATED ADVERSE EVENTS[3]

Historically, postmarket monitoring for adverse events has been accomplished through passive surveillance. While this system may be

[3]This section is based on the presentation of Richard Platt, Professor and Chair, Harvard Medical School and Harvard Pilgrim Health Care.

capable of detecting rare serious adverse events, it has several limitations, including underreporting, biased reporting, and difficulties in attributing an adverse event to a specific drug. In addition, the data accumulate slowly, and answering important safety questions can take years. With the technological advances that have occurred in recent years, numerous groups and stakeholders have embarked on the establishment of active surveillance systems to monitor for adverse drug events, with the aim of identifying drug safety issues more quickly than is possible with standard passive surveillance. Dr. Platt described what a national active surveillance system might look like and elaborated on the benefits and challenges it would entail.

Using Claims Databases for Surveillance

A large percentage of Americans' medical records and history of prescription drug use can be accessed by using health care claims, making this an ideal platform for launching a national active surveillance system. The backbone database of such a system would comprise routinely collected administrative health care claims enhanced by supplemental information, such as links to full-text medical records in either electronic or paper form, laboratory results, and pharmacy records. Claims databases have the important features of covering defined groups of individuals and containing information on all reimbursed care. Thus they can provide both numerators (e.g., how many people experienced event X after taking drug Y) and denominators (e.g., the total number of people who took drug Y) for events, avoiding biases in systems based solely on medical records.

To test how well such a claims database might work, Platt and colleagues in the FDA and the Centers for Education and Research in Therapeutics (CERTs) program at the Agency for Healthcare Research and Quality performed a retrospective study to determine how early it might have been possible to uncover the association between Vioxx (rofecoxib) and acute myocardial infarction. They looked at several years' worth of claims from a group of health plans with an aggregate population of about 7 million and plotted the observed number of myocardial infarctions among rofecoxib users versus the expected number, based on a comparison group composed of naproxen (brand name Aleve) users. They concluded there was a statistically significant signal of excess acute myocardial infarction when 28 heart attacks had been recorded among rofecoxib users, data that took 34 months to accumulate in the group of health plans with which they were working. If the researchers had had data for 100 million people available, the signal might have been evident from only about 3 months' worth of data. While the data are never available immediately—it takes

a while to obtain them and to transfer them into analyzable form—this example illustrates that working with large data sets can make it possible to identify phenomena of interest relatively quickly.

An FDA reviewer questioned Platt's assertion that an active surveillance system would have been able to detect the safety signal from Vioxx much sooner. She noted that Platt had an advantage in picking the outcome to study and the comparator, whereas in real time, if other outcomes had been monitored or a different comparator had been used, the myocardial infarction events might have been masked. In other words, because Platt's study was retrospective (it was already known that myocardial infarction was the problem), it was possible to monitor specifically for that event. In a real-life situation, researchers might not know which events to monitor closely for, and therefore it might take longer or be more difficult to identify the unanticipated serious adverse event than was demonstrated with Platt's example.

Using Claims Databases for Hypothesis Testing

A valuable use for claims databases would be to test hypotheses that have been raised in some other way. For example, a hypothesized connection between the Menactra meningococcal conjugate vaccine and Guillain-Barré syndrome is currently of substantial interest to both the FDA and the Centers for Disease Control and Prevention (CDC). Shortly after the vaccine was approved in 2005, the Advisory Committee on Immunization Practice recommended that it be used for all adolescents; within a year or so, 15 spontaneous reports of Guillain-Barré syndrome occurring within 6 weeks of immunization had been filed. At the time, it was estimated that approximately 6 million people had been immunized. A number of questions were raised, such as whether an excess risk is associated with the vaccine; if so, how great; and whether this is a high-risk subgroup.

The Vaccine Safety Datalink (VSD) project, a CDC-supported program that operates in eight health plans of the HMO Research Network, quickly became involved and analyzed the risk using its database of 7 million health plan members. At the end of a year, approximately 100,000 doses had been administered, and no cases of Guillain-Barré syndrome had appeared among those receiving the vaccine. However, this did not rule out a connection: since the background rate of Guillain-Barré syndrome is only about 1 to 2 cases per 100,000 person-years, it could take several years for the connection to be observable in the VSD database.

The FDA currently has postmarket contracts with the same eight HMO Research Network plans that are involved in the VSD project, as well as UnitedHealthcare, two state Medicaid databases, and the Veterans Health Administration system; altogether, these organizations represent about 26

million people. The information they provide includes details about the diagnoses that are assigned, about the procedures people undergo, and about the drugs dispensed through pharmacies, and all of this information can be linked to full-text medical records. While these are the systems used most often in the United States for surveillance purposes, however, they are insufficient for ensuring timely identification of new adverse events or timely follow-up on safety signals. Platt asserted that in addition, linked databases from Medicare Parts A, B, and D, Medicaid in most large states, and private health plans need to be accessible and included in a national surveillance network.

To complement the VSD project's effort to test whether the vaccine increased the risk of Guillain-Barreé syndrome, a one-time collaboration of four health plans with 40 million members was established to conduct a study that would use linked automated resources to identify potential cases by their diagnosis codes, obtain the medical records of potential cases for review and abstraction, and have an expert panel adjudicate all abstracted cases. The study results will be reported frequently to the FDA, CDC, and the vaccine's manufacturer, and will eventually be made available to the public.

These efforts have helped stimulate the creation of a standing consortium called the Health Plan Consortium for Public Health, which Platt and colleagues are working to develop. The consortium would be run under the auspices of CERTs and would have a target population of 100 million covered people. Its goal would be to improve the safety and use of marketed vaccines and prescription drugs. While there is no guarantee that the consortium will be accomplished, active planning is under way.

Characteristics of an Effective Active Surveillance System

Assuming such a consortium could be established, Platt described a number of characteristics that he would expect the active surveillance system to have. It would function as a distributed network, with the data residing at and belonging to the individual health plans. Typically, the data would be accessed via computer programs that would be distributed to each health plan, which would run the programs on their own data. Results would be returned to a coordinating center, combined with results from other health plans, and then analyzed. Although most of the surveillance analysis could be performed using deidentified information, thus ensuring protection of confidential personal information, it might be necessary for the health plans to provide individual patient-level data for a small fraction of individuals with specific diagnosis codes or other characteristics requiring additional evaluation. However, this information would be provided under full Health Insurance Portability and Account-

ability Act (HIPAA) protections. Platt commented that the VSD project works this way.

In addition to claims data, the system should ideally have access to a variety of other clinical information. For example, many large national health plans now receive laboratory test results for their members, and electronic medical records are becoming increasingly available and should eventually become a critical component of a national surveillance system. Platt stressed that being able to access full-text medical records would be crucial for this system. Because the data involved in the system would belong to individual health plans, plans should have the option to opt in and out of specific uses of the data. Further, during development of the system, transparency to the public would be important: protocols should be offered for public comment before being finalized and available to the public when a study begins, and results should be provided to the public when a study is completed.

Data Ownership and Decision Making

Robert Califf, of Duke University, urged caution in response to Platt's description of an integrated active surveillance network in which data would belong to individual health plans, companies, regulators, etc., and groups could opt in and out of specific uses of the data. He warned against every stakeholder having its own data sets, completing its own analyses, and making its own decisions about what drugs are dangerous or safe. Platt responded that the signal detection he described represents only the beginning of the decision process. Currently, these systems cannot be used to make assertive decisions, as researchers in the community are in the midst of working to establish best practices and are debating methods and the interpretation of results. Platt asserted further that decision making should fall to regulators and then to the community.

Selection of Outcomes to Monitor

An active surveillance system could be used in two ways: (1) to watch for potential adverse outcomes specified in advance, or (2) to evaluate signals arising from spontaneous reports or other sources. Thoughtful selection is necessary in choosing outcomes of interest to monitor by active surveillance. Spontaneous reports collected through passive surveillance indicate that the choice of outcome may be problematic, however. With active surveillance, it is more difficult to determine what outcomes should be monitored because data exist for every outcome that has occurred to every person in the database. While it would also be possible to use data mining approaches of the kind described by DuMouchel, Platt suggested

that the primary use for the system should be to focus on adverse outcomes for which the FDA already has cause for concern. He contended that the large majority of postmarket safety problems are caused by a relatively small set of candidates, so the first goal would be to conduct prospective surveillance looking for signals related to these candidates. Judy Racoosin, FDA, seconded Platt's suggestion that the experience gained during clinical trials should help guide the active surveillance program for a drug. During the FDA's preapproval safety conference, when the drug review division meets with the Office of Surveillance and Epidemiology (OSE), the participants discuss issues that have arisen during the development process. Racoosin explained that it is not uncommon to encounter a few cases of worrisome events and a greater number of more common events. Some events are difficult to understand because of the limited number of subjects in whom a drug has been tested. Some of the more obscure and confounded premarket data could be useful when selecting outcomes to monitor. Almenoff added that this is exactly what GSK does: every program, beginning in early development, has a risk/benefit management plan. From this plan is created a list of items of special interest that are monitored throughout the drug's life.

Approaches for Conducting Active Surveillance

Platt described two approaches for conducting active surveillance for prespecified events. The first is to wait until a sufficiently large number of exposures has occurred and then conduct a study. The weaknesses of this approach are that it is difficult to define a sufficiently large number, and it could take a long time to acquire the data. A second approach is to conduct sequential analysis—periodic data accumulation followed by periodic analysis, with each new analysis adding to the existing ones. This approach requires a method that allows for repeated testing on the same data.

There are a variety of ways to look for a signal in accumulating data. One standard method is the sequential probability ratio test (SPRT). A weakness of this method is that a threshold must be specified ahead of time for what constitutes an excess risk, and it is difficult to know what risk to specify in advance. If the correct excess risk threshold is chosen, the test can be highly effective and verify a risk very quickly, but if the wrong risk threshold is chosen, it may mask real risks.

Martin Kullforff, a statistician working with the VSD project, developed a variant of the SPRT called the maximized SPRT, which tests the null hypothesis (no excess risk) against the compound alternative hypothesis (a relative risk greater than one). The trade-off, from a statistical point of view, is that while this method can be used to test for any increased

risk, it is somewhat less efficient than would be the case if the excess risk were known.

The VSD project is using this technique to perform surveillance on new vaccines, including the meningococcal vaccine. In the latter case, at the end of 95 weeks, there had been two cases of thrombocytopenia relatively soon after immunization. That number does not exceed the signal threshold. But, Platt said, if the second case had occurred by week 12, that would have represented a statistically significant excess, and the project would have been faced with the question of what those two cases really meant.

A workshop participant expressed his opinion that SPRT is greatly limited for signal detection. Platt agreed that SPRT is not an appropriate method for signal detection and said that at present, no one knows what the best method is. He reiterated that maximized SPRT may prove to be useful, but suggested that researchers will need to explore and debate different methods until they find a better one. DuMouchel added that SPRT is designed to test a predefined hypothesis that is followed over time, and is not intended to be used when one is screening multiple drugs and multiple hypotheses. The next step will be to gain a better understanding of how sequential signal detection methods work under conditions that mimic real-life use.

