
# Evolution of Translational Omics: Lessons Learned and the Path Forward

ISBN
978-0-309-22418-5

300 pages
6 x 9
PAPERBACK (2012)

Christine M. Micheel, Sharly J. Nass, and Gilbert S. Omenn, Editors; Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care Services; Board on Health Sciences Policy; Institute of Medicine

# Evolution of Translational Omics

## Lessons Learned and the Path Forward

Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials

Board on Health Care Services
Board on Health Sciences Policy

Christine M. Micheel, Sharyl J. Nass, and Gilbert S. Omenn, *Editors*

INSTITUTE OF MEDICINE
*OF THE NATIONAL ACADEMIES*

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

***PREPUBLICATION COPY: UNCORRECTED PROOFS***

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

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

This study was supported by Contract Nos. HHSN261200900003C (National Cancer Institute); HHSF223201010692P and HHSF22301018T (Food and Drug Administration); and 200-2011-38807 and 200-2005-13434 (Centers for Disease Control and Prevention). This study was also supported by the U.S. Department of Veterans Affairs, the American Society for Clinical Pathology, and the College of American Pathologists. 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 0-309-XXXXX-X (Book)
International Standard Book Number 0-309- XXXXX -X (PDF)

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; 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). 2012. *Evolution of Translational Omics: Lessons Learned and the Path Forward.* Washington, DC: The National Academies Press.

***PREPUBLICATION COPY: UNCORRECTED PROOFS***

*"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.**

***PREPUBLICATION COPY: UNCORRECTED PROOFS***

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**

***PREPUBLICATION COPY: UNCORRECTED PROOFS***

## COMMITTEE ON THE REVIEW OF OMICS-BASED TESTS FOR PREDICTING PATIENT OUTCOMES IN CLINICAL TRIALS

**GILBERT S. OMENN** (*Chair*), Professor of Internal Medicine, Human Genetics and Public Health; Director, University of Michigan Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor

**CATHERINE D. DEANGELIS,** Professor of Pediatrics, Johns Hopkins School of Medicine, Professor of Health Policy and Management, Johns Hopkins School of Public Health, and Editor-in-Chief Emerita, *Journal of the American Medical Association,* Baltimore, MD

**DAVID L. DEMETS,** Professor of Statistics and Biostatistics, University of Wisconsin, Madison

**THOMAS R. FLEMING,** Professor of Biostatistics, Statistics, University of Washington, Seattle

**GAIL GELLER,** Professor of Medicine, Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD

**JOE GRAY,** Gordon Moore Endowed Chair, Department of Biomedical Engineering, Center for Spatial Systems Biomedicine, Oregon Health & Science University Knight Cancer Institute, Portland

**DANIEL F. HAYES,** Clinical Director of the Breast Oncology Program, and Stuart B. Padnos Professor of Breast Cancer Research, University of Michigan Comprehensive Cancer Center, Ann Arbor

**I. CRAIG HENDERSON,** Adjunct Professor of Medicine, University of California, San Francisco and Helen Diller Family Comprehensive Cancer Center

**LARRY KESSLER,** Professor and Chair, Department of Health Services, University of Washington School of Public Health, Seattle

**STANLEY LAPIDUS,** Founder, President, and CEO, SynapDx Corporation, Southborough, MA

**DEBRA LEONARD,** Professor and Vice Chair of Laboratory Medicine, and Director of the Clinical Laboratories, Weill Medical College of Cornell University, New York, MY

**HAROLD L. MOSES,** Director Emeritus, Hortense B. Ingram Professor of Molecular Oncology, and Professor of Cancer Biology, Medicine, and Pathology, Vanderbilt-Ingram Cancer Center, Nashville, TN

**WILLIAM PAO,** Associate Professor of Medicine, Cancer Biology, and Pathology, Microbiology, Immunology, Ingram Associate Professor of Cancer Research, Director of the Division of Hematology/Oncology, and Director of Personalized Cancer Medicine, Vanderbilt University School of Medicine, Nashville, TN

**REBECCA D. PENTZ,** Professor of Hematology and Oncology in Research Ethics, Emory School of Medicine, Atlanta, GA

**NATHAN D. PRICE,** Associate Professor, Institute for Systems Biology, Seattle, WA

**JOHN QUACKENBUSH,** Professor of Computational Biology and Bioinformatics, Dana-Farber Cancer Institute, Boston, MA

**ELDA RAILEY,** Cofounder, Research Advocacy Network, Plano, TX

**DAVID RANSOHOFF,** Professor of Medicine, Clinical Professor of Epidemiology, University of North Carolina, School of Medicine and Gillings School of Global Public Health, Chapel Hill

**E. ALBERT REECE,** Vice President for Medical Affairs, University of Maryland, and John Z. and Akiko K. Bowers Distinguished Professor and Dean, University of Maryland School of Medicine, Baltimore

**DANIELA M. WITTEN,** Assistant Professor of Biostatistics, University of Washington, Seattle

*Staff*

**CHRISTINE M. MICHEEL,** Study Director (through November 2011)

**SHARYL NASS,** Director, National Cancer Policy Forum; Study Director (from December 2011)

**LAURA LEVIT,** Program Officer (from June 2011)

**ERIN BALOGH,** Associate Program Officer

**SARAH DOMNITZ,** Christine Mirzayan Science and Technology Policy Graduate Fellow (from August to December 2011)
**JULIA E. DOOHER,** Christine Mirzayan Science and Technology Policy Graduate Fellow (from January to May 2011)
**NIHARIKA SATHE,** Research Assistant (from June 2011)
**MICHAEL PARK,** Senior Program Assistant
**PATRICK BURKE,** Financial Associate
**ROGER HERDMAN,** Director, Board on Health Care Services
**ANDREW POPE,** Director, Board on Health Sciences Policy

*Consultant*
**JOHN BAILAR,** Scholar in Residence

## 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:

**WYLIE BURKE,** Professor and Chair, Department of Bioethics and Humanities, University of Washington
**ADAM M. CLARKE,** Founder, MedTran Health Strategies, LLC
**SUSAN S. ELLENBERG,** Professor of Biostatistics, Associate Dean for Clinical Research, Perelman School of Medicine, University of Pennsylvania
**CHARIS ENG,** ACS Clinical Research Professor, Chair, and Director, Genomic Medicine Institute, Cleveland Clinic Foundation, Case Western Reserve University School of Medicine
**MARCUS FELDMAN**, Professor of Biology, Stanford University
**DAVID B. FLANNERY**, Chief of Medical Genetics, Professor of Pediatrics, Vice Chair for Administration of the Department of Pediatrics, Georgia Health Sciences University
**STEPHEN FRIEND,** President and CEO, Sage Bionetworks
**LARRY GOLD,** Chairman and Chief Executive Officer, SomaLogic, Inc.
**STEVEN GOODMAN**, Associate Dean for Clinical and Translational Research, Professor of Medicine & Health Research and Policy, Stanford University School of Medicine
**ROBERT GRAY**, Professor of Biostatistics, Dana Farber Cancer Institute
**STEPHEN GRUBBS,** Principal Investigator, Delaware Christiana Care Community Clinical Oncology Program
**DAVID KORN**, Vice Provost for Research, Professor of Pathology, Harvard University
**MARC LADANYI,** William J. Ruane Chair in Molecular Oncology, Memorial Sloan-Kettering Cancer Center
**BERNARD LO,** Professor of Medicine and Director, Program in Medical Ethics, University of California, San Francisco
**DAVID MADIGAN,** Professor of Statistics, Columbia University
**BETTIE SUE SILER MASTERS,** The Robert A. Welch Distinguished Professor in Chemistry, University of Texas Health Science Center at San Antonio
**CHARLES E. PHELPS,** Provost Emeritus, Professor of Political Science and of Economics, University of Rochester
**DAN RODEN,** Professor of Medicine, William Stokes Chair in Experimental Therapeutics, Professor of Pharmacology, Vanderbilt University Medical Center
**LARRY SHAPIRO,** Executive Vice Chancellor for Medical Affairs and Dean, Washington University School of Medicine
**PETER SHIELDS,** Professor and Chief, Deputy Director, Comprehensive Cancer Center and Professor, College of Medicine, The Ohio State University Medical Center
**STEVE TEUTSCH,** Chief Science Officer, Los Angeles County Public Health
**DAVID WONG,** Professor of Bioengineering, Director of the Dental Research Institute, University of California, Los Angeles

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by **KRISTINE GEBBIE,** Adjunct Professor, Flinders University School of Nursing and Midwifery and **LAWRENCE D. BROWN,** Professor, Department of Statistics, The Wharton School, University of Pennsylvania. Appointed by the Institute of Medicine and the National Research Council, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

# Acknowledgments

The committee is grateful to many individuals who provided valuable input and information for the study, either through formal presentations or through informal communications with study staff and committee members:

Patrick Anquetil, SynapDx
Keith Baggerly, MD Anderson Cancer Center
John Bailar, Institute of Medicine
William Barry, Duke University
Robert Becker, The Food and Drug Administration
Robert Califf, Duke University
Ned Calonge, The Colorado Trust
Michael Cuffe, Duke University
Jeffrey Drazen, Editor in Chief, *New England Journal of Medicine*
John Falletta, Duke University
Geoffrey Ginsburg, Duke University
Alberto Gutierrez, The Food and Drug Administration
Scott Henderson, Vermillion
Michael Kelley, Duke University
Katrina Kelner, *Translational Medicine* and *Science* Magazine
Véronique Kiermer, Nature Publishing
Sally Kornbluth, Duke University
Sumithra Mandrekar, Mayo Clinic
Anne-Marie Mazza, The National Academies
Ross McKinney, Duke University
Lisa McShane, National Cancer Institute
Stephen A. Merrill, The National Academies
Mitch Nelles, XDx
Joseph Nevins, Duke University
Nicole Osmer, Nicole Osmer Healthcare Communications
Harold Paz, Pennsylvania State University
Charles Perou, The University of North Carolina at Chapel Hill
Peter Pronovost, Johns Hopkins University
William Sellers, Novartis Institutes for BioMedical Research
Steven Shak, Genomic Health
Richard Simon, National Cancer Institute
Ed Stevens, Pathwork Diagnostics
Susan Su, United States Patent and Trademark Office
Paul Uhlir, The National Academies
Laura van ‘t Veer, Helen Diller Family Comprehensive Cancer Center
Pablo Whaley, United States Patent and Trademark Office
Scott Zeger, Johns Hopkins University

We thank the National Cancer Institute, the Food and Drug Administration, the U.S. Department of Veterans Affairs, the Centers for Disease Control and Prevention, the American Society for Clinical Pathology and the College of American Pathologists for supporting this study.

# Contents

# Boxes, Figures and Tables

*Boxes*



*Figures*



*Tables*

# Summary

"Omics" is a term encompassing multiple molecular disciplines, which involve the characterization of global sets of biological molecules such as DNAs, RNAs, proteins, and metabolites. For example, genomics investigates thousands of DNA sequences, transcriptomics investigates all or many gene transcripts, proteomics investigates large numbers of proteins, and metabolomics investigates large sets of metabolites.

Omics research generates complex high-dimensional data; these data are often generated through measurement of many more variables per sample than the total number of biological samples used to generate the dataset. These data can be used to produce a computational model[1] that potentially distinguishes a health-related characteristic of clinical significance and is intended for eventual analysis of individual patient specimens in a clinical setting. High-dimensional data are particularly prone to overfitting; as a result, a computational model emerging from the research and discovery phase may function well on the samples used for the discovery research, but is inaccurate on any other sample. A carefully designed and executed series of studies is necessary to develop a clinically useful omics-based test for patient management and care, with the goal of improving patient outcomes.

Several characteristics distinguish omics-based tests from other medical technologies, including a different regulatory oversight process, the difficulty in defining the biological rationale behind a test based on multiple individual biomarkers, the complexity of data sharing with other scientists, and the high degree of hope placed in the promise of omics-enabled technologies and medical care.

Omics-based tests, and indeed all clinical laboratory tests, are subject to a different regulatory framework than drugs. Specifically, there are more pathways for regulation of in vitro diagnostic test devices—the category under which omics-based tests fall—than there are for drugs. Tests can be developed, validated, and placed into clinical use either through review by the Food and Drug Administration (FDA) or through validation and performance in a specific laboratory, also called laboratory developed tests (LDTs). Any clinical laboratory that reports tests for clinical management of patients falls under the purview of the Clinical Laboratory Improvement Act (a CLIA-certified clinical laboratory) that provides a baseline level of oversight with respect to test development and the quality of laboratory operations. While the Food and Drug FDA has the authority for regulatory oversight of all tests used in patient care, the FDA has not defined a regulatory framework that includes oversight of LDTs and has only reviewed LDT tests determined to be of high complexity and therefore high risk to patients. This alternate LDT pathway is not possible for drug development, and all drugs must be approved by the FDA. It is precisely this LDT pathway that allows academic medical centers to move omics-based tests from discovery to clinical use without external regulatory review of the new test, and places a new and mostly unrecognized demand on academic institutions to provide proper

[1]The series of computational steps performed in processing the raw data, as well as the mathematical formula or formulas used to convert the data into a prediction of the phenotype of interest, all precisely defined in written form.

oversight for omics-based test development, validation, and clinical implementation. While pharmaceutical and medical device companies follow well-established medical product development pathways and have many process controls in place for strong oversight of development, clinical validation, and manufacturing, academic institutions are not as accustomed to overseeing the development of medical products.

The frequent lack of a clear biological rationale further distinguishes omics-based tests from most other clinical laboratory tests based on a single analyte. The biological rationale behind a single-analyte test is frequently quite evident: The test is useful because the gene, RNA, protein, or metabolite plays an understood role in the disease pathology or other biological process under investigation. Examples of single-analyte tests include human epidermal growth factor receptor 2 (HER2) testing of breast cancers or measuring low-density lipoprotein (LDL) cholesterol level for cardiac risk assessment. In contrast, the biological rationale for the set of biomarkers in an omics-based test frequently is not well defined scientifically. This difference puts an additional burden on the statisticians and bioinformatics experts involved in test validation to ensure that the biological data and computational model are scientifically sound. Due to the increased risk of overfitting large data sets in the development of the computational model, the need for rigor, validation, and accountability is even higher than for other single biomarker-based tests.

The complexity of omics research also makes data provenance more challenging and makes sharing of the complex data sets and computational models difficult, which limits the ability of other scientists to replicate and verify the findings and conclusions of omics research studies. Database repositories for genomic data sets are available, but data sharing is not routine, and without access to the data sets or a precisely defined computational model, replication and verification are more difficult than for single biomarker tests. While independent confirmation studies are expensive, the need for replication is beneficial in the omics field given the data complexities that can lead to errors, from simple data management errors to incorrectly designed computational models. This level of complexity does not exist for single-biomarker test research, development, and validation.

Despite the nearly complete identification of the human genome sequence in 2001, development of omics-based products that influence or improve patient health has been slower than expected. One possible reason for this limited progress is that there has not been a widely agreed-upon process for translation of omics discoveries into clinical omics-based tests intended to improve patient outcomes and care. Many hope that the promise that omics science holds for medicine and public health will be realized. With the creation of high-throughput measurement technologies, it is now feasible to take a snapshot of a patient's molecular profile at specific stages in the progression of disease pathology or at a given location in the body. However, the complexity of these technologies and of the resulting high-dimensional data introduces major challenges for the scientific community, as rigorous statistical, bioinformatics, laboratory, and clinical procedures are required to develop and validate these tests and evaluate their clinical usefulness.

The failure of scientific collaboration, review processes by journals, regulatory oversight, institutional systems for protection of patient-participants, and institutional systems for management of conflicts of interest in a recent case involving the premature use of gene expression-based tests in clinical trials at Duke University led the National Cancer Institute (NCI) to request establishment of this Institute of Medicine (IOM) committee. The committee's charge was to develop recommendations to clarify and improve the pathway from discovery to

first use of omics-based tests in a clinical trial, to assess the potential for new omics-based tests to benefit patients.

## STUDY SCOPE

Recent events have highlighted the lack of clarity about best practices for omics-based test validations and the failure of current oversight systems. In July 2010, NCI Director Harold Varmus received a letter from more than 30 statisticians and bioinformatics scientists expressing concerns over several genomics-based predictive tests already in use in clinical trials at Duke University to predict the type of chemotherapy that individual cancer patients were most likely to benefit from. As a result, an IOM committee was convened to help clarify questions about how to effectively develop omics-based tests to enable progress toward improving patient outcomes. The IOM study was focused on making recommendations useful to investigators in the broader field of omics-based test development, rather than simply examining what went wrong in the test development process at Duke University. With support from NCI, FDA, the Centers for Disease Control and Prevention, the U.S. Department of Veterans Affairs, the American Society for Clinical Pathology, and the College of American Pathologists, the IOM committee's charge was to recommend sound principles for appropriate development and evaluation for translating omics-based tests from the research laboratory into clinical trials, with the ultimate goal of guiding therapeutic decisions to improve patient outcomes. The complete charge to the committee can be found in Chapter 1.

## FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The committee considered its task in the context of the scientific processes of discovery, confirmation, validation, and evaluation for clinical use of candidate omics-based tests and in relation to the many parties responsible for the discovery and development of omics-based tests. The primary investigators, who often work in interdisciplinary teams, bear the greatest responsibility and accountability for the scientific rigor of the discovery research and test development. Academic institutions, other non-profit research organizations, and for-profit companies that support the development of omics-based tests also bear responsibility for proper oversight of the discovery, translational, and clinical research conducted and reported by their faculty or research staff seeking to generate successful omics-based tests. Although these institutions depend on the rigor and integrity with which individual investigators perform and defend their work, they also have a significant role to play in providing necessary infrastructure, supporting scientific integrity, and organizing and conducting investigations of allegations of improper or incorrect research and reporting practices.

The evaluation process recommended in this report defines the best practices for translation of an omics-based discovery into a validated omics-based test for use in a clinical trial, and focuses on the responsibilities of the investigators (Recommendations 1-3; Box S-1), with additional recommendations for other responsible parties, particularly institutions, but also funding agencies, journals, and the FDA (Recommendations 4-7; Box S-2). Throughout its recommendations, the committee emphasized the importance of transparency in reporting—making data, metadata (information about a data set and how it was generated), prespecified analysis plans, computer code, and fully specified computational models available for external evaluation or confirmation. This reinforces recommendations made in several National Research Council reports (NRC, 2003, 2005, 2006).

## Development and Evaluation Process

The committee's recommended development and evaluation process for omics-based tests is summarized in Figure S-1. The two major stages of test development and evaluation entail (1) discovery and test validation phases, and (2) evaluation of clinical utility and use. The discovery phase includes complete definition of the computational model to be used for data analysis in a clinical test and independent confirmation of that model. At this point, the fully specified computational procedures should be locked down—recorded and no longer changed. The candidate omics-based test from the research laboratory is then transferred to a CLIA-certified clinical laboratory for development of the clinical testing methods followed by analytical validation and clinical/biological validation. The final stage is assessment of the clinical utility and use of the validated omics-based test within a clinical trial, with multiple design options depending on the intended clinical use of the test and availability of specimens from previous clinical trials. Statistics and bioinformatics validation occurs throughout both development stages. Overfitting of statistical models derived from omics data is common and many published gene expression results have been difficult to replicate.