What Is Needed

Creating an effective active surveillance system to monitor large numbers of therapeutic agents and outcomes will require consideration of several factors. Researchers will need to determine

- how to select the outcomes that will be monitored;
- how many outcomes can realistically be monitored;
- how to define outcomes in the terms in which they exist in the data sets;
- how often to look for those outcomes;
- what the appropriate comparators should be; and
- what the statistical approach should be.

Researchers will also need to develop rapid and effective ways of determining which signals represent real problems that require public health or regulatory action.

The maximized SPRT currently used by the VSD project has many desirable properties, but other sequential analysis methods should also be tested to determine which works best. Before making a decision, it will be necessary to evaluate each method to determine what its performance

characteristics are—in particular, how common false positives (signals when there is no real excess risk) and false negatives (failure to detect an excess risk that is actually present) are, and once this information is known, how to trade off between the two. It will also be necessary to decide how much error in each direction can be tolerated.

Problems to Overcome

Implementation of active surveillance will require that researchers overcome several barriers. For example, they will need to be able to determine whether the outcomes in a data set are real problems and not simply artifacts of the recording or analysis methods employed. Platt described the detection of a signal involving excess gastrointestinal bleeding associated with a new vaccine. After substantial time and effort, the signal was proven to be spurious, resulting from a change over time in the way the health plans' clinicians used certain diagnostic codes (more common use of codes that suggested gastrointestinal bleeding); the signal was highlighted by the change in documentation practices.

This is one example of many ways in which dynamic data systems developed to support health care delivery and payment can pose major challenges when one attempts to use them for surveillance purposes. If a signal appears not to be an artifact of the data systems, it will usually be necessary to validate the accuracy of the coded diagnoses for the cases by obtaining additional information from the associated medical records. Review of medical records will also be important in those cases to determine whether other factors are present that contributed to the outcome. Furthermore, even after it has been determined that there are more confirmed adverse outcomes than would be expected by chance, it will be necessary to disentangle the contribution of the drug or vaccine in question from other potential contributors, such as the underlying illness that was the indication for treatment or concomitant treatments. Finally, active surveillance will raise the issue of balancing benefits and risks to a new level of visibility. Because active surveillance will reveal risks of a drug that would otherwise have taken longer to detect—or perhaps would not have been detected at all—it will force a decision as to whether the benefits of continuing to use a drug in the way it has been used outweigh the risks uncovered by the surveillance.

Benefit of Maintaining the AERS Once a National Active Surveillance Network Has Been Established

Moderator Paul Seligman, FDA, asked Almenoff, DuMouchel, and Platt to comment on the value of maintaining a record system based on

voluntary reports of adverse events when an active surveillance network encompassing 100 million people is available. DuMouchel explained that while he is enthusiastic about the idea of an active surveillance network, he believes choosing the correct outcomes to monitor will be challenging. When data are reported to the AERS, a qualified health care provider has already decided that the event is important and should be explored. DuMouchel cautioned that without spontaneous reports, data could be entered into the system without undergoing such scrutiny, and therefore important outcomes could be missed. Although the AERS has a number of limitations as described earlier, until confidence in the ability of an active surveillance system to match the sensitivity of the AERS is established, spontaneous reports should not be abandoned. Platt agreed that spontaneous reports will be needed for the foreseeable future.

Almenoff suggested that an ideal way to approach this issue would be to include in electronic medical records a box that could be checked to indicate that the health care provider believed the occurrence was an adverse event, thereby flagging the event. Responding to this suggestion, Platt said his group is experimenting with "elicited surveillance," an electronic medical record system including a field designed to prompt clinicians to indicate when an event has occurred (diagnosis or laboratory result) that would not be expected. Using vaccines, this method was tested through comparison with the baseline reporting of the AERS. A five- to six-fold increase in the number of reported events was seen when clinicians were told that they had entered a diagnosis that would be unexpected for an individual who had recently been immunized, and asked whether this might be an adverse event for which they wanted to submit an AERS report. Though Platt believes this might be a good way of soliciting such information from clinicians, he expressed concern that many clinicians are hesitant to attribute unexpected outcomes to drugs, and therefore events could be missed.

9

Integration

As the workshop made clear, much of the work being done in safety science is focused in two areas: understanding and predicting toxicity in the discovery and preclinical stages of development, and spotting increased risk as soon as possible in the postmarket stage. As Janet Woodcock, Deputy Commissioner and Chief Medical Officer, U.S. Food and Drug Administration (FDA), noted in her introductory comments, however, the two areas should not be treated as separate spheres. Indeed, she said, one of the most significant challenges of emerging safety science will be to bring the two together. Ultimately, safety science research must be iterative, with insights from one area being used to provide direction for investigations in the other. The long-term vision is to have all areas of safety science fully integrated, from the discovery through the postmarket stages. To that end, several speakers described various ways in which such integration is being accomplished now and offered visions of what it might look like in the future.

AN INTEGRATION TOOL AT GLAXOSMITHKLINE[1]

Almenoff noted that although the FDA receives about 400,000 adverse event reports each year for marketed products, there is very limited systematic feedback of that clinical safety information to the discovery pipeline. And while there is likely a great deal of information embedded in

[1]This section is based on the presentation of Dr. Almenoff.

those reports that could be of value to researchers developing drugs, these data have not been leveraged. For example, adverse event reports may contain information that could help in discerning which chemical structures may and may not be associated with particular problems. Almenoff described one approach that GlaxoSmithKline (GSK) has taken to attempt to extract and utilize that unrevealed information.

Molecular Clinical Safety Program

GSK developed the Molecular Clinical Safety Program (MCSP) as a way of closing the knowledge gap among various disciplines and helping to minimize both safety risks and attrition in the drug pipeline. The program consists of a large data warehouse that will ultimately house information on about 21,000 compounds, plus a set of tools for working with those data. Each compound is anchored to a chemical structure that can be linked to all the information in the database: class and dose information, physical properties, toxicology, pharmacokinetics, and metabolism and bioassay data. Each compound is also linked to human safety data, including disproportionality analysis (DPA) scores calculated from data in the FDA's Adverse Event Reporting System (AERS), drug labeling information, and literature submissions. Ideally, the warehouse will also contain data and results from clinical trials, but because those data are more difficult to include, the system does not yet contain them.

The tools for interrogating the data include various statistical tools, such as recursive partitioning, as well as molecular mapping, visualization, and query and search tools. Starting with a particular drug, for example, it is possible to obtain its structure and then ask the database for a list of all other compounds with a similar structure, or perhaps all other compounds containing a particular substructure that can be used to compare a drug of interest with a reference drug. After a list of compounds has been generated, bioassay data can be obtained on all of those compounds, or the evaluator can examine those compounds that are on the market and evaluate their safety data. It is also possible to query for all the compounds that bind a particular molecule and examine the safety data associated with those molecules.

Example: Nelarabine and Neurotoxicity

Almenoff offered an example of how GSK researchers used this query tool to answer a retrospective question about nelarabine (a chemotherapy drug marketed by GSK as Arranon) that had been shown to cause demyelination in primate studies. Researchers wanted to determine whether interrogating the MCSP database could have provided them with information that might have made them proceed differently.

As noted, the MCSP database allows researchers to enter a drug of interest and ask for a listing of all similar compounds and any relevant toxicities. In this case, the researchers asked the system for marketed drugs with a chemical structure similar to that of nelarabine and any toxicities associated with those drugs. For three of the six compounds that resulted from the similarity search, there were serious neurotoxicity signals—such as demyelination and polyneuropathy—rated on the EBGM (empirical Bayes geometric mean) scale Almenoff had described earlier (see Chapter 8).

As it turned out, nelarabine did show some dose-related clinical neurotoxicity during development, but it is approved for use as a second- or third-line drug for refractory T cell lymphomas and leukemias. While it does have a very favorable risk/benefit balance for lymphoma and leukemia, it carries a black box warning for the demyelinating toxicity. If the MCSP tool had been available during the development of nelarabine, researchers might have seen the potential neurotoxicity problem early on and might have made a different decision about the drug's development, perhaps deciding to proceed with a backup drug, for instance. Even having early knowledge of the neurotoxicity issues, researchers might have decided to move forward with the compound because they believed it was still the best option; even so, however, they would have been aware of the types of outcomes that would ultimately need to be monitored for. Almenoff argued that the MCSP is one more tool in the decision-making process: it does not necessarily cause a change in the development path of a drug, but it does provide more information for use in the decision-making process.

Example: A Receptor Associated with Tardive Dyskinesia

A great many preclinical screens exist, and more are continually being developed. An important question is which of these screens are most predictive of clinical toxicity. Almenoff offered an example of how the MCSP tool was used to model toxicity data by linking human clinical safety data with molecular targets.

GSK researchers examined tardive dyskinesia (TD), which is a major adverse effect of schizophrenia treatment. Working from the entire AERS database, they screened for bioassay results that were most strongly associated with TD. The analysis was performed with 600 marketed drugs because those were the drugs for which GSK had compound profiling data at the time. Working from the EBGM scores that measured the drugs' statistically estimated risk factor for TD, the analysis looked for assays that discriminated between drugs with high and low TD signals. They found that certain catecholamine receptor subtypes, particularly at very high potency, were extremely strong predictors for TD. Thus, this analysis

was able to pinpoint the relationship between catecholamine inhibition and TD.

But the analysis also identified a strong association with another receptor. Although it is commonly believed that the action at the dopamine receptors explains the connection between antipsychotic drugs and TD, the analysis found six to eight compounds for which there was a high signal score, but the signal was better explained by this second receptor.

Almenoff warned that while the findings are still preliminary, the researchers believe they have discovered an association between a receptor and TD that was previously unknown. Antipsychotic drugs bind with this second receptor very avidly; furthermore, this receptor is localized in an area of the brain that modulates movement. Although this association has not been verified, the important point is that the MCSP tool can be used to discover new and potentially valuable information with data that already exist.

THE ELSEVIER DATABASE (PHARMAPENDIUM)[2]

Dr. MacLaughlin described a different approach to integration: an information tool that integrates preclinical, clinical, and postmarket safety data and makes it possible to look for patterns and connections among these data. In 1999 he was the principal investigator in a cooperative research and development agreement between Elsevier and the FDA to predict toxicological and adverse event end points. Leveraging historical data, the researchers sought to establish a comprehensive database comprising well-organized and integrated preclinical, clinical, and postmarket data using FDA approval packages.

Accumulating the Data

FDA approval packages, or summary basis of approval packages, consist of the medical reviews, pharmacology reviews, and summaries of other data collected on a compound reviewed during the FDA's approval process. These packages offer a rich source of information, but because of the format in which the FDA presents them publicly (e.g., nonindexed, nonsearchable paper; microfiche; bitmap formats), they are not readily accessible. Therefore the data cannot be queried for such information as class, target, and effect. To transform the data into a more usable format, Elsevier scanned roughly 750,000 pages character by character. Given the complexity of the data, it was necessary to have MDs and PhDs review

[2]This section is based on the presentation of Philip MacLaughlin, Senior Product Manager, Pharmaceutical Development, Elsevier.

and interpret them once they had been scanned. While this was a difficult and labor-intensive process, the final product yielded data in a format that could be accessed as a modern electronic document (i.e., it could be copied, searched, and manipulated electronically). The resulting database included records dating back to 1992 and sometimes earlier, totaling about 35,000 approval packages.

Organization and Context

One barrier encountered by the researchers was inconsistencies in the terminology used throughout the approval packages. Different packages used different terms for the same thing. For example, "electrocardiogram QT prolonged," "long QT," "QT increased," "QT interval prolonged," "prolonged ventricular repolarization," and "increased QT interval" are just a few of the terms used in the approval packages for a prolonged QT interval. Overcoming this barrier required a careful review of each package, followed by mapping of each term to a standard term, such as that given in MedDRA, the Medical Dictionary for Regulatory Activities.

Another barrier involved hierarchical terms. For example, if one is interested in retrieving data about ventricular arrhythmias from the database, it will be important for the database to recognize that there are many different types of ventricular arrhythmias: premature ventricular contraction, multifocal ventricular tachycardia, wide complex ventricular tachycardia, ventricular flutter, etc. This barrier can be overcome by organizing all of the terms into hierarchical thesauruses, so that it is possible not only to know when two terms mean the same thing, but also to understand their relationships—when one is a restricted case of another, for example. Such thesauruses had to be defined for all the different types of terms found in the approval packages, including adverse events, drugs, and targets. Overcoming this terminology barrier was a difficult process, and creating these thesauruses required qualified and properly trained scientists, of whom there is currently a shortage. Once the thesauruses were completed, however, the team had a way to classify and find relationships for every term found in the approval packages (see Figure 9-1). The lack of standardized terminology throughout the drug development industry will continue to pose a major barrier to the integration of datasets.

Strategic organization of the database is critical. One method for efficiently organizing the data is to use drug names and structural chemistry as the foundation for anchoring all other data. If this method is used in conjunction with the drug thesauruses described above, then regardless of what compound name is queried—generic or trade name—the correct compound with its structure and all the other links to related information will be retrieved. Once the thesauruses had been built and the data

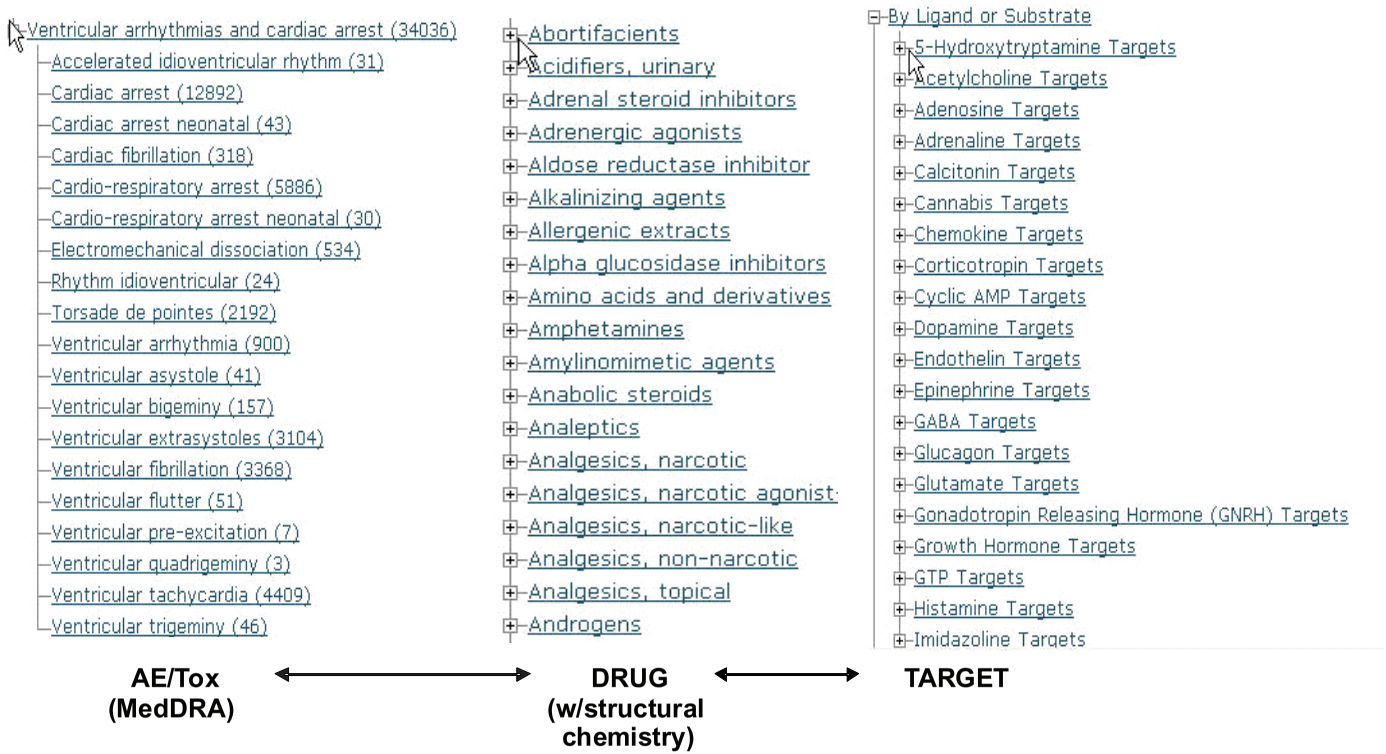


FIGURE 9-1 Example of hierarchical thesauruses for adverse events, drugs, and targets.
NOTE: MedDRA = Medical Dictionary for Regulatory Activities.
SOURCE: MacLaughlin, 2007.

extracted from the now-digitized approval packages, the resulting database, with all the original information in standardized form, could be interrogated.

MacLaughlin described the work of a pharmaceutical company that looked for comparative exposure data, particularly in humans, but also in animals. After 3 years of combing the literature and validating experiments in an effort to create its own database of a single end point, the company had identified a total of 600 drugs with this end point. If the database described above had performed this search, it would have been possible to retrieve within moments data on more than twice as many drugs—1,400 instead of just 600—and with a full population of parameters, not just a single data point.

Summary

In summary, MacLaughlin offered several take-away messages and next steps:

- There is no easy solution to the integration of data sets; however, properly planning for the future will facilitate the effort. The key is to have a uniform, standardized database spanning the entire development process.
- The lack of standardized terminology throughout the drug development industry and the health care community is a major barrier.
- Using this database, it is possible to look at preclinical, clinical, and postmarket data simultaneously; identify all compounds with a certain substructure and a certain target; and list all the toxicological effects of a particular type.
- The database described above is available to reviewers at the FDA today, but more generally, its creation can serve as a template for similar efforts—for example, for the collection and organization of the kinds of toxicogenomics data described in Chapter 4.

AN FDA PERSPECTIVE[3]

Dr. Frueh offered an FDA perspective on emerging safety science, with a particular focus on what is needed to integrate the various stages of drug development. Within the field, he explained, there are two main questions the FDA is interested in answering: At the preclinical stage, is it possible for tests to screen out compounds that have the potential

[3]This section is based on the presentation of Felix Frueh, Associate Director for Genomics, Office of Clinical Pharmacology, Center for Drug Evaluation and Research, FDA.

to induce toxicity in humans? And at the clinical stage, is it possible to develop tests or diagnostics that can measure the probability of drug-induced toxicity?

As described by the workshop's presenters, researchers are working to improve safety at each stage of drug development. At the preclinical stage, for example, gene expression analyses are being used to predict toxicity. In addition to various technologies under development, Frueh emphasized that it will be vital to develop feedback loops through which information from one stage of development is used to make safety assessments at another. At each stage of development, researchers should be thinking about how to apply the information generated at that stage to early development so they can establish better tests, enhance the usefulness of biomarkers employed in the drug development process, and improve decision making.

The Need for Bridging Biomarkers

While it is important to create feedback loops that enable data from later in the process to inform technologies and decisions in earlier stages, it is also important to create forward connections. Frueh stressed the importance of having bridged biomarkers from the preclinical to the clinical stage in the event that preclinical information could predict what might happen in the clinical stage (see Figure 9-2).

While a number of traditional nonclinical test systems are used to assess safety—animal tests, tissue slices, cell cultures, *in silico* models, etc.—they are not always good predictors of toxicity in humans. Furthermore, many of these systems rely on the signal generated by a toxicity state, and it would be preferable to have tests that can identify toxicity before it manifests and is detectable by histopathology assessments.

In contrast to these traditional methods, the biomarkers described by many speakers during the workshop are, in a sense, surrogates for toxicity—they indicate when toxicity will develop but before it actually does so. A number of such markers are available today, and many others are being developed. While they may work well for compound selection and early characterization, however, they are not necessarily good predictors of a toxic event in humans.

Creation of a more efficient drug development paradigm will require the establishment of biomarkers that can bridge or translate early preclinical findings to the clinical stage. These biomarkers would be of the same type in preclinical models and in humans and would represent quantifiable indicators of normal biological processes, pathophysiological states, and responses to therapeutics. In some fields, studies to identify and validate such bridging biomarkers are already under way. For example, the

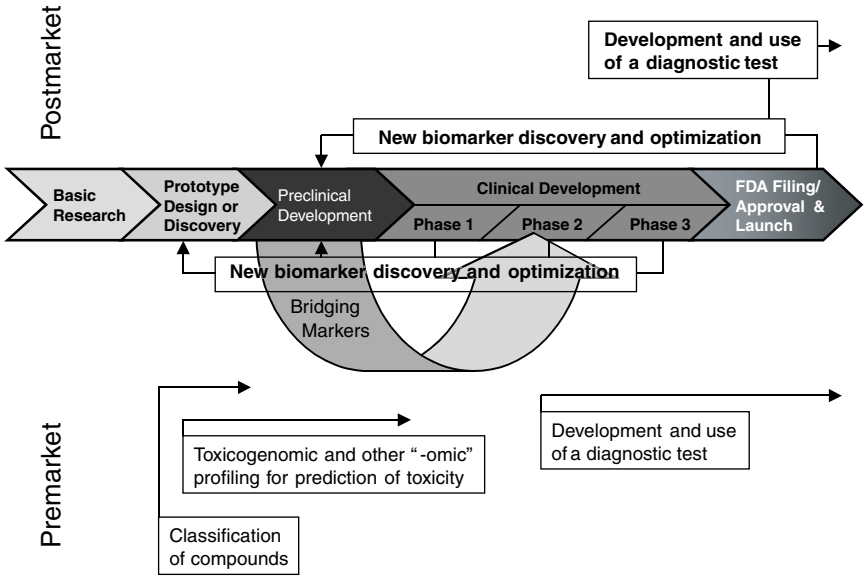


FIGURE 9-2 Drug development as an iterative process. The figure illustrates how biomarkers may be used to bridge, or translate, early preclinical findings to clinical findings, and how clinical findings may be used to inform and corroborate the basic science. Many of the workshop participants emphasized that the ultimate goal of applying these new technologies in safety science is to create a continual iterative process in which the basic scientific data can help to inform and predict clinical outcomes.
SOURCE: Frueh, 2007.

Predictive Safety Testing Consortium, a partnership between the FDA's Critical Path Initiative and a number of large pharmaceutical companies that is aimed at predicting the safety of new treatments before their use in humans, is investigating the analytical validation of a set of nephrotoxicity markers, and in the near future it will consider methods for establishing clinical qualification.