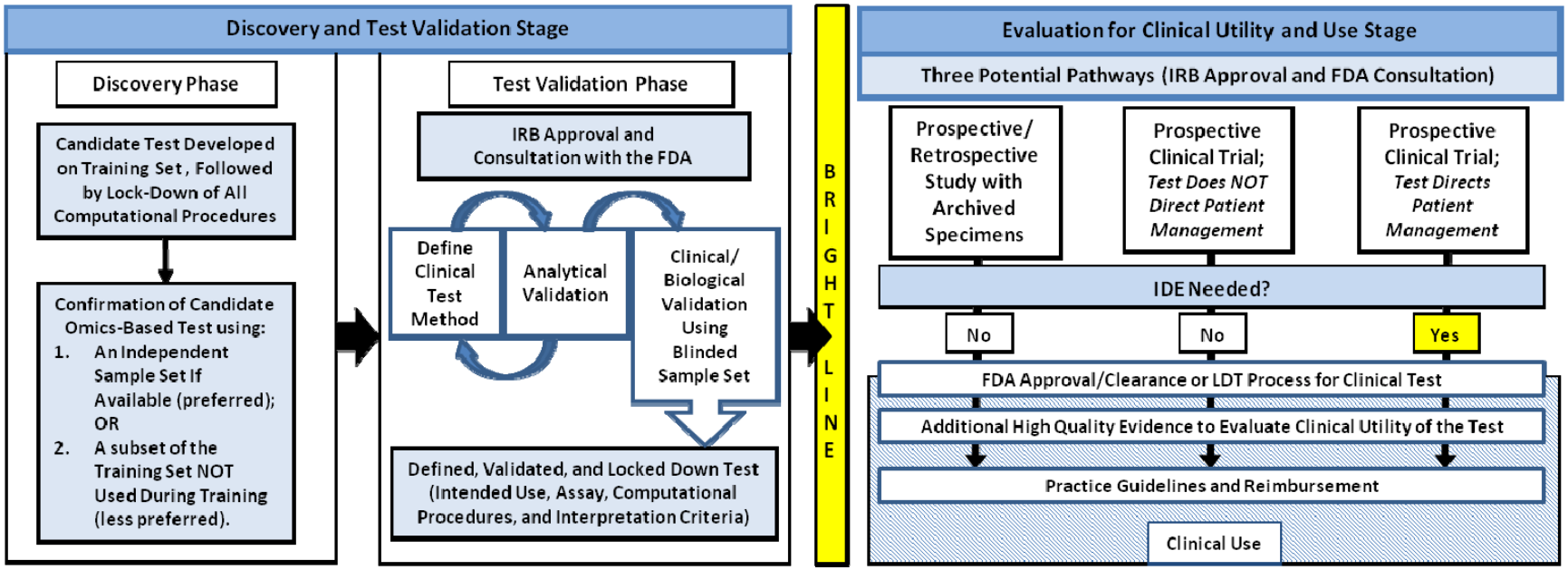


**FIGURE S-1** Omics-based test development process. In the first stage of omics-based test development, there are two phases: discovery and test validation. In the discovery phase, a candidate test is developed and confirmed. The fully specified computational procedures are locked down in the discovery phase and should remain unchanged in all subsequent development steps. Ideally, confirmation should take place on an independent sample set. Under exceptional circumstances it may be necessary to move into the test validation phase without first confirming the candidate test on an independent sample set if using an independent test set in the discovery phase is not possible, but this increases the risk of test failure in the validation phase. In the test validation phase, the omics-based test undergoes analytical and clinical/biological validation. The bright line signifies the point in test development where a fully defined, validated, and locked down clinical test (analytical and clinical/biological validation) is necessary. Changes to the test after the bright line is crossed require a return to the test validation phase, approval by the Institutional Review Board, and possibly consultation with the Food and Drug Administration. In the second stage of test development, the fully defined, validated, and locked down omics-based test undergoes evaluation for its intended clinical use. Evaluation of clinical utility and use is a process that often continues after initial adoption into clinical use. Statistics and bioinformatics validation occurs throughout the discovery and test validation stage as well as the stage of evaluation for clinical utility and use.
NOTE: FDA = Food and Drug Administration, IDE = investigational device exemption, IRB = Institutional Review Board, LDT = laboratory-developed test.

**BOX S-1**
**Development and Evaluation Process Recommendations**

**Discovery and Test Validation Stage: Discovery Phase**

1. **When candidate omics-based tests from the discovery phase are intended for further clinical development, the following criteria should be satisfied and fully disclosed (for example, through publication or patent application) to enable independent verification of the findings:**
   a. Candidate omics-based tests should be confirmed using an independent set of samples, not used in generation of the computational model and, when feasible, blinded to any outcome or other phenotypic data until after the computational procedures have been locked down and the candidate omics-based test has been applied to the samples;
   b. The data and metadata used for development of the candidate omics-based test should be made available in an independently managed database (e.g., the databases of Genotypes and Phenotypes [dbGaP]) in standard format;
   c. The computer code and fully specified computational procedures used for development of the candidate omics-based test should be made sustainably available[2]; and
   d. The candidate omics-based test should be defined precisely, including the molecular measurements, the computational procedures, and the intended clinical use of the test, in anticipation of the test validation phase.

**Discovery and Test Validation Stage: Test Validation Phase**

2. An omics-based test consists of both the data-generating assay and the fully specified computational procedures used for analysis of the assay data. The committee recommends that both components of omics-based tests used to direct patient management in a clinical trial setting should be validated during the test validation phase using the following steps:
   a. The candidate omics-based test and its intended use should be discussed with the Food and Drug Administration (FDA) prior to initiation of validation studies.
   b. Test validation should be performed in a CLIA-certified clinical laboratory, beginning with a defined candidate omics-based test from the discovery phase.
   c. The CLIA-certified laboratory should design, optimize, validate, and implement the omics-based test under current clinical laboratory standards.
   d. If the omics-based test will be performed in more than one CLIA-certified laboratory for a clinical trial, analytical validation and CLIA requirements for the same omics-based test should be met by each laboratory, working with the primary laboratory.

**Evaluation for Clinical Utility and Use Stage**

3. For investigators conducting a clinical trial to assess the clinical utility and use of an omics-based test that has been confirmed and validated as described above (Recommendations 1-2), the committee recommends that:
   a. Investigators should communicate early with the FDA regarding the investigational device exemption (IDE) process and validation requirements.

[2]For publicly funded research, code and fully specified computational procedures should be made publicly available either at the time of publication or at the end of funding. For commercially developed tests, this information would be submitted for FDA review if seeking approval or clearance, or would be described in a publication in the case of a Laboratory Developed Test.

b. Omics-based tests should not be changed during the clinical trial without a protocol amendment and discussion with the FDA. A substantive change to the omics-based test may require that the study be restarted.

The committee was charged with the task of recommending an evaluation process for determining when predictive tests utilizing omics-based technologies are fit for use in clinical trials, especially those in which the assay is used to determine patient management. Review of published work and several case studies of tests—including some that have been adopted into clinical use and some that did not achieve that goal—helped the committee outline and recommend a process for developing and evaluating omics-based tests. The committee's recommendations aim to reinforce, clarify, and build on the current regulations and processes that apply to medical devices such as in vitro diagnostic multivariate index assays and laboratory-developed tests. The quality and quantity of data and information available for making this determination depends in part on the test discovery process. Because omics-based tests rely on interpretation of high-dimensional datasets, methods that avoid overfitting the data in development of the computational model should be used throughout the test development process. Overfitting due to use of improper statistical methods leads to a computational model that fits the training samples well, but will perform poorly on independent samples not used in the discovery phase. Confirmation of all computational models and candidate omics-based tests on an independent sample set appropriate for the intended used of the test, blinded to assay results and clinical information linked to the specimens, is the "gold standard" method to assess test validity and to avoid taking an overfit computational model into subsequent development steps. In addition, the complexities of the data management and analyses of the large omics datasets highlight the need for availability of the data and computer code used in the discovery phase of omics-based test development for external evaluation and confirmation. The result of the discovery process, outlined in Figure S-1 and Recommendation 1 (see Box S-1) and fully described in Chapter 2, is a candidate omics-based test with fully specified and locked down computational procedures.

The candidate omics-based test is then transitioned into the test validation phase to assess analytical and clinical/biological validation. There is a wealth of existing work on best practices for the development of clinical laboratory tests, and much of this can be applied to omics-based tests. A candidate omics-based test method should be optimized for performance in a clinical laboratory, and then should undergo analytical validation and clinical/biological validation. The clinical test methods can be optimized based on feedback from the analytical validation performance characteristics, but must be fully defined, completely validated analytically and biologically/clinically with acceptable performance of the test, and locked down prior to assessment of the clinical utility and use of the test in a clinical trial. The optimal process for validating a candidate omics-based test is the same process used for validating any clinical test to be performed in clinical laboratories. This process is outlined in Figure S-1 and Recommendation 2 (see Box S-1) and fully described in Chapter 3.