Such bridging biomarkers would likely be highly useful in exploratory IND (investigational new drug) work. If these markers are indeed mechanistic and provide information about underlying toxicity from a molecular mechanistic point of view, it should be possible to discern toxicity signals very early in humans—before harm to organs occurs—and at very low doses. Frueh hypothesized that the availability and measurement of a variety of bridging markers could well result in the creation of a

completely new "phase 0" safety package. Although this idea is at present only conceptual, he believes there is true potential for the application of such bridging biomarkers in very early-stage drug development.

Dealing with Idiosyncratic Events

Frueh explained that in addition to developing biomarkers to predict preclinical and clinical events, it is necessary to establish biomarkers for dealing with idiosyncratic events—random, unexpected, often dose-independent adverse drug reactions—that occur during clinical testing and once a drug is on the market. These events are generally caused by an interplay between the properties of a drug and the predispositions—genetic and otherwise—of the patient. In such cases, the only option is to learn after the fact, studying the event to understand what caused it.

Withdrawal of drugs from the market because of the occurrence of idiosyncratic events is harmful not only to patients suffering the adverse events, but also to patients not at risk who would otherwise benefit from a drug. However, the only way to reduce the occurrence of idiosyncratic events is to develop processes and invest in research that can lead to a reduction in all adverse events, both serious and nonserious. As an example of this approach, Frueh described work on drug-induced long QT syndrome. This is an idiosyncratic, rare event, but after hepatotoxicity, it is the top reason for drug withdrawals. The effect is generally reversible—if a patient stops taking the drug in question, he or she reverts to a normal QT—but it can sometimes lead to torsades de pointes, a condition that can be fatal. While not all QT prolongation leads to torsades de pointes, it is impossible to predict when this will occur, so it is necessary to regard QT prolongation as a marker for the potential development of that fatal effect.

Since drug-induced long QT syndrome occurs in a wide variety of structurally diverse compounds, it is impossible to make a class prediction. Even though researchers believe they understand some of the mechanisms involved—namely, many of the drugs that induce long QT syndrome are KCNH2 (HERG) blockers—other factors clearly play a role as well. For instance, many drugs that block the same channel do not induce long QT syndrome, and therefore it is not easy to predict which drugs will do so.

To predict which drugs may induce long QT, researchers would need to design a study that could identify new genetic biomarkers that could be used to determine whether a drug had the potential to cause prolonged QT. Such a study would need to consider the influence of external factors, such as medications or other exposures, but it would also need to look closely at genetic factors. Although some people have a genetic predispo-

sition to congenital long QT syndrome, the same mutation can also lead to different phenotypes. Some with the mutation have QT prolongation in the absence of drugs, while others have it only in the presence of a drug; this genetic mutation is responsible for about 10 percent of cases of drug-induced long QT syndrome. The study would also need to consider other relevant genotypes, such as CYP2D6 and drug-metabolizing enzymes.

The bottom line is that many genetic factors likely play a role, and it will be necessary to identify these factors and perhaps formulate some risk pattern that would make it possible to assess an individual's risk of experiencing long QT syndrome when given various drugs. Frueh hypothesized that most likely, a genome-wide single nucleotide polymorphism (SNP) analysis conducted in a large number of patients using a variety of different drug classes and taking all other factors into consideration will be required to develop a hypothesis about the phenotype–genotype association underlying this phenomenon. Once this has been accomplished, the association can be qualified with a separate data set, and ultimately, causation can be established, molecular mechanisms mapped out, and the resulting understanding applied in the clinical setting to avoid drug-induced long QT syndrome in as many patients as possible.

Monitored Release

Regardless of how proficient scientists become with bridging biomarkers and understanding of idiosyncratic events, it will never be possible to know for certain whether a drug is totally safe. The main problem is that the safety databases generated during drug development are generally too small to highlight rare events successfully. The largest amount of safety data is actually produced once a drug is on the market, but current tools do not allow scientists to capture this information effectively and capitalize on its potential. Typically, there are tens of patients in phase I trials, tens to hundreds in phase II, and hundreds to thousands in phase III. If an adverse event occurs in 1 of every 5,000 or 10,000 patients, it may be impossible to detect such an event prior to the drug's approval.

Frueh suggested that the problem could be addressed by instituting a system of monitored release (see Figure 9-3), which would be invoked after a drug's initial but before its final approval. In such a system, the first 100,000 patients (or whatever number was selected) to take a new drug would be monitored for adverse events. Samples would be collected from any patients who experienced such events, along with samples from a matched group of controls, and these samples would be analyzed to identify the factors leading to the adverse event. Once it was possible

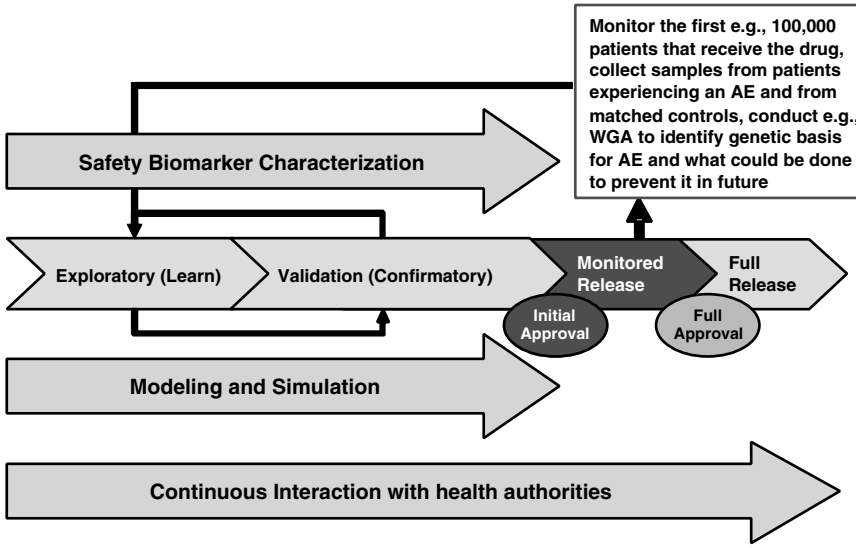


FIGURE 9-3 Illustration of a product development timeline that includes a monitored released phase. This phase would be invoked after a drug's initial but before its final approval. Throughout monitored release, samples would be collected from patients to enable study of the genetic basis of adverse events.
NOTE: AE = adverse event; WGA = whole genome association [studies].
SOURCE: Frueh, 2007.

to identify at-risk patients and remove them from the population being prescribed that drug, the drug could proceed to final approval.

Summary

Frueh summarized the potential of these emerging technologies and some of the future barriers scientists must overcome:

- The emergence of new molecular biomarkers for drug safety will make it possible to better bridge the safety gap between the preclinical and clinical stages; the hope is that eventually, having true translational biomarkers will transform the process into a continuum.
- Because a drug's safety can never be completely proven, researchers will always have to rely on the absence of signals; however, the process of looking for such signals can be significantly improved.
- The development of better characterizations for toxicity will lead

to better markers for toxicity, but accomplishing this will require a true interdisciplinary approach involving experts in all relevant fields.

- Emerging safety science has already made it possible to better classify compounds through new genomic and other technologies.
- Researchers need to qualify markers for bridging studies, as well as those for addressing idiosyncratic events.
- Nephrotoxicty biomarkers will be submitted to the FDA for review this year, and a process for reviewing these markers is being established.[4]
- Genomic association studies (including, for example, whole-genome SNP scanning) have the potential to identify markers for rare adverse events, but access to well-characterized samples remains a problem.
- New mechanisms and processes for studying clinical and postmarket safety need to be explored.

[4]Following the workshop, the Predictive Safety Testing Consortium's nephrotoxicity biomarker package was submitted to the FDA for review.

10

The Future of Safety Science[1]

During the general discussion sessions at the end of each day of the workshop, participants summarized and synthesized the presentations made, discussed gaps and needs for the future, and suggested next steps. Much of the session on the first day was devoted to issues surrounding prediction, while the session on the second day focused mainly on issues concerning surveillance. There was significant overlap, however, as well as a good deal of discussion of how best to integrate the two areas. In the workshop's final session, Dr. Krall considered the presentations and discussions that had taken place during the workshop and offered some general observations about the field and the future of safety science in these areas.

PREDICTION

Prediction of the safety and efficacy of drugs is paramount to revitalizing the present drug development paradigm. Prediction that can detect potential problems in advance of clinical testing or market approval will allow for safer delivery of medicines, vaccines, and medical devices, and even the performance of safer surgeries. Traditional drug safety detection methods have generally depended on animal testing, with the assumption that the results of these tests can be indicative of what will happen

[1]This chapter is based on the presentation of Ronald Krall, Senior Vice President and Chief Medical Officer, GlaxoSmithKline, and the contributions of several other workshop participants.

in humans. Emerging safety science holds promise for enriching this traditional approach.

Much of emerging safety science is predicated on going beyond observations in whole animals or in organ systems to look at what is happening within a cell—to understand which pathways are perturbed, for example. Adding this information to the traditional approach could even allow researchers in the future to sidestep the traditional animal experiments altogether because of the ability to predict directly what will happen in humans. One advantage of studying actual human cells and human pathways is that it obviates the need to extrapolate from other species. Furthermore, as various presentations at the workshop demonstrated, this approach offers an explanatory power that is lacking with the traditional animal experimentation. By providing information such as gene transcription data, emerging safety science techniques can offer insight into what pathways, targets, or receptors have been perturbed. This information can then be applied to understand and predict the kinds of events that can be expected in humans who take a drug.

Thomas Caskey, of the University of Texas Health Science Center at Houston, identified two areas he believes will be important for prediction in the future but were not discussed as thoroughly at the workshop as some others. The first is the use of protein assays. He argued that although proteomics technology is more challenging and not as well developed as transcriptomics or metabolomics, proteomics assays can be easy to conduct when one knows what to measure, and will eventually prove to be important. The second area is imaging. Alluding to Westwick's description of an imaging technology used in a human cell–based screening technique (see Chapter 3), he suggested that such imaging technologies will likely prove to be very powerful because they can be extended from cells to whole animals, and thus be used to determine whether what is seen in individual cells is also seen in more complex systems.

The techniques described during the workshop, such as gene transcription and metabolomics, are already being used to discriminate among drug candidates. Throughout development, they are being used to help select targets, classes, or doses. There are also examples of their being used to prevent adverse events in humans and of markers being identified to help monitor for effects in humans, thus minimizing the chances of drug-induced injury. Finally, these techniques are being used to gain additional information about cellular pathways and signaling, thereby increasing understanding of why certain events occur and offering insights into potential new targets.

SURVEILLANCE

Traditional surveillance approaches that rely primarily on data collected from spontaneously reported adverse events are a valuable source of information, and with recent advances in information technology, these approaches hold additional potential. Nonetheless, much of the discussion at the workshop focused on active surveillance and the identification of better ways to track clinical experience. Although advances are being made that may help in predicting responses prior to use, active surveillance will always be necessary, since it will never be possible to predict with certainty what will happen when a new drug is introduced in the market and large numbers of people begin taking it.

Utility of Active Surveillance Systems

A system capable of detecting increases in classic events that could lead to the withdrawal of drugs from the market would be tremendously valuable. Krall referred to programs that GlaxoSmithKline (GSK) has implemented to actively monitor large health care system databases and detect patterns of postmarket adverse events once a drug is on the market.