Once the analytical and clinical/biological performance of the defined omics-based test is acceptable based on intended clinical use, the test is ready for a clinical study to assess clinical utility and use. Ideally, determination of clinical utility should be derived from a prospective randomized clinical trial, but in some circumstances that may not be feasible. Figure S-1 shows several possible pathways to evaluate clinical utility and use. Depending on the choice of study design, investigators may need to obtain an investigational device exemption (IDE) from the

FDA. Regardless of which pathway is chosen, however, the committee strongly recommends consulting with the FDA prior to initiation of a clinical trial. In the case of a trial where patient management will be influenced by the omics-based test findings, this is a legal requirement. In other cases, the committee recommends consultation with the FDA because the requirement for an IDE based on the trial design is not always clear. In addition, if the test will later need clearance or approval from the FDA before clinical use, the study design and analysis will be subject to FDA review. A pre-IDE consultation can assist both the test developer and the FDA in agreeing on the appropriate pathway for FDA review and the data necessary for FDA approval or clearance. Critical considerations for moving a fully defined omics-based test into clinical trials for assessing clinical utility and use are outlined in Figure S-1 and Recommendation 3 (see Box S-1) and fully described in Chapter 4.

**Ensuring Adoption and Adherence to the Development and Evaluation Process**

The committee's recommendations to ensure adoption of appropriate research and development practices (Recommendations 4-7) are put forth in the spirit of "best practices." There could be numerous approaches to ensuring adherence, and the optimal approach will need to be determined by each of the stakeholders.

Investigators and institutions that conduct omics-based research with the goal of improving patient care have responsibilities for supporting that research and test development. Both contribute to the scientific research culture in which omics-based research is conducted; investigators control the culture of individual laboratories, while institutions put policies and procedures in place that support scientific integrity and ensure sound and ethical practices for clinical research. To avoid adding new barriers to innovation in this promising field, the committee's recommendations aim to emphasize and enhance institutional awareness of existing responsibilities to ensure the integrity of the scientific process. The committee recognized that these recommendations might increase the oversight requirements for omics research in some institutions, but agreed that these potential costs were offset by the added safeguards for the integrity of this research. If an institution does not have the infrastructure or capability to follow the recommended Test Development and Evaluation Process defined in this report, then the committee believes that institution should consider not engaging in the translation of omics-based discoveries into validated tests intended for clinical use.

As the Duke case study clearly demonstrates, existing procedures in some institutions may not adequately ensure the scientific integrity of translational omics. For example, although most institutions have clear policies and procedures for financial conflicts of interest for individuals, there is often less clarity when handling institutional conflicts, both financial and non-financial. An institution might appear so conflicted in certain situations that an outside body should be asked to take responsibility for an investigation.

The committee also addressed responsibilities of funders, the FDA, and journals in ensuring rigorous development of omics-based tests (Box S-2). Funders play a leadership role in encouraging a culture of integrity and transparency in science, while they seek to accelerate progress through discovery, translation, and clinical applications. The committee highlighted the importance of funders supporting independent confirmation as well as validation in a CLIA-certified clinical laboratory of candidate omics-based tests because funders have generally not supported such work, as they do not consider it to be original, innovative science. Without this support, confirmation and validation studies will be difficult, and the field will be left with promising ideas published in journals that may be used prematurely in clinical trials. The

responsibilities of journal editors with respect to the adoption and adherence to the recommended omics-based test development and evaluation process are complicated by the wide spectrum of adopted policies and resources available to individual journals, but the committee recommends several actions—all specified in Box S-2—to improve the transparency and reproducibility of published research. Finally, there are several steps that the FDA should take to improve understanding of the regulatory requirements for omics-based tests, by directly communicating with investigators and academic institutions and by developing a guidance or regulation that clarifies the relevant requirements in this dynamic field.

**BOX S-2**
**Recommendations on Appropriate Actions to Ensure Adoption and Adherence**

**Institutions**

4.a. Institutions are responsible for establishing, supporting, and overseeing the infrastructure and research processes for omics-based test development and evaluation as well as best practices for clinical trials and observational research, including those incorporating omics technologies, and should assure that the evaluation process outlined in this report is followed for omics-based test development and evaluation at their institution.

4.b. Given the complexity of omics research and omics-based tests, the multidisciplinary nature of omics research, and the potential for conflicts of interest in developing and evaluating tests for clinical use, institutional leaders should pay heightened attention to providing appropriate oversight and promoting a culture of scientific integrity and transparency. They should designate:

i. A specific Institutional Review Board (IRB) member(s) to be responsible for considering investigational device exemption (IDE) and investigational new drug (IND) requirements as a component of ensuring the proper conduct of omics-based clinical research
ii. An institutional official who is responsible for comprehensive and timely documentation, disclosure, and management of financial and non-financial conflicts of interest, both individual and institutional
iii. An institutional official who is responsible for establishing and managing a safe system for preventing, reporting, and adjudicating lapses in scientific integrity, to enhance patient safety
iv. An institutional official who is responsible for establishing clear procedures for response to inquiries and/or serious criticism about the science being conducted at the institution. For example, this individual would be the responsible official for journals to contact with a serious concern about a manuscript, ensure that relevant information is provided to external scientists to help resolve issues of transparency of methods and data, and inform funders when an investigation of potential scientific misconduct is initiated

4.c. Institutions that conduct biomedical omics research, including test development and clinical trials, should train, recognize, and support the faculty-level careers of individuals from the multiple collaborating disciplines, including biostatistics, bioinformatics, pathology, omics technologies, and clinical trials, and ensure that they are:

i. Treated as equal co-investigators and co-owners of responsibility

ii. Represented on all relevant review and oversight bodies within the institutions
iii. Intellectually independent, preferably reporting to an independent mentor and/or department chair as well as to the project leader

**Funders**

5.a. All funders of omics-based translational research should:

i. Require investigators to make all data, metadata, prespecified analysis plans, code, and fully specified computational procedures publicly available and readily interpretable either at the time of publication or, if not published, at the end of funding, and funders should financially support this requirement;
ii. Provide continuing support for independent repositories to guarantee ongoing access to relevant omics and clinical data;
iii. Support test validation in a CLIA-certified clinical laboratory and consider the usefulness of an independent confirmation of a candidate omics-based tests prior to evaluation for clinical use;
iv. Designate an official to alert the institutional leadership when serious allegations or questions have been raised that may warrant an institutional investigation; if the funder (e.g., the National Institutes of Health) has initiated that question, then the funder and institution should communicate during the investigation; and
v. Establish lines of communication with other funders to be used when serious problems appear to involve interdependent research sponsored by another funder along the omics-based test development process.

5.b. Federal funders of omics-based translational research should have authority to exercise the option of investigating any research being conducted by a funding recipient after requesting an investigation by the institution.

**FDA**

6.a. In order to enable investigators and institutions to have a clear understanding of their regulatory responsibilities, the FDA should develop and finalize a risk-based guidance or a regulation on:

i. Bringing omics-based tests to the FDA for review
ii. Oversight of laboratory-developed tests (LDTs)

6.b. The Food and Drug Administration (FDA) should communicate the IDE requirements for use of omics-based tests in clinical trials to the Office of Human Research Protections (OHRP), IRBs, and other relevant institutional leadership.

**Journals**

7. Journal editors should:

a. Require authors submitting manuscripts describing clinical evaluations of omics-based tests to:

i. Register all clinical trials at clinicaltrials.gov or another trial registry acceptable to the journal
ii. Make data, metadata, prespecified analysis plans, code, and fully specified computational procedures publicly available in an independently managed database (e.g., the databases of Genotypes and Phenotypes [dbGAP]) in standard format
iii. Provide the journal with the sections of the research protocol relevant to their manuscript
iv. Identify each author's role in the development, conduct, analysis, writing, and editing of the manuscript. Require the lead and senior authors to attest to the integrity of the study and the co-authors to confirm shared responsibility for study integrity
v. Use appropriate guidelines (e.g., the Consolidated Standards of Reporting Trials [CONSORT] and the Reporting recommendations for tumor MARKer prognostic studies [REMARK]) and submit checklists to certify guideline use

b. Develop mechanisms to resolve possible serious errors in published data, metadata, code, and/or computational models and establish clear procedures for management of error reports.
c. Alert the institutional leadership and all authors when a serious question of accuracy or integrity has been raised.

## Case Studies

The committee examined several case studies of tests whose development histories provide lessons learned and illustrate the committee's recommendations. These include the series of genomics-based predictive tests used in clinical trials at Duke University; the commercial tests OncotypeDx, MammaPrint, Ova1, AlloMap Testing, CorusCAD, and the Tissue of Origin Test; and the first OvaCheck test, which did not reach clinical use due to errors discovered in the methods used to develop the test. HER2 testing also is included as a case study to illustrate the challenges associated with a single-biomarker test, which could be magnified in omics-based tests. The committee was charged with presenting findings related to the genomics tests used in the three Duke University clinical trials named in the statement of task. Published papers describing the development of those tests have been retracted, and, thus, it is now widely accepted that the clinical trials should not have used the omics-based tests for patient management decisions.

The events at Duke University captured the attention of biological and quantitative scientists around the world. The committee gathered information about the series of events leading to the inappropriate use of the genomics-based predictive tests for patient management decisions in clinical trials at Duke University. Unfortunately, multiple systems put in place by Duke University to ensure the integrity and rigor of the scientific process failed. However, Duke University is not unique. Many of these failures stemmed from problems that may exist at other institutions: unclear lines of accountability, lack of consistently strong data management, lack of confirmation of the omics discovery using an independent sample set, lack of definition or locking down of the specific assay and computational analysis methods, lack of analytical and clinical/biological validation of the omics-based test prior to commencing clinical trials, and

individual and institutional conflicts of interest, both financial and non-financial. As a result, public trust in the scientific and medical systems and patient-participant safety have been put at risk.

During the 10 years since the research leading to the erroneous predictive tests was initiated, omics science and the regulation of omics-based tests have evolved. Institutions are better equipped now to answer investigator questions about appropriate development processes. Nonetheless, the committee identified needs for improvement. The committee believes the problems at Duke University could have been prevented had its recommendations been available and followed. Furthermore, the committee believes that scientific progress in omics test development will improve if these recommendations are broadly adopted because they ensure wide availability of data and computational models for the scientific community to explore, clarify the regulatory steps that must be followed along the process, and clarify responsibilities for the parties involved in this process.