To illustrate the utility of such a system, Krall used the example of a drug that was ultimately withdrawn from the market. Without identifying the drug or the adverse event it caused, he explained that a disproportionality analysis of data from the Adverse Event Reporting System (AERS) database indicated that the event was occurring with this particular drug much more often than with other drugs in the database. The excess was apparent from the first year of marketing and continued for as long as the drug was on the market. To see whether they could detect the same event using large health care system databases, GSK researchers chose two databases—the Integrated Health Care Information Services claims database and an electronic medical records database from General Electric. The databases were large, one containing 40 million people and the other 5 million. The researchers found they were able to detect the event in question, and Krall exhibited a graph that showed the rate calculated in terms of events per 10,000 patients (see Figure 10-1). The confidence intervals were relatively narrow, and the graph showed that the event was taking place at a rate that was at least double that for other drugs in the same class. This example showed it is possible to search for and identify events of interest in large health care databases using a prescribed set of tools and methodologies, just as events are found using data collected from spontaneously reported adverse events. Furthermore, such an analysis can expand on the findings from a spontaneous adverse event reporting system.

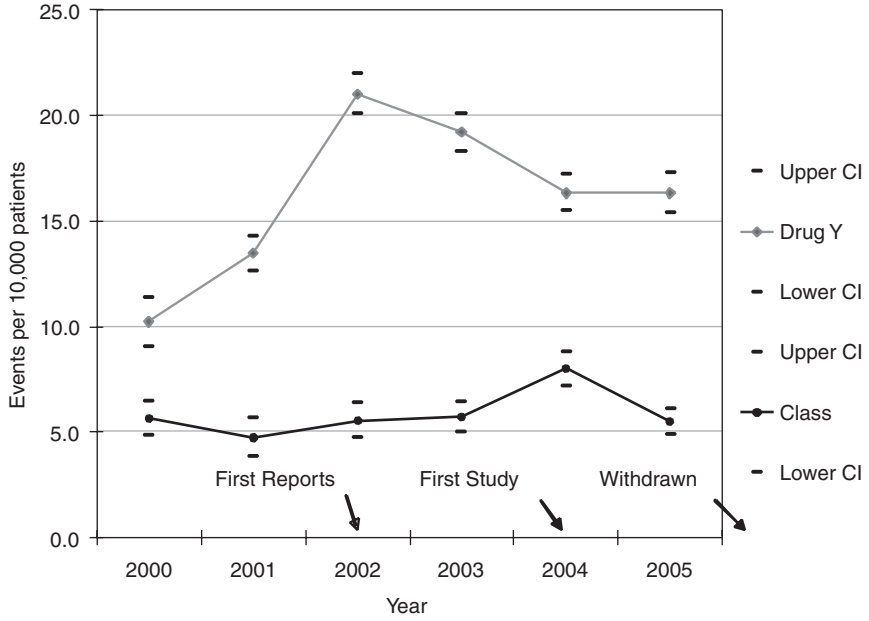


FIGURE 10-1 Use of large health care databases to identify events of interest. The graph describes the rates of occurrence of event X using observational data. The results showed that the event was taking place at a rate that was at least double that for other drugs in the same class. The same results were achieved by performing disproportionality analysis of data from the Adverse Event Reporting System.
SOURCE: Krall, 2007.

Achieving an Active Surveillance System

Because the largest amount of safety information is produced when a drug is on the market, it is important to capture those data. Yet current tools do not make it possible to capitalize adequately on this information. Krall said that stakeholders have an obligation to share their knowledge with the larger society. Looking to the future, Krall and several other workshop participants reviewed a number of ways to meet the challenges involved in realizing the goal of an active surveillance system.

Enhancing Data Sharing

Peter Corr, retired from Pfizer, echoed the importance of sharing data and technology. He suggested that the best way to move forward quickly would be for companies to combine their efforts. This is already

happening, for example, in toxicology with the Predictive Safety Testing Consortium. Corr argued that this sort of openness should be expanded to include other areas. Further, while a great deal of meaningful work on individual compounds was described during the workshop, a comprehensive understanding of the relationship between molecular structures and toxicity will demand the study of many diverse compounds. A large amount of data already exists in various pharmaceutical companies—far more than is ever submitted to the U.S. Food and Drug Administration (FDA) with drug applications—but the data are not shared. Acknowledging arguments for keeping the data proprietary, Corr suggested that there are even better arguments for sharing the data. Combining forces would have a huge effect on the diversity of available data and thus on the ability to understand the relationships of interest.

Paul Seligman, of the FDA's Center for Drug Evaluation and Research, echoed the need to share information. One of the greatest challenges facing the field is achieving access to all of the information accumulated, particularly that on products that fail during their development. Dissemination of negative data could enable more efficient drug development paradigms. Almenoff added that obtaining data on failed compounds is crucial. To date, most data mining has been done on "honor role molecules"—those that made it through the various testing phases and have some promising attributes. Data mining with failed compounds would be useful as it could help improve prediction.

Standardizing Nomenclature

As discussed earlier, different databases and even different records within the same database use varying names for the same drug. They also use varying names or descriptions for the same medical condition.

Several participants emphasized the importance of creating standard formats for information about drugs and their biological properties and actions. For example, Ana Szafrman, FDA, called for unique names for drug products. Giving drugs unique names and using those names consistently would make it much easier to link information from different databases.

Another workshop participant from GSK expanded on this point. Having been involved over the past 2 years in a GSK initiative aimed at linking quantitative clinical data with basic science information, he has found the biggest challenge to be the lack of data standards within basic science databases and the lack of consistent names for drugs being studied. He noted that even working with data from GSK's own databases has been difficult, in part because the various companies that merged to form GSK each had their own formats, and standardizing the data has been

tedious. He suggested that an industrywide FDA-recommended standard for collecting and describing data would help prevent these problems in the future. He emphasized that the field of toxicology in particular could benefit from such standards because currently, some toxicology data are provided in qualitative text strings, which are very difficult to compile.

Mary Prince Panaccio, of Merck, spoke to the lack of standardization in the postmarket stage. Researchers must work with spontaneous reports that have no common structure or set of details. Standardizing the terminology used to describe postmarket events would make it easier to feed that information back into the basic science work being done on prediction.

Improving the Comprehensiveness and Linkage of Data Sources

Data sources seldom provide all the data needed. While there may be individual records of data collected from a hospital stay or an outpatient visit, rarely are the data sets linked. Furthermore, the medical record data may not be linked to X-ray data, laboratory data, or pharmacy data that would indicate whether prescriptions were actually filled. In addition, most data sources are missing medical content—content that is often available only in doctors' charts or notes and in forms not easy to access. As a result, there is no continuous record of data on an individual in any of these systems, so it is impossible to capture a person's life experience. Because each of these systems captures only a slice of that experience, it is difficult to create longitudinal records.

The multiplicity of data sources also hinders the development of an active surveillance system. Claims data are very different from health record data, and health record sources differ from one organization to another. The way an electronic health record is implemented in a health care system has a great influence on which data actually exist and on how easy it is to find associations in those data.

Instituting Electronic Record Keeping

Krall suggested that to capture medically important information that is not being captured in current health care record systems, the best approach would be to institute electronic health records. With such records, it would be possible not only to get more from the data that exist, but also to obtain more data. The benefits of having electronic health records were indeed amply demonstrated during the workshop. With such records, it would become possible, for example, to link what is being learned about cellular pathways and cellular signaling with clinical information about a disease and various interventions used to treat

the disease. Instituting electronic health records will be an enormous challenge but will ultimately pay off in many ways, including ones that cannot even be imagined today. One barrier to be overcome, however, is that the education of health care practitioners generally does not cover how to keep such records.

Conducting Research on Analytical Methodology

Presentations made throughout the workshop demonstrated the power of various analytical methodologies developed to date, but more work in this area is needed. It is important to continue to learn about how to discern and to evaluate and assess the signals that appear in health care system databases. Research on analytical methodologies should be built into any approach to active drug surveillance.

Addressing Issues Inherent in Data Sources

Panaccio stressed the importance of understanding how to interpret information collected from a health care plan database. Because the population of that database will not be the same as the general population, it is useful to examine cohorts within the health care plan in an effort to understand what the patients in these cohorts looked like before the drug of interest was marketed—for example, what sorts of adverse events were reported.

John Jenkins, of the FDA, reinforced Krall's comment about using biological understanding to hypothesize the types of events that should be monitored once a drug has been marketed. Having a better idea of what events to monitor for in the postmarket setting can help identify those questions that should be answered before approval and those that can be answered after approval. The focus is generally on serious adverse events, which fall into two categories. The first is rare serious adverse reactions, such as hepatotoxicity, that in general will not be detected in clinical trials; the AERS does a fairly good job of picking these up, although it could probably be improved with a more active monitoring program. The second category consists of drugs that cause an increase in the rate of common adverse events, such as heart attacks or strokes. Determining the best way to detect these types of events will require serious thought and discussion, weighing an active postmarket surveillance system against very large controlled clinical trials.

Incorporating Phenotyping into Routine Clinical Trials

Krall suggested that one improvement to the current surveillance system would be to gather more information about patients. In general, research on the effects of drugs has been approached from the point of view of the clinical trial, where researchers compare the results for a test group with those for a control group and look for differences between the two. It would be very valuable, however, to phenotype all of the subjects in those trials in such a way that it would be possible to discriminate among groups and identify biomarkers that could be used to predict how different people will respond to a drug.

On a similar note, Caskey said it will be important to use "scanning markers" in postmarket surveillance programs as a way of picking up signals of impending toxicity. These will be different from the biomarkers used in the premarket phase, when researchers are studying a particular target and looking for a response. In the postmarket surveillance phase, it may not be clear which targets may be involved, so it will be necessary to have some general scanning markers that measure various aspects of metabolism. Over time, as data are accumulated, it should become possible to zero in on markers that are associated with—and preferably predictive of—the eventual appearance of an adverse event.

Integrating Basic and Clinical Science

Referring to the presentation by Almenoff, who described her company's Molecular Clinical Safety Program (see Chapter 9), Edward Holmes, of the A*Star Biomedical Research Council, asked how many other examples exist of attempts to link basic science data with clinical data. Integration of these two areas remains at this point more hope than reality, but there was some discussion of what might be needed to achieve such integration in the future.

Mikhail Gishizky, of Entelos, said that one of the major challenges will be dealing with the overwhelming amount of data. This sort of data challenge has been met in other industries, he said, and it will be important to look at these other industries and learn how they have been successful through the use of computer modeling and other technologies. Gishizky also suggested that an appropriate metaphor for what is needed to link the basic sciences to the clinical setting is the Rosetta stone: researchers must find some way to translate information from the basic sciences into the clinical setting and vice versa. Once again, a number of researchers commented that a vital first step in this process of translation will be to develop standardized terminology. If information is to be shared across the life cycle of a drug, basic researchers and clinicians must at the very least be speaking the same language.

Robert Califf questioned whether the integration of the basic sciences and the clinical setting could ever be realized. The current system identifies many preventable adverse drug events, events caused by situations that are very well described, yet adequate clinical systems to deal with them are not in place. In prescribing antithrombotic drugs, for example, the wrong dose is given about a third of the time. In such cases, one must administer a test whose results are not available for 2 days, and then figure out how to deal with the problem. Woodcock disagreed, suggesting that if a new technology is introduced with explicit instructions for its use, the health care system will apply it. As an example, she pointed to the experience with abacavir (see Chapter 6).