The committee hopes this report will provide a guide to the entire pathway for the development of omics-based tests, from discovery to clinical trials, to assist the many parties contributing to this translational research in understanding the complete pathway and not just their focused contributions. This broader perspective may help the whole investigative team to understand the entire pathway and the pitfalls of each stage, with the hope of avoiding future problems in translating omics-based discoveries into clinical tests for the benefit of improved patient care.

Envisioning the improvement of omics-based test development through the implementation of its recommendations, the committee joins patients, clinicians, and scientists in seeking revolutionary new omics-based tools for improving patient care.

## REFERENCES

National Research Council (NRC). 2003. *Sharing publication-related data and materials: Responsibilities of authorship in the life sciences*. Washington, DC: The National Academies Press.

NRC. 2005. *Catalyzing inquiry at the interface of computing and biology*. Washington, DC: The National Academies Press.

NRC. 2006. *Reaping the benefits of genomic and proteomic research: Intellectual property rights, innovation, and public health*. Washington, DC: The National Academies Press.

# 1
# Introduction

The completion of the sequence of the human genome in 2001 and the technologies that have emerged from the Human Genome Project have ushered in a new era in biomedical science. Using technologies in genomics, proteomics, and metabolomics, together with advanced analytical methods in biostatistics, bioinformatics, and computational biology, scientists are developing a new understanding of the molecular and genetic basis of disease. By measuring, in each patient sample, thousands of genetic variations, mutations, or changes in gene and protein expression and activity, scientists are identifying previously unknown, molecularly defined disease states and searching for complex biomarkers that predict responses to therapy and disease outcome.

This new understanding is beginning to shape both in the way disease is managed and how new drugs and tests are being developed and used. For example, Onco*type* DX (Paik et al., 2004) is a multiparameter gene expression test that helps determine which patients with early stage breast cancer are at higher risk of recurrence and thus may be more likely to benefit from chemotherapy, while allowing women at lower risk to safely forgo chemotherapy. These patients avoid the toxicities, cost, and quality-of-life issues associated with treatment. Increasingly, drugs are being developed to target specific disease subtypes or mutations, and companion diagnostic tests are being developed to identify the subsets of patients most likely to respond or least likely to suffer serious side effects.

Despite great promise, progress in translating such "omics-based" tests into direct clinical applications has been slower than anticipated. This has been attributed to the time-consuming, expensive, and uncertain development pathway from disease biomarker discovery to clinical test; the underdeveloped and inconsistent standards of evidence to assess biomarker validity; the heterogeneity of patients with a given diagnosis; and the lack of appropriate study designs and analytical methods for these analyses (IOM, 2007). Some also have questioned the excitement afforded omics-based discoveries, suggesting that advancements will have primarily modest effects in patient care (Burke and Psaty, 2007).

Nevertheless, patients themselves recognize the promise of molecularly-driven medicine and are looking to the scientific community to provide validated, reliable clinical tests that accurately measure and predict response to treatment and provide more effective ways of screening for disease. Among scientists and clinicians, omics-based tests are seen as presenting opportunities for important new clinical trial design strategies and hopefully reducing the time and cost of developing new treatments (Macconaill and Garraway, 2010).

## ORIGIN OF THE TASK

As is true in all areas of scientific research, rigorous standards must be applied to assess the validity of any study results, particularly if the study involves patients. Recently, the scientific community raised serious concerns about several omics-based tests developed to predict sensitivity to chemotherapeutic agents, developed by investigators at Duke University. The initial papers describing these omics-based tests garnered extensive attention because results

suggested a potential major advance in the discovery and use of genomic signatures to direct choice of therapy for individual cancer patients. Almost from the time of initial publication, however, concerns were raised about the validity of these gene expression–based tests; Keith Baggerly and Kevin Coombes of MD Anderson Cancer Center first approached the Duke University Principal Investigators, Anil Potti and Joseph Nevins with questions on November 8, 2006 (Baggerly, 2011), soon after the October 22 electronic publication of the article (PubMed, 2006), because clinical investigators at their institution were interested in using the methods, but the statisticians could not reproduce the results with the publicly available data and information. These concerns were heightened upon the publication of an article by Baggerly and Coombes (2009), detailing several errors in the development of the tests, inconsistencies between primary data and data used in the articles, and a failure to reproduce results reported by the investigators. In addition, in July 2010, a letter to the director of the National Cancer Institute (NCI) signed by a group of more than 30 respected statisticians and bioinformatics scientists brought additional scrutiny to these concerns, especially because these omics-based tests were being used in clinical trials to direct patient care (Baron et al., 2010). Between October 2007 and April 2008, three cancer clinical trials were launched at Duke University, in which patients with lung cancer or breast cancer were assigned to a chemotherapy regimen on the basis of the test results (see Appendix B for additional details).

Dr. Varmus asked the IOM to conduct an independent analysis of the omics-based tests developed at Duke and define evaluation criteria for ensuring high standards of evidence for the development of omics-based tests prior to their use in clinical trials. In an interview for *The Cancer Letter*, Dr. Varmus summarized the committee's task:

> The Duke episode, from my perspective, was simply another way of illustrating the dangers of not doing it right, not having the right kinds of safeguards. And with my various colleagues, including colleagues at Duke, I asked the Institute of Medicine to do a study. The intention was not to investigate wrongdoing, because that was going to be taken care of in other ways, but to think about what needs to be in place to ensure that correct evaluation of new approaches to cancer care had been undertaken, that we met competing standards, and that the evidence base for changing diagnosis itself or evaluation of responses or, more importantly, choice of therapies—was based on good evidence. I asked the IOM … to think carefully about what kinds of hoops people need to jump through before new information about cancer is actually used in the clinical setting. The risks are high here. (Goldberg, 2011, p. 4)

NCI biostatistician Lisa McShane provided further motivation for the committee's work:

> I have witnessed the birth of many omics technologies and remain excited about their potential for providing important biological insights and their potential to lead to clinical tests that might improve care for cancer patients. It is important, however, that we understand the challenges and potential pitfalls that can be encountered with use of these technologies. Some unfortunate events at Duke University involving the use of genomic predictors in cancer clinical trials were a major impetus for the formation of this committee. We need to take a step back to evaluate the process by which tests based on omics technologies are developed and determined to be fit for use as a basis for clinical trial designs in which they may be used to determine patient therapy. (McShane, 2011, p 1-2)

The scientific community needs to address these gaps if we are to realize the full potential of omics research in patient care. Omics technologies hold great promise, but also pose substantial risks if not properly developed and validated for clinical use.

## COMMITTEE APPOINTMENT AND CHARGE

With support from NCI, the Food and Drug Administration, the Centers for Disease Control and Prevention, the U.S. Department of Veterans Affairs, the American Society for Clinical Pathology, and the College of American Pathologists, an IOM committee was charged to identify appropriate evaluation criteria for developing clinically applicable omics-based tests and to recommend an evaluation process for determining when predictive tests using omics-based technologies are fit for use in clinical trials, especially those in which the assay is used to direct patient care (Box 1-1). The IOM appointed a 20-member committee with a broad range of expertise and experience, including experts in discovery and development of omics-based technologies, clinical oncology, biostatistics and bioinformatics, clinical pathology, ethics, patient advocacy, development and regulation of diagnostic tests, university administration, and scientific publication.

**BOX 1-1**
**Committee Statement of Task**

An ad hoc committee will review the published literature to identify appropriate evaluation criteria for tests based on "omics" technologies (e.g., genomics, epigenomics, proteomics, and metabolomics) that are used as predictors of clinical outcomes. The committee will recommend an evaluation process for determining when predictive tests based on omics technologies are fit for use as a basis for clinical trial design, including stratification of patients and predicting response to therapy in clinical trials. The committee will identify criteria important for the analytical validation, qualification, and utilization components of test evaluation.

The committee will apply these evaluation criteria to predictive tests used in three cancer clinical trials conducted by Duke University investigators (NCT00509366, NCT00545948, NCT00636441). For example, the committee may assess the analytical methods used to generate and validate the predictive models, examine how the source data that were used to develop and test the predictive models were generated or acquired, assess the quality of the source data, and evaluate the appropriateness of the use of the predictive models in clinical trials.

The committee will issue a report with recommendations regarding criteria for using models that predict clinical outcomes from genomic expression profiles and other omics profiles in future clinical trials, as well as recommendations on appropriate actions to ensure adoption and adherence to the recommended evaluation process. The report will also include the committee's findings regarding the three trials in question.

Before the IOM convened for its first meeting, investigators at Duke concluded that the omics-based tests used in the three clinical trials were invalid. They terminated the clinical trials, and began the process of retracting the papers describing the development of the tests. As a result, the committee did not undertake a detailed analysis of the data and computer code used in

the development of those tests. Rather, the committee focused on how errors in the development process resulted in those tests being used in clinical trials before they were fully validated, and on developing best practices that would prevent invalid tests from progressing to the clinical testing stage in the future.

A rigorous process was undertaken in the development of the committee's recommendations that included a review of the field of omics-based research, the processes necessary for verification and validation of omics-based tests, examination of what transpired in the development of the omics-based tests listed in the statement of task as well as other case studies of omics-based test development selected by the committee, and identification of the parties responsible for funding, oversight, and publication of results. Recommendations developed by the committee should be considered a roadmap critical to omics-based test development. The recommendations cover the roles and responsibilities of all partners involved in the process, including individual scientists, their institutions, funding agencies that support the work, journals that publish the results of these studies, and the Food and Drug Administration (FDA), which ultimately helps to define how these tests will make their way to clinical application.