Along the same lines, Frazier commented that health care providers will not actively follow the search for biomarkers and adopt each as it is discovered. Instead, when a biomarker is validated as being clinically useful, doctors will adopt it. If doctors are provided with a useful bottom line, they will apply it.

Caskey said that while analyzing their compounds, many of the large pharmaceutical companies use different tests to measure the same outcome. Thus the decision that is made at Pfizer will not be the same as that made at Merck or that made at Abbott. Caskey suggested that the FDA undertake a research initiative to determine which of these tests are most effective in predicting clinical safety. When a drug was approved and launched, it could be subjected to the testing systems proposed by each of the companies, and the actual clinical results could be compared with those of the various testing systems. Caskey suggested that partial funding for these efforts could come from the National Institutes of Health.

SUMMARY

In her concluding comments, Woodcock said it will be important to keep an eye on the long-term goal. That goal is not just to fix problems that occur when a drug enters the market. Rather, it is to move medicine to a more scientific basis, something for which the necessary tools exist. What is lacking is the system to make it happen. Summarizing the workshop's take-away messages, Woodcock said that efforts to create standards should be greatly intensified, especially in areas in which data from different sources will be linked. She emphasized that the science is emerging, and the community needs to ensure that it is put to the best use as quickly as possible; the next steps need to be considered and discussed now.

References

Almenoff, J. 2007. *Signal management through the product lifecycle*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 24, 2007, White Oak, MD.

Butcher, E. 2007. *Drug evaluation in human cell systems-based models of physiology and disease*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

Cockett, M. 2007. *Toxicogenomics and assessment of drug pharmacology using microarrays*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

DiMasi, J. A., and H. G. Grabowski. 2007. The cost of biopharmaceutical R&D: Is biotech different? *Managerial and Decision Economics* 28(4–5):469–479.

DuMouchel, W. 2007. *Statistical issues in the analysis of spontaneous report databases.* Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 24, 2007, White Oak, MD.

FDA (Food and Drug Administration). 2004. *Innovation or stagnation: Challenge and opportunity on the critical path to new medical products*. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf (accessed July 11, 2007).

FDA. 2006. *Innovation or stagnation: Critical path opportunity list*. http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_list.pdf (accessed October 25, 2007).

Frazier, K. 2007. *ALK5: Targeted investigations of a targeted therapy using laser capture microdissection, flow cytometry, immunohistochemical approaches and genomics.* Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

Frueh, F. 2007. *The transition from preclinical to clinical application of safety related genomics.* Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

GAO (Government Accountability Office). 2006. *Drug safety: Improvement needed in FDA's postmarket decision-making and oversight process*. http://www.gao.gov/new.items/d06402.pdf (accessed January 17, 2008).

Halbert, D. 2007. *Application of toxicogenomics to drug discovery and to preclinical safety assessment*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

IOM (Institute of Medicine). 2007. *The future of drug safety: Promoting and protecting the health of the public*. Washington, DC: The National Academies Press.

Krall, R. 2007. S*umming up: Improving safety science to make better medicines*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 24, 2007, White Oak, MD.

Lai, E. 2007. *Pharmacogenetic investigation of drug associated adverse events*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

MacDonald, M. L., J. Lamerdin, S. Owens, B. H. Keon, G. K. Bilter, Z. Shang, Z. Huang, H. Yu, J. Dias, T. Minami, S. W. Michnick, and J. K. Westwick. 2006. Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nature Chemical Biology* 6(2):329–337.

MacLaughlin, P. 2007. *Integration of preclinical, clinical, and postmarketing safety data.* Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

Milburn, M. 2007. *Metabolomics as an emerging technology in drug safety assessment.* Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

Spear, B. 2007. *Practical application of toxicogenomics in early drug discovery*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

Vonderscher, J. 2007. *Biomarker of drug induced kidney injury qualification for regulatory decision making*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

Westwick, J. 2007. *Tapping the power of cellular networks for improved drug discovery*. Speaker presentation at the Institute of Medicine Workshop on Emerging Safety Science, April 23, 2007, White Oak, MD.

Appendix A

Workshop Agenda

EMERGING SAFETY SCIENCE

April 23–24, 2007
FDA White Oak Conference Center
Silver Spring, MD

Workshop Objective: Safety science is a rapidly changing field, leading to new developments in the methods and technologies used to detect and interpret safety signals. Predictive tools are needed to accurately screen out candidates, early in development, that are most likely to have serious, undesired effects and identify those most likely to become safe and effective treatments. The goal of this workshop is to present and collaboratively discuss novel, cutting-edge methodologies and techniques that are being used by academicians, researchers, drug manufacturers, and regulatory scientists. It will also explore how this new knowledge and technology may be applied for both drug development and postapproval regulatory/safety review processes. Day 1 (Emerging Safety Science: Biology of Adverse Events) will cover primarily preclinical safety issues, and Day 2 (Emerging Safety Science: Data Mining from the Medical Experience) will focus on postmarket surveillance/pharmacovigilance topics.

Monday, April 23: The Biology of Adverse Events

Welcome and Opening Remarks

8:30–8:35 **Edward Holmes**, Workshop Chair
Co-Chair, Forum on Drug Discovery, Development, and Translation
Executive Deputy Chairman
A*Star Biomedical Research Council,
National University of Singapore

Meeting Objectives: Merging New Science and Drug Review

8:35–8:45 **STEVE GALSON**
Member, Forum on Drug Discovery, Development, and Translation
Director, Center for Drug Evaluation and Research
U.S. Food and Drug Administration

JANET WOODCOCK
Member, Forum on Drug Discovery, Development, and Translation
Deputy Commissioner and Chief Medical Officer
U.S. Food and Drug Administration

Human Cell System–Based Approaches to Signaling and Biology

8:45–9:40 Moderator: **DAVID JACOBSON-KRAM**
Associate Director for Pharmacology and Toxicology
Office of New Drugs
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

20 min. Drug Evaluation in Human Cell-Systems: Biology-Based Models of Physiology and Disease

EUGENE BUTCHER
Cofounder and Chair of the Scientific Advisory Board, Bioseek
Professor, Department of Pathology
Stanford University School of Medicine

20 min. High-Throughput, High-Content Cellular Screening for Definition of Drug Mechanisms, Selectivity, and Safety

JOHN K. WESTWICK
President and CSO
Odyssey Thera, Inc.

15 min. **DISCUSSION**

Biomarkers of Toxicity in Drug Discovery and Development

9:40–12:30 Moderator: **Federico Goodsaid**
Senior Staff Scientist in Genomics
Office of Clinical Pharmacology, Office of Translational Sciences
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

20 min. Toxicogenomics and Assessment of Drug Pharmacology Using Microarrays

Mark Cockett
Vice President, Applied Genomics
Bristol-Myers Squibb Company

20 min. Application of Toxicogenomics to Drug Discovery and to Preclinical Safety Assessment

Don Halbert
Executive Vice President of Research and Development
Iconix Pharmaceuticals

10:20–10:35 **Break**

20 min. Practical Application of Toxicogenomics in Early Drug Discovery

Brian Spear
Director, Genomic and Proteomic Technologies
Abbott Laboratories

20 min. Gene Expression Profiling in Rat Exploratory Toxicology Studies: Why and Where Is It Useful?

Eric Blomme
Project Leader, Cell and Molecular Toxicology
Abbott Laboratories

20 min. Integration of Preclinical, Clinical, and Postmarket Safety Data

PHILIP MACLAUGHLIN
Senior Product Manager, Pharmaceutical Development
Elsevier

20 min. Qualification of Drug-Induced Nephrotoxicity Biomarkers

JACKY VONDERSCHER
Vice President, Head of Exploratory Development in Europe
Novartis

20 min. The Transition from Preclinical to Clinical Application of Safety-Related Genomics

FELIX FRUEH
Associate Director for Genomics
Office of Clinical Pharmacology
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

15 min. **DISCUSSION**

12:30–1:30 **Lunch**

Metabolomics in Drug Safety

1:30–2:25 Moderator: **SHIEW-MEI HUANG**
Deputy Director
Office of Clinical Pharmacology
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

20 min. Metabolomics as an Emerging Technology in Drug Safety Assessment

MICHAEL MILBURN
Chief Scientific Officer
Metabolon, Inc.

20 min. Targeted Metabolomics in Pharmacodynamics and Toxicology

KLAUS M. WEINBERGER
Chief Scientific Officer
Biocrates Life Sciences

15 min. **DISCUSSION**

Targeted Therapy

2:25–3:20 Moderator: **GEORGE ROCHESTER**
Lead Mathematical Statistician
Quantitative Safety & Pharmacoepidemiology Group
Office of Biostatistics
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

30 min. ALK5: Targeted Investigations of a Targeted Therapy—Using Laser Capture Microdissection, Flow Cytometry, Immunohistochemical Approaches, and Genomics

KENDALL FRAZIER
Director of Cellular & Molecular Pathology
GlaxoSmithKline, Safety Assessment

25 min. **DISCUSSION**

3:20–3:35 **Break**

Abacavir: A Working Example of PGx Investigation of Drug-Related Adverse Events

3:35–4:15 Moderator: **KENDALL MARCUS**
Medical Team Leader, Division of Antiviral Products
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

30 min. Abacavir: A Working Example of PGx Investigation of Drug-Related Adverse Events

ERIC LAI
Vice President
PGx Experimental Project Coordination and Analysis
GlaxoSmithKline

10 min. **DISCUSSION**

Where Is the Science Taking Us?

4:15–4:45 Moderator: **EDWARD HOLMES**

Drug Forum Panel Discussion:
JANET WOODCOCK, U.S. Food and Drug Administration
MIKHAIL GISHIZKY, Entelos, Inc.
PETER CORR, Pfizer, Inc. (retired)
THOMAS CASKEY, University of Texas HSC at Houston

EMERGING SAFETY SCIENCE

April 23–24, 2007
FDA White Oak Conference Center
Silver Spring, MD

Tuesday, April 24: Data Mining from Medical Experience

Welcome and Opening Remarks

8:30–8:45 **EDWARD HOLMES**, Workshop Chair
Co-Chair, Forum on Drug Discovery, Development, and Translation
Executive Deputy Chairman
A*Star Biomedical Research Council, Singapore

JANET WOODCOCK
Member, Forum on Drug Discovery, Development, and Translation
Deputy Commissioner and Chief Medical Officer
U.S. Food and Drug Administration

Approaches to Pharmacovigilance and Signal Detection

8:45–11:00 Moderators:
PAUL SELIGMAN
Associate Director for Safety Policy and Communication
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

GERALD DAL PAN
Director, Office of Surveillance and Epidemiology
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

30 min. Signal Management Through the Product Life Cycle

JUNE S. ALMENOFF
Vice President, Safety Evaluation and Risk Management
Global Clinical Safety and Pharmacovigilance
GlaxoSmithKline

30 min. Statistical Issues in the Analysis of Spontaneous Report Databases

WILLIAM DUMOUCHEL
Chief Statistical Scientist
Lincoln Technologies

30 min. Active Surveillance for Anticipated Adverse Events: Opportunities and Challenges

RICHARD PLATT
Professor and Chair
Harvard Medical School and Harvard Pilgrim Health Care

10:15–10:30 **Break**

30 min. Panel Discussion
JUNE ALMENOFF, GlaxoSmithKline
WILLIAM DUMOUCHEL, Lincoln Technologies
RICHARD PLATT, Harvard Medical School
ANA SZARFMAN, U.S. Food and Drug Administration
JUDITH RACOOSIN, U.S. Food and Drug Administration

Where Is the Science Taking Us?