## Outside the Scope

The processes and criteria for adoption and use of omics-based tests in standard clinical practice are outside the scope of this report. The process of taking an omics-based test into clinical trials to evaluate a test for clinical utility and use is described, but no recommendation is made on how, finally, to take a test from the clinical trial setting into clinical practice. However, discussion of this step is critical for understanding the recommendations of the committee because this step may involve using an omics-based test to direct patient management in clinical trials, which is within the charge of the committee. Regardless, if the omics-based test is to be considered for use in clinical practice, one of three pathways needs to be followed to determine clinical utility, and all of these require a fully specified and validated omics-based test. When considering the parties responsible in the development of omics-based tests, the committee considered international funders to be outside the scope of the recommendations. Issues specific to tests that fall outside the committee's definition of omics-based tests, such as single gene tests and whole genome sequencing, are also not addressed.

It is important to note that the IOM's study is in no way linked to the concurrent scientific misconduct investigation at Duke University, and that inquiries about misconduct were not within this committee's purview.

## Definitions

Precise definitions and use of correct terminology are important for ensuring understanding, especially given the complexity of the rapidly expanding field of omics. The committee defined terminology that was central to its deliberations and recommendations (Box 1-2). Where possible, the committee used widely accepted definitions, such as those from the Biomarkers Definition Working Group. The terms "analytical validation," "clinical validation," and "clinical utility" have been adapted from the widely used definitions of the Evaluation of Genomic Applications in Practice and Prevention initiative, established by the Centers for Disease Control and Prevention (Teutsch et al., 2009). The committee has adapted this terminology by incorporating statistics and bioinformatics validation through use of the term "clinical/biological validation."

**BOX 1-2**
**Important Definitions**

**Analytical Validation:** Traditionally, "assessing [an] assay and its measurement performance characteristics, determining the range of conditions under which the assay will give reproducible and accurate data."[a] With respect to omics, assessing a test's "ability to accurately and reliably measure the …analyte[s]…of interest in the clinical laboratory, and in specimens representative of the population of interest."[b]

**Biomarker:** "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n] …intervention."[c]

**Clinical Utility:** "Evidence of improved measurable clinical outcomes, and [a test's] usefulness and added value to patient management decision-making compared with current management without [omics] testing."[b]

**Clinical/Biological Validation:** Assessing a test's "ability to accurately and reliably predict the clinically defined disorder or phenotype of interest."[b]

**Cross-validation:** a statistical method for preliminary confirmation of a computational model's performance using a single data set, by dividing the data into multiple segments, and iteratively fitting the model to all but one segment and then evaluating its performance on the remaining segment.

**Effect Modifier:** A measure that identifies patients most likely to be sensitive or resistant to a specific treatment regimen or agent. An effect modifier is particularly useful when that measure can be used to identify the subgroup of patients for whom treatment will have a clinically meaningfully favorable benefit-to-risk profile.

**High-Dimensional Data:** Large datasets characterized by the presence of many more predictor variables than observations, such as datasets that result from measurements of hundreds to thousands of molecules in a relatively small number of biological samples. The analysis of such datasets requires appropriate computing power and statistical methods.

**Omics:** Scientific disciplines comprising study of related sets of biological molecules. Examples of omics disciplines include genomics, transcriptomics, proteomics, metabolomics, and epigenomics.

**Omics-Based Test:** An assay composed of or derived from many molecular measurements and interpreted by a fully specified computational model to produce a clinically actionable result.

**Overfitting:** Occurs when the model-fitting process unintentionally exploits characteristics of the data that are due to noise, experimental artifacts, or other chance effects that are not shared between data sets, rather than to the underlying biology that is shared between data sets. Overfitting leads to a statistical or computational model that exhibits very good performance on the particular data set on which it is fit, but poor performance on other data sets Although not unique to omics research, the chance of overfitting increases when the model has a large number of measurements relative to the number of samples.

**Preanalytical Variables:** Aspects of sample collection and handling that need to be standardized and documented prior to test development and use.

**Predictive Factor:** An effect modifier of treatment.

**Prognostic Factor:** A measure correlated with a clinical outcome in the setting of natural history or a standard of care regimen; It is a variable used to estimate the risk of or time

to clinical outcomes.

**Statistics and Bioinformatics Validation:** Verifying that the omics-based test can perform its intended task. Ideally, this involves assuring that the test can accurately predict the clinical outcome of interest in an independent set of samples that were not used in developing the test. Such validation is particularly important as omics tests typically involve computational models whose parameters can be overfit in any single dataset, leading to an overly optimistic sense of the test's accuracy.

SOURCES: [a]Wagner (2002); [b]Teutsch et al. (2009); [c]Biomarkers Definitions Working Group (2001).

The committee provides additional scientific and technical definitions in Chapters 2, 3, 4, the glossary, and Appendix C.

## Introduction to Biomarkers

The set of biological information measured and analyzed in a validated omics-based test is an example of a biomarker. This section introduces the concept and history of biomarkers. The scientific literature provides definitions of the term "biomarker" as well as some of the principal uses of biomarkers. A widely used definition of a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n] … intervention" (Biomarkers Definitions Working Group, 2001). A recent IOM report on *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease* provided the following description of biomarkers:

> Biomarkers are measurements of biological processes. Biomarkers include physiological measurements, blood tests and other chemical analyses of tissue or bodily fluids, genetic or metabolic data, and measurements from images. Cholesterol and blood sugar levels are biomarkers, as are blood pressure, enzyme levels, measurements of tumor size from MRI or CT, and the biochemical and genetic variations observed in age-related macular degeneration. Emerging technologies have also enabled the use of simultaneously measured "signatures," or patterns of co-occurring sets, of genetic sequences, peptides, proteins, or metabolites as biomarkers. These signatures can also be combinations of several of these types of measurements; ideally, each component of a signature is identified. (IOM, 2010, p 2-3)

Biomarkers can be measurements of macromolecules (DNA, RNA, proteins, lipids), cells, or processes that describe a normal or abnormal biological state in an organism. Biomarkers may be detected and analyzed in tissue, in circulation (blood, lymph), and in body fluids (urine, stool, saliva, sputum, breast nipple aspiration, etc.). Biomarkers have many important potential roles in settings such as discovery research, clinical practice, and public health practice; these and other biomarker uses are described in Table 1-1 (IOM, 2010).

**TABLE 1-1** Categories of Biomarker Use

| Use | Description |
|---|---|
| Discovery | Identification of biochemical, image, or other biomarkers associated with a disease, condition, or behavior of interest; biomarkers identified may be screened for many potential uses, including as a target for intervention to prevent, treat, or mitigate a disease or condition |
| Early product development | Biomarkers used for target validation, compound screening, pharmacodynamic assays, safety assessments, and subject selection for clinical trials, and as endpoints in early clinical screening (i.e., Phase I and II trials) |
| Surrogate endpoints for claim and product approvals | Biomarkers used for Phase III clinical testing or to substantiate claims for product marketing when the effect of treatment on the biomarker reliably predicts the effect of treatment on a direct measure of how a patient feels, functions or survives. |
| Clinical practice | Biomarkers used by clinicians for uses such as risk stratification, disease prevention, screening, diagnosis, prognosis, therapeutic monitoring, and posttreatment monitoring |
| Clinical practice guidelines | Biomarkers used to make generalized recommendations for healthcare practitioners in the areas of risk stratification, disease prevention, treatment, behavior/lifestyle modifications, and more |
| Comparative efficacy and safety | Biomarkers used in clinical studies looking at the relative efficacy, safety, and cost effectiveness of any or all interventions used for a particular disease or condition, including changes in behavior, nutrition, or lifestyle; these studies are a component of comparative effectiveness research |
| Public health practice | Biomarkers used to track public health status and make recommendations for prevention, mitigation, and treatment of diseases and conditions at the population level |

SOURCE: Adapted from IOM (2010).

Uses intended for clinical practice include risk assessment, screening, diagnosis, prognosis, prediction of response to therapy (effect modifiers), prediction of clinical outcome (surrogate endpoints), and patient monitoring during and after treatment (Table 1-2).

**TABLE 1-2** Use of Biomarkers in Clinical Practice

| Clinical Biomarker Use | Clinical Objective |
|---|---|
| Disease risk stratification | Assess the likelihood that disease will develop (or recur) |
| Screening | Detect and treat early-stage disease in the asymptomatic population |
| Diagnosis/Differential Diagnosis | Definitively establish the presence and precise description of disease |
| Classification[a] | Classify patients by disease subset |
| Prognosis | Estimate the risk of or the time to clinical outcomes. |
| Prediction/treatment stratification[a] | Predict response to particular therapies and choose the drug that is mostly likely to yield a favorable response in a given patient. |
| Therapy-related risk management | Identify patients with a high probability of adverse effects of a treatment |
| Therapy monitoring[b] | Determine whether a therapy is having the intended effect on a disease and whether adverse effects arise |
| Posttreatment monitoring | Early detection and treatment of advancing disease or complications |

[a] Companion diagnostic biomarkers include features from several of these categories. These tests identify whether an individual's molecular profile associated with a disease pathophysiology is likely to respond favorably to a particular therapeutic. Examples include KRAS–cetuximab, HER2–herceptin, and estrogen receptor–tamoxifen.
[b] Dose optimization is a subset of this category.
SOURCE: Adapted from IOM (2007, 2010).