11:00–11:30 Moderator: **EDWARD HOLMES**

30 min. Summing Up: Improving Safety Science to Make Better Medicines

RONALD KRALL
Member, Forum on Drug Discovery, Development, and Translation
Senior Vice President and Chief Medical Officer
GlaxoSmithKline

11:30–12:00 Panel Discussion
JANET WOODCOCK, U.S. Food and Drug Administration
PAUL SELIGMAN, U.S. Food and Drug Administration
RONALD KRALL, GlaxoSmithKline
JOHN JENKINS, U.S. Food and Drug Administration
MARY PRINCE PANACCIO, Merck

12:00 **Adjourn**

Appendix B

Speaker Biographies

JUNE S. ALMENOFF, MD, PHD, received her bachelor's degree with honors from Smith College. She graduated from the MD–PhD program at the Mount Sinai School of Medicine. She did her residency training in internal medicine, followed by an infectious diseases fellowship at Stanford University School of Medicine. Following this, she completed a fellowship in molecular pathogenesis at the Howard Hughes Medical Institute at Stanford. In 1993, she joined the faculty at Duke University Medical Center, where she directed a research program in molecular pharmacology. In 1997, she joined the clinical safety group at GlaxoSmithKline (GSK) (formerly GlaxoWellcome). Dr. Almenoff is currently Vice President of Safety Evaluation and Risk Management at GSK, where she manages a therapeutic portfolio. She also leads the GSK team that has developed two pioneering, award-winning systems for detecting safety issues in pharmaceutical products. These systems (Web VDME and Online Signal Management), which were developed to enhance the protection of public safety, have since been implemented at regulatory agencies such as the U.S. Food and Drug Administration and the UK Medicines Healthcare Regulatory Agency, as well as numerous pharmaceutical companies. Dr. Almenoff is a fellow of the American College of Physicians and is an associate faculty member at Duke Medical School. She has authored 40 publications and served on the editorial board of the *Drug Information Association Journal*. She currently co-chairs the collaborative PhRMA–FDA working group for safety signal detection, and was lead author of its benchmark publication on quantitative signal detection. Dr. Almenoff is also a member of the

Council for International Organizations of Medical Sciences (CIOMS) VIII Working Group on Safety Signal Detection.

ERIC BLOMME, DVM, PHD, is currently leader of the cellular and molecular toxicology group at Abbott Laboratories.

EUGENE C. BUTCHER, MD, is a professor in the Department of Pathology at Stanford University and a staff physician in the Veterans Affairs Palo Alto Health Care System. He received a BS in chemistry from the Massachusetts Institute of Technology and an MD from the Washington University School of Medicine, St. Louis, Missouri. His work has focused on the cellular and molecular mechanisms of leukocyte trafficking in immunity and inflammation, and on systems-level insights into mechanisms of cell–cell recognition and function. He has been elected to the Association of American Physicians and has been awarded the Warner Lambert/Parke Davis Award by the American Association of Pathologists, the AAI-Huang Foundation Meritorious Career Award by the American Association of Immunologists, and an Outstanding Inventor Award from Stanford University. He received the Crafoord Prize from the Swedish Academy of Sciences in 2004 for the scientific discovery of mechanisms of leukocyte trafficking contributing to the treatment of arthritis and inflammatory diseases. Dr. Butcher has been active in biotechnology, most recently cofounding and serving as chair of the Scientific Advisory Board of Bioseek, Inc. He previously helped found Leukosite, Inc., and has served on the scientific advisory boards of Millennium, Medimmune, and Thios Pharmaceuticals.

MARK I. COCKETT, PHD, joined Bristol-Myers Squibb (BMS) in January 2000, and is responsible for functional genomics and bioinformatics applied to preclinical research and development. His group manages and supports key strategic alliances with Lexicon, Artemis, Xenogen, Athersys, Pharmagene/Asterand, Exelixis, Iconix, and the Broad Institute, and is a centralized resource supporting all therapeutic areas at BMS. Before joining BMS, Dr. Cockett worked for 7 years in the neuroscience group at Wyeth, ultimately as Director, Molecular and Cell Biology, and for 10 years in the biotechnology industry for Celltech PLC, where he worked on mammalian gene expression technology and in oncology. While at Celltech, he obtained his PhD in collaboration with the Strangeway Research Laboratory, Cambridge, United Kingdom, working on the involvement of matrix metalloproteinases in tumor cell invasion. Dr. Cockett has published more than 40 peer-reviewed articles in the fields of recombinant gene expression in mammalian cells; the biochemistry and function of several matrix metalloproteinase enzymes and their role

in disease; and, more recently, heterotrimeric G protein signaling, and genomics in the pharmaceutical industry.

WILLIAM DUMOUCHEL, PHD, is currently Chief Statistical Scientist at Lincoln Technologies. From 1996 to 2004, he served as a Senior Scientist at AT&T Labs and was Professor of Biostatistics and Medical Informatics at Columbia University from 1994 to 1996. His professional interests include Bayesian statistics, data mining, pharmacovigilance, clinical data analysis, and meta-analysis. Dr. DuMouchel served as chair of the Section on Statistical Graphics for the American Statistical Association from 1996 to 1997 and as chair of the Societal Institute of the Mathematical Sciences from 1999 to 2001. He has received awards for the Best Application Paper, KDD-2001, and the Best Application Paper, KDD-2003. He was elected as a fellow of the American Statistical Association in 1981 and as a fellow of the Institute of Mathematical Statistics in 1986. Dr. DuMouchel received his PhD in statistics from Yale University in 1971.

KENDALL FRAZIER, DVM, PHD, is Director of Cellular and Molecular Pathology for Safety Assessment at GlaxoSmithKline (GSK) in King of Prussia, Pennsylvania. He received his DVM degree from Kansas State University in 1987 and a PhD in molecular biology from the University of Miami School of Medicine. He completed a residency and National Institutes of Health–funded fellowship in comparative pathology at the University of Miami Jackson Memorial Hospital in 1996 and served on the faculty of the University of Georgia College of Veterinary Medicine as assistant and associate professor of pathology prior to joining GSK. He has coauthored more than 80 peer-reviewed scientific articles and abstracts.

FELIX FRUEH, PHD, holds the position of Associate Director for Genomics in the Office of Clinical Pharmacology in the Center for Drug Evaluation and Research (CDER) at the U.S. Food and Drug Administration (FDA) and also chairs the FDA-wide Interdisciplinary Pharmacogenomics Review Group. Prior to his appointment at the FDA, Dr. Frueh was Managing Partner at Stepoutside Consulting, LLC, and served as a special government employee to the FDA and as a consultant to the Centers for Disease Control and Prevention's (CDC) National Health and Nutrition Examination Survey (NHANES) project. He held the position of Research Director for Pharmacogenetics at Transgenomic, Inc., managing the expansion of the business into new program areas for the diagnosis of genetic disorders. Previously, Dr. Frueh was Assistant Director for Biology at Protogene Laboratories, Inc., responsible for application development based on novel, in situ synthesized DNA microarray technology. He held an appointment as Assistant Professor at Georgetown University,

Washington, DC, in the Departments of Pharmacology and Medicine, and was a postdoctoral fellow at Stanford University and at the Biocenter of the University of Basel, Switzerland.

STEVEN K. GALSON, MD, MPH (*Drug Forum Member*), was named Director of the FDA's Center for Drug Evaluation and Research (CDER) in July 2005. He provides leadership for the center's broad national and international programs in pharmaceutical regulation. Dr. Galson began his public health service (PHS) career as an epidemiological investigator at the Centers for Disease Control after completing a residency in internal medicine at the Hospitals of the Medical College of Pennsylvania. He has held senior-level positions at the Environmental Protection Agency (EPA); the Department of Energy, where he was Chief Medical Officer; and the Department of Health and Human Services. Prior to his arrival at the FDA, Dr. Galson was Director of the Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances, at the EPA. He joined the FDA in April 2001 as CDER Deputy Director. Dr. Galson is the recipient of numerous PHS awards, including the Outstanding Service Medal for his leadership and management of CDER while serving as Acting Center Director from November 2001 to February 2002. He is also the recipient of three Secretary of Energy Gold Awards. Dr. Galson is a board member of the National Board of Medical Examiners and a regular peer reviewer for medical journals. He holds a BS from the State University of New York at Stony Brook, an MD from Mt. Sinai School of Medicine, and an MPH from the Harvard School of Public Health. He is board certified in preventive medicine and public health and occupational medicine.

DONALD N. HALBERT, PHD, joined Iconix Biosciences in March 2005 as Executive Vice President of Research and Development. Iconix works with a wide range of pharmaceutical clients to apply gene expression profiling as well as biomarker discovery and validation to understanding mechanisms of drug toxicity and improving preclinical drug safety. Dr. Halbert came to Iconix from Abbott Laboratories, where he was most recently Director of Genomics, Proteomics and Bioinformatics in the Global Pharmaceutical Research and Development Group. Beginning in 1991 at Abbott, Dr. Halbert was responsible for the development and integration of molecular biology and bioinformatics in the Pharmaceutical Division. In 1997 he established Genomics within the Advanced Technology Group, with responsibility for the application of human genomics, genetics, proteomics, and bioinformatics across all therapeutic areas. In 2001 he assumed additional responsibility for the Cell and Molecular Toxicology Group and was instrumental in establishing Abbott as a recognized industry leader in the application of gene expression analysis to

the understanding and prediction of compound toxicity. In his 13-year tenure at Abbott, Dr. Halbert served on and chaired numerous biotechnology scientific collaboration committees and was head of the Abbott Corporate Genomics Task Force. Prior to joining Abbott, he held scientific and management positions at Becton Dickinson in molecular biology and diagnostics, and was a cofounder of the molecular diagnostics company Gene-Trak Systems in Framingham, Massachusetts. Dr. Halbert has published extensively and holds five patents related to his work. He earned a bachelor's degree in biology from the State University of New York at Buffalo, and a doctoral degree in molecular virology from Washington University in St. Louis. He completed his postdoctoral fellowship in molecular virology with Dr. Thomas Shenk at the State University of New York at Stony Brook.