It is important to understand a key distinction between two types of biomarkers: prognostic factors and effect modifiers. Prognostic factors are correlated with a clinical outcome in the setting of a specified clinical regimen. They are used to estimate the risk of or the time to clinical outcomes. However, a pure prognostic factor does not predict whether future, additional patient management strategies or therapies will be effective. Conversely, an effect modifier identifies patients most likely to be sensitive or resistant to a specific treatment regimen or agent. Effect modifiers are particularly useful when they can be used to identify the subgroup of patients for whom treatment will have a clinically meaningful favorable benefit-to-risk profile. In oncology, effect modifiers are also referred to as predictive factors, treatment-guiding biomarkers, or treatment selection biomarkers (Henry and Hayes, 2006; McGuire et al., 1990). While many people frequently use the term "predictive factor" rather than "effect modifier," the use of this term is problematic because most dictionaries indicate that the adjectives predictive and prognostic have very similar meanings. This report uses the term "effect modifier." It should be noted that a biomarker can be both prognostic and an effect modifier.

More detailed descriptions of biomarker types and examples, as well as the types of clinical studies and trials in which biomarkers are developed, are given in Appendix C.

### *Evaluation of Biomarkers and Surrogate Endpoints*

As outlined in the statement of task, this committee was charged with identifying appropriate evaluation criteria for tests based on omics technologies, including criteria for the

analytical validation, qualification, and utilization components of test evaluation. The terminology of analytical validation, qualification, and utilization stems from the IOM consensus report, *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease* (IOM, 2010). The 2010 committee recommended a three-step biomarker evaluation framework consisting of analytical validation, qualification, and utilization, and intended the framework to be applicable to a diverse range of biomarker uses, including panels of biomarkers. The qualification step of biomarker evaluation is parallel to this report's clinical/biological validation step.

The 2010 IOM report emphasized the importance of a test's intended use when making determinations in the utilization stage of biomarker evaluation. If a test's validation did not reach the level needed for its intended use, the test would be sent back for further development. The interdependence of the steps in the evaluation process is highlighted in Figure 1-1.

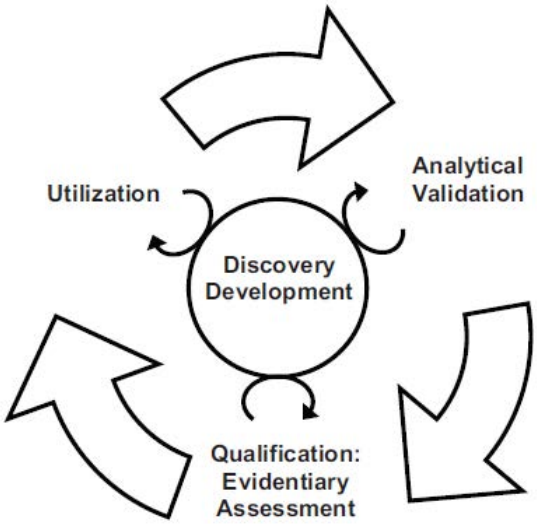


**FIGURE 1-1** The steps of the biomarker evaluation are interdependent.
SOURCE: IOM (2010).

This report's process for discovery and development of omics-based tests can be viewed as an example of how the process above can be applied in a more specific case. The 2010 report covered all types of biomarkers and surrogate endpoints, including single and multiple analyte, molecular or imaging, and quantitative or qualitative biomarkers. Omics-based biomarkers are generally quantitative, involve measurement of multiple analytes, and involve use of computational models.

### *Omics-Based Biomarkers and Omics-Based Tests*

Omics-based tests can be considered a complex form of a biomarker test, using a defined set of measurements combined with a precisely defined computational model as a clinical test, for any of the purposes defined above for biomarkers. Several features distinguish omics-based biomarkers and omics-based tests from other biomarkers and biomarker-based tests. Most importantly, an omics-based test is derived from complex high-dimensional data; these data are often generated through measurement of many more variables per sample than the total number

of biological samples used to generate the dataset. These data are used to produce a computational model[1] that can be used to analyze samples from individual patients. High-dimensional data are particularly prone to overfitting, which can result in a computational model that functions well on the samples used for test development, but is inaccurate on any other sample. With careful analysis and a series of studies leading to a valid test, an omics-based test can be used to help a clinician make a decision about a patient's care.

Several other characteristics distinguish omics-based tests from other medical technologies, including regulation and oversight of the development process and the difficulty in defining the biological rationale behind the test.

Omics-based tests and other clinical laboratory tests are subject to a different regulatory framework than drugs; for example, there are more pathways for regulation of devices—the regulatory category under which omics-based tests fall—than there are for drugs. Also, test development is more likely to occur in an academic setting than for drugs. Because the regulatory and oversight requirements for clinical laboratory tests are both different and less clear than for drugs, a greater burden is placed on the institutions to oversee biomarker-based test research and development. While pharmaceutical companies follow well-established drug development pathways and have many process controls in place for strong oversight of drug development and manufacturing, academic institutions are not as accustomed to overseeing the development of medical products. More work may be needed in academic institutions to reach an appropriate level of oversight.

The frequent lack of a clear biological rationale further distinguishes omics-based tests from other biomarker-based tests. It is usually possible to explain the biological rationale behind a single-biomarker test: The test is useful because the biomarker plays a role in disease pathology or other biological process under investigation. Examples of single-biomarker tests include single-gene tests such as HER2 or tests such as blood levels of LDL cholesterol. For omics-based tests, however, the opposite is often true: It is generally not possible to explain the biological reasons why the test works. This difference puts an additional burden on the statisticians and bioinformatics experts involved in test validation. Due to the risk of overfitting involved with omics-based test development, the need for rigor, validation, and accountability is even higher than for other biomarker-based tests.

## ENGAGEMENT OF STAKEHOLDERS AND IMPLEMENTATION OF THE RECOMMENDATIONS

The future of omics-based tests and the promise they hold may well depend on the adoption of the recommendations put forth in this report. This report will be relevant to multiple audiences, including the responsible parties to whom the recommendations are directed and the various scientific disciplines and professions involved in the discovery and translation of omics-based biomarkers and tests. These responsible parties include the biomedical and clinical research community, investigators, institutions—public and private, commercial and non-profit—funders of omics-based research, and journals that publish omics-based research and clinical trials. Finally, the general public, as potential clinical trial participants and the beneficiaries of the products developed through this research, also may be interested. However,

[1] Includes all data processing steps, normalization techniques, weights, parameters, and other aspects of the model, as well as the mathematical formula or formulas used to convert the data into a prediction of the phenotype of interest.

the technical aspects of this report are intended for a scientific audience. The omics test development process is complex and thus calls for a complex and rigorous methodology.

Many different scientific disciplines and professions are involved in the discovery, development, and validation of omics-based tests. Investigators that conduct the studies and the institutions that oversee their research are key to adoption and implementation of these recommendations. Laboratory scientists in many fields engage in omics-based research. Quantitative scientists, including those trained in biostatistics and bioinformatics, are an essential part of the scientific team in omics-based research because of the need to manage and use large datasets and to generate and validate the predictive models. Clinical researchers are responsible for the design and implementation of clinical trials that assess the clinical utility of new tests and can play a major role in the adoption of these recommendations. Organizations that fund this research and scientific journals that publish the results are also integral to advancing this research. The FDA, as a regulatory agency, has a substantial role in oversight of this science as well.

## ORGANIZATION OF THE REPORT

Chapters 2 through 4 describe the recommended omics-based test evaluation process that the committee was charged with developing.

Chapter 2 provides an overview of the science and technology underpinning omics-based research and the recommended discovery and confirmation process prior to development and validation of an omics-based clinical test.

Chapter 3 describes the processes for defining the clinical test method and for assessing analytical and clinical/biological validation prior to use of an omics-based test in a clinical trial to evaluate the clinical utility and use of the test.

Chapter 4 describes the process for assessing the clinical utility and use of a new omics-based test.

Chapter 5 examines the roles of responsible parties in the development of omics-based tests: the roles of investigators and institutions in ensuring high-quality discovery and test development, the roles of journals and funders in the publication and financing of research to develop omics-based tests, and the role of the FDA.

Chapter 6 presents an overview of lessons learned from the case studies, which are described in detail in Appendix A and B.

Recommendations are presented in Chapters 2-5 and in the Summary.

## REFERENCES

Baggerly, K. A. 2011. *Forensics bioinformatics.* Presented at the Workshop of the IOM Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30-31.

Baggerly, K. A., and K. R. Coombes. 2009. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Annals of Applied Statistics* 3(4):1309-1334.

Baron, A. E., K. Bandeen-Roche, D. A. Berry, J. Bryan, V. J. Carey, K. Chaloner, M. Delorenzi, B. Efron, R. C. Elston, D. Ghosh, J. D. Goldberg, S. Goodman, F. E. Harrell, S. Galloway Hilsenbeck, W. Huber, R. A. Irizarry, C. Kendziorski, M. R. Kosorok, T. A. Louis, J. S. Marron, M. Newton, M. Ochs, J. Quackenbush, G. L. Rosner, I. Ruczinski, S. Skates, T. P. Speed, J. D. Storey, Z. Szallasi, R. Tibshirani, and S. Zeger. 2010. Letter to Harold Varmus: Concerns about prediction models used in Duke clinical trials. Bethesda, MD, July 19. Available online at http://www.cancerletter.com/categories/documents (accessed January 18, 2012).

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 69(3):89-95.

Burke, W., and B. M. Psaty. 2007. Personalized medicine in the era of genomics. *Journal of the American Medical Association* 298(14):1682-1684.

Goldberg, P. 2011. A year at NCI: Harold Varmus reflects on provocative questions, Duke scandal, financial disaster and grant review. *The Cancer Letter* 37(29):1-7.

Henry, N. L., and D. F. Hayes. 2006. Uses and abuses of tumor markers in the diagnosis, monitoring, and treatment of primary and metastatic breast cancer. *Oncologist* 11(6):541-552.