EDWARD W. HOLMES, MD (*Drug Forum Co-Chair and Workshop Chair*), was appointed a Howard Hughes Medical Investigator at Duke University School of Medicine in 1974 and later became James B. Wyngaarden Professor of Medicine. He was recruited to the University of Pennsylvania School of Medicine in 1991 as Chair of the Department of Medicine and Frank Wister Thomas Professor of Medicine and Genetics. In 1997 he became Joseph Grant Professor in the School of Medicine, Senior Associate Dean for Research, Vice President of Translational Medicine and Clinical Research, and Special Counsel to the President of the University on Biomedical Research at Stanford University. In January 1999, Dr. Holmes returned to Duke University as Dean of the School of Medicine and Walter Kempner Professor in Medicine and Genetics. He was appointed Vice Chancellor for Health Sciences and Dean of the School of Medicine at the University of California, San Diego, in fall 2000 and served in this role until October 2006. He is currently a Distinguished Professor of Medicine at the University of California, San Diego, and Vice Chancellor/Dean of Health Sciences, Emeritus, at the University of California, San Diego. Dr. Holmes became Executive Deputy Chairman of the Biomedical Research Council and Executive Chairman of the National Medical Research Council in Singapore in October 2006, and he also holds an appointment as Lien Ying Chow Professor of Medicine at the Yong Loo Lin School of Medicine, National University of Singapore. He has engaged in basic biomedical and clinical research throughout his academic career, and his laboratory work has focused on the molecular bases of human disease. Dr. Holmes has served on the Council of Advisors for the National Institute for Diabetes, Digestive, and Kidney Diseases of the National Institutes of Health, and he currently serves as Chair of the Research Advisory Board of GlaxoSmithKline. He has received Distinguished Alumnus Awards from the University of Pennsylvania and Duke University. He has been

elected to membership in the American Society for Clinical Investigation and the Association of American Physicians, is a fellow of the American Association for the Advancement of Science, and is a member of the Institute of Medicine of the National Academy of Sciences.

RONALD L. KRALL, MD (*Drug Forum Member*), is Senior Vice President and Chief Medical Officer for GlaxoSmithKline (GSK). He is responsible for all matters of human safety for all GSK compounds in development and medicinal and vaccine products, and for pharmaceutical regulatory affairs and GxP compliance. Dr. Krall joined GSK in 2003. Previously, he held positions at AstraZeneca Pharmaceuticals, Abbott Laboratories, and Lorex Pharmaceuticals. He earned a bachelor's degree in mathematics from Swarthmore College and an MD from the University of Pittsburgh, trained as a staff associate at the National Institutes of Health Epilepsy Branch, and completed his training in neurology and clinical pharmacology at the University of Rochester. He is board certified in neurology, and is a former member of the board of directors of the National Sleep Foundation, a member of the Board of Directors of the Delaware Valley Science Fairs, a member of the University of Pennsylvania Center for Bioethics Advisory Board, and a past trustee of the American Academy of Pharmaceutical Physicians.

ERIC LAI, PHD, is Vice President of Pharmacogenetics (PGx) Experimental Project Coordination and Analysis at GlaxoSmithKline (GSK). The research activities in this unit include PGx experimental design, planning and coordination; sample management and storage; genotyping; and genetics data analysis. From 2003 to 2006, Dr. Lai was Vice President, Discovery and Pipeline Genetics Division. From 2000 to 2003 he was Vice President, SNP Capability, Discovery Genetics Division, and project leader of the experimental team formed to study genetic factors associated with abacavir hypersensitivity reactions. Dr. Lai also played a major role in the planning and creation of The SNP Consortium (TSC) and was co-leader of the TSC scientific management team. He received his BSc (Hon) from the University of Waterloo and his PhD in pharmacology from Columbia University College of Physicians and Surgeons in 1989 under the guidance of Dr. Elvin Kabat. He did his postdoctoral training at the California Institute of Technology under the supervision of Dr. Leroy Hood. His postdoctoral accomplishments include the development of pulsed field gel electrophoresis and the cloning and mapping of human and mouse T-cell receptor loci. Prior to joining GlaxoWellcome in 1995, Dr. Lai was an assistant professor at the University of North Carolina at Chapel Hill. His previous works include cloning and mapping of human chromosome 2 using bacterial artificial chromosomes and extrachromo-

somal mini-chromosomes. Dr. Lai and other GSK scientists are leading the way in single nucleotide polymorphism (SNP) discovery, genotyping technology, and the use of SNP maps in the search for susceptibility genes and in pharmacogenetics.

PHILIP MACLAUGHLIN, MD, PHD, has been designing software and content for data modeling over the last 9 years. He spent 3 years with SciVision and the last 6 with Elsevier MDL, and is now with Elsevier Pharmaceutical Development Group. He serves as Principal Investigator for Elsevier MDL on an existing cooperative research and development project with the FDA's Center for Drug Evaluation and Research that involves modeling toxicity and adverse events by assembling data sets and creating quantitative structure-activity relationships (QSAR) models. The latest efforts in this area include the release of a new, widely used product on drug safety, PharmaPendium.

MICHAEL MILBURN, PHD, has served as Chief Scientific Officer for Metabolon, Inc., since 2005. Previously, he was most recently Senior Vice President of Research and Corporate Development at Sirtris Pharmaceuticals. At Sirtris, he led the preclinical/clinical development of projects in the areas of metabolic disease and neurodegeneration. Prior to his work at Sirtris, Dr. Milburn was Senior Vice President of Research at Plexxikon, where he was responsible for the development of the company's proprietary high-throughput cocrystallography drug discovery platform. He has also held positions at Structural Genomix and GlaxoSmithKline. Dr. Milburn received his PhD in structural chemistry at the University of California, Berkeley, and was a research fellow at Harvard Medical School for his postdoctoral work.

RICHARD PLATT, MD, MSC, is Professor and Chair of the Department of Ambulatory Care and Prevention, Harvard Medical School. He is an internist trained in infectious diseases and epidemiology. He is a member of the Association of American Medical Colleges Advisory Panel on Research and the Institute of Medicine's Roundtable on Evidence Based Medicine, and he currently chairs the FDA's Drug Safety and Risk Management Advisory Committee. He chaired the Executive Committee of the HMO Research Network, was co-chair of the Board of Scientific Counselors of the Centers for Disease Control and Prevention's (CDC's) Center for Infectious Diseases, chaired the National Institutes of Health's study section Epidemiology and Disease Control 2, and chaired CDC's Office of Health Care Partnerships Steering Committee. His research focuses on developing multi-institution automated record linkage systems for use in pharmacoepidemiology and in population-based surveillance, reporting, and

control of both hospital- and community-acquired infections, including bioterrorism events. He is Principal Investigator of the CDC-sponsored Center of Excellence in Public Health Informatics (www.phiconnect.org) and the Agency for Healthcare Research and Quality (AHRQ)–sponsored HMO Research Network Center for Education and Research in Therapeutics (CERT [www.certs.hhs.gov]), and co–Principal Investigator of the Modeling Infectious Disease Agent Study (MIDAS [http://www.nigms.nih.gov/Initiatives/MIDAS]) and the CDC-sponsored Eastern Massachusetts Prevention Epicenter.

Brian B. Spear, PhD, is Research Director within Abbott Laboratories' Global Pharmaceutical Research and Development Division, with responsibility for genomics, pharmacogenetics, cell and molecular toxicology, and bioinformatics. Previously, he was Director of Technology Assessment and Acquisition in the Abbott Diagnostics Division, and he has held R&D management positions in diagnostics, agricultural products, and corporate molecular biology. Dr. Spear graduated from Amherst College with honors and received his PhD from Yale University. He has held positions at the University of Colorado and Northwestern University and carried out research in chromosome structure and genome organization. Dr. Spear's recent publications have addressed applications of pharmacogenomics in drug development and patient management, and ethical and regulatory issues relating to pharmacogenetics.

Jacky Vonderscher, PhD, obtained an engineering degree in biological chemistry from the National Institute of Applied Sciences (INSA, Lyon, France). He joined the Biopharmaceutical Department of Sandoz as a research fellow investigating new drug delivery systems and in 1986 was awarded a PhD in biochemistry from the University of Geneva. Continuing at Sandoz, he and his group dealt with all biopharmaceutical aspects of various drug administration routes (oral, parenteral, dermal, nasal, and pulmonary), with techniques ranging from cell culture to healthy human subject trials. After the creation of Novartis, he became Head of the Drug Metabolism and Pharmacokinetics Department in Europe. In 2002, he was named Global Head of a new function called Integrative Compound and Product Profiling, working at the R&D interface to improve the value of the drug pipeline by making optimal use of pharmacogenomics, biomarkers, and *in silico* and *in vitro* profiling assays. In October 2002, he was nominated as Head of Drug Development for Novartis Institutes in Cambridge, Massachusetts, in addition to his global duties in preclinical safety. He is a member of the extended Development Management Board of Novartis Pharma, Ltd., and of the Discovery Board at Novartis Institutes. Dr. Vonderscher is co-author of more than 50 publications and

presentations, and co-inventor on several formulation and drug substance patents. In 1995, he shared the Golden Sandoz Triangle award for the development of Sandimmune Neoral, the new microemulsion formulation of cyclosporine. He is also a co-inventor of the recently developed immunosuppressant Myfortic.

KLAUS M. WEINBERGER, PHD, is a biomedical scientist with particular expertise in metabolomics, infectious diseases, public health, and immunology, and serves as Chief Scientific Officer and member of the management board for BIOCRATES Life Sciences. Before joining BIOCRATES in January 2003, he led a research group specializing in clinical virology, infection immunology, and molecular epidemiology at the Institute for Medical Microbiology and Hygiene at the University of Regensburg, Germany. His personal research focus was on hepatitis viruses and on the implementation of innovative technologies (e.g., quartz crystal biosensor analyses) and bioinformatics tools in biomedical research and routine diagnostics. Dr. Weinberger holds an MSc in biophysics, biochemistry, and microbiology and a PhD in medical microbiology from the University of Regensburg. He serves as a referee for scientific journals and for organizations of public science funding. He is a member of several scientific societies and was awarded a 5-year-scholarship by the Bavarian Ministry of Education (Munich, 1989–1994). He holds memberships in the World Health Organization's (WHO) Collaborating Centre for Virus Associated Cancer (Regensburg, 1994) and the WHO Reference Centre for Viral Hepatitis (Regensburg, 1996), and has received the triennial Saul Krugman Award for substantial contributions to molecular and clinical hepatitis virology (Atlanta, 2000).

JOHN K. WESTWICK, PHD, has worked in the field of cell signaling for more than 23 years, and he has been responsible for research and development at Odyssey Thera since 2002. He was previously Associate Director of Cell Signaling and Target Discovery at Celgene Corp. Prior to his work with Celgene, he was Group Leader and Project Team Leader of multiple projects at Signal Pharmaceuticals. Dr. Westwick holds a BA in biology from the University of California at San Diego and a PhD in molecular pathology from the University of California at San Diego School of Medicine. He performed postdoctoral studies in medicine and pharmacology at the University of North Carolina, Chapel Hill, School of Medicine, and was subsequently named Lineberger Cancer Center Fellow at that institution. Dr. Westwick's work is focused on signal transduction, the cellular mechanisms of cancer, inflammatory and metabolic disease, and the development of novel technologies for translating this knowledge into improved therapeutics. He has authored or co-authored more than 60 scientific articles and patents.

JANET WOODCOCK, MD (*Drug Forum Member*), is Deputy Commissioner for Operations and Chief Operating Officer, FDA. She is responsible for overseeing agency operations and cross-cutting regulatory and scientific processes. Dr. Woodcock served as Director of the FDA's Center for Drug Evaluation and Research from 1994 to 2005. She previously served in other positions at the agency, including Director, Office of Therapeutics Research and Review, and Acting Deputy Director, Center for Biologics Evaluation and Research. Dr. Woodcock received her MD from Northwestern Medical School, and completed further training and held teaching appointments at the Pennsylvania State University and the University of California at San Francisco. She joined the FDA in 1986.