IOM (Institute of Medicine). 2007. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

IOM. 2010. *Evaluation of biomarkers and surrogate endpoints in chronic disease*. Washington, DC: The National Academies Press.

Macconaill, L. E., and L. A. Garraway. 2010. Clinical implications of the cancer genome. *Journal of Clinical Oncology* 28(35):5219-5228.

McGuire, W. L., A. K. Tandon, D. C. Allred, G. C. Chamness, and G. M. Clark. 1990. How to use prognostic factors in axillary node-negative breast cancer patients. *Journal of the National Cancer Institute* 82(12):1006-1015.

McShane, L. 2010. *NCI address to the Institute of Medicine committee convened to review omics-based tests for predicting patient outcomes in clinical trials*, Washington, DC. December 20.

NRC (National Research Council). 2009. *On being a scientist: A guide to responsible conduct in research*, 3rd ed. Washington, DC: The National Academies Press.

Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, and N. Wolmark. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine* 351(27):2817-2826.

PubMed. 2006. PubMed entry for *Genomic signatures to guide the use of chemotherapeutics* by Potti et al., *Nature Medicine,* 2006. http://www.ncbi.nlm.nih.gov/pubmed/17057710 (accessed October 18, 2011).

Teutsch, S. M., L. A. Bradley, G. E. Palomaki, J. E. Haddow, M. Piper, N. Calonge, D. Dotson, M. P. Douglas, and A. O. Berg. 2009. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group. *Genetics in Medicine* 11(1):3-14.

Wagner, J. A. 2002. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 18(2):41-46.

# 2
# Omics-Based Clinical Discovery: Science, Technology, and Applications

Since the process of mapping and sequencing the human genome began, omics has been a burgeoning field of research. In the past 10 to 15 years, new technologies have made it possible to obtain a huge number of molecular measurements within a tissue or cell. These technologies can be applied to a biological system of interest, to obtain a snapshot of the underlying biology at a resolution that has never before been possible. Broadly speaking, the scientific fields associated with measuring such biological molecules in a high-throughput way are called "omics."

Many areas of research can be classified as omics. Examples include proteomics, transcriptomics, genomics, metabolomics, lipidomics, and epigenomics, which correspond to global analyses of proteins, RNA, genes, metabolites, lipids, and methylated DNA or modified histone proteins in chromosomes, respectively. There are many motivations for conducting omics research. One common reason is to obtain a comprehensive understanding of the biological system under study. For instance, one might perform a proteomics study on normal human kidney tissues to better understand protein activity, functional pathways, and protein interactions in the kidney. Another common goal of omics studies is to associate the omics-based molecular measurements with a clinical outcome of interest, such as prostate cancer survival time, risk of breast cancer recurrence, or response to therapy. The rationale is that by taking advantage of omics-based measurements, there is the potential to develop a more accurate predictive or prognostic model of a particular condition or disease—namely, an omics-based test (see definition in the Introduction)—that is more accurate than can be obtained using standard clinical approaches.

This report focuses on the development of omics-based tests prior to use to direct treatment choice in a clinical trial. In this chapter, the discovery phase (see Figures 2-1 and S-1) of the recommended omics-based test development process is discussed, beginning with examples of specific types of omics studies and the technologies involved, followed by the statistical, computational, and bioinformatics challenges that arise in the analysis of omics data. Some of these challenges are unique to omics data, whereas others relate to fundamental principles of good scientific research. The chapter concludes with discussion of emerging directions for omics research as they relate to the discovery and future development of omics-based tests for clinical use.

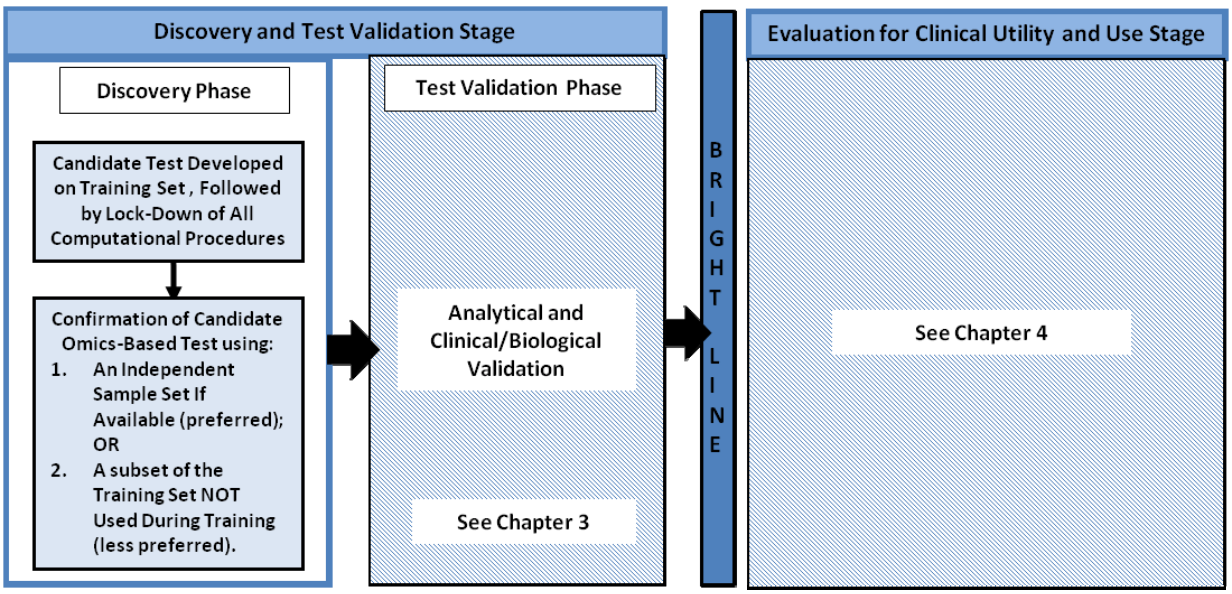


**FIGURE 2-1** Omics-based test development process, highlighting the discovery phase. In the discovery phase, a candidate test is developed, precisely defined, and confirmed. The computational procedures developed in this phase should be fully specified and locked down through all subsequent development steps. Ideally, confirmation should take place on an independent sample set. Under exceptional circumstances it may be necessary to move into the test validation phase without first confirming the candidate test on an independent sample set if using an independent test set in the discovery phase is not possible, but this increases the risk of test failure in the validation phase. Statistics and bioinformatics validation occurs throughout the discovery and test validation stage as well as the stage for evaluation of clinical utility and use.

## TYPES OF OMICS DATA

Examples of the types of omics data that can be used to develop an omics-based test are discussed below. This list is by no means meant to be comprehensive, and indeed a comprehensive list would be impossible because many new omics technologies are rapidly developing.

### Genomics

The genome is the complete sequence of DNA in a cell or organism. This genetic material may be found in the cell nucleus or in other organelles, such as mitochondria. With the exception of mutations and chromosomal rearrangements, the genome of an organism remains essentially constant over time. Complete or partial DNA sequence can be assayed using various experimental platforms, including single nucleotide polymorphism (SNP) chips and DNA sequencing technology. SNP chips are arrays of thousands of oligonucleotide probes that hybridize (or bind) to specific DNA sequences in which nucleotide variants are known to occur. Only known sequence variants can be assayed using SNP chips, and in practice only common variants are assayed in this way. Genomic analysis also can detect insertions and deletions and copy number variation, referring to loss of or amplification of the expected two copies of each gene (one from the mother and one from the father at each gene locus). Personal genome sequencing is a more recent and powerful technology, which allows for direct and complete sequencing of genomes and transcriptomes (see below). DNA also can be modified by methylation of cytosines (see Epigenomics, below). There is also an emerging interest in using genomics technologies to study the impact of an individual's microbiome (the aggregate of

microorganisms that reside within the human body) in health and disease (Honda and Littman, 2011; Kinros et al., 2011; Tilg and Kaser, 2011).

### Transcriptomics

The transcriptome is the complete set of RNA transcripts from DNA in a cell or tissue. The transcriptome includes ribosomal RNA (rRNA), messenger RNA (mRNA), transfer RNA (tRNA), micro RNA (miRNA), and other non-coding RNA (ncRNA). In humans, only 1.5 to 2 percent of the genome is represented in the transcriptome as protein-coding genes. The two dominant classes of measurement technologies for the transcriptome are microarrays and RNA sequencing (RNAseq). Microarrays are based on oligonucleotide probes that hybridize to specific RNA transcripts. RNAseq is a much more recent approach, which allows for direct sequencing of RNAs without the need for probes. Oncotype DX, MammaPrint, Tissue of Origin, AlloMap, CorusCAD, and the Duke case studies described in Appendix A and B all involve transcriptomics-based tests.

### Proteomics

The proteome is the complete set of proteins expressed by a cell, tissue, or organism. The proteome is inherently quite complex because proteins can undergo posttranslational modifications (glycosylation, phosphorylation, acetylation, ubiquitylation, and many other modifications to the amino acids comprising proteins), have different spatial configurations and intracellular localizations, and interact with other proteins as well as other molecules. This complexity can lead to challenges in proteomics-based test development. The proteome can be assayed using mass spectrometry and protein microarrays (reviewed in Wolf-Yadlin et al., 2009). Unlike RNA transcripts, proteins do not have obvious complementary binding partners, so the identification and characterization of capture agents is critical to the success of protein arrays. The Oval and Ovacheck tests discussed in Appendix A are proteomics-based tests.

### Epigenomics

The epigenome consists of reversible chemical modifications to the DNA, or to the histones that bind DNA, and produce changes in the expression of genes without altering their base sequence. Epigenomic modifications can occur in a tissue-specific manner, in response to environmental factors, or in the development of disease states and can persist across generations.. The epigenome can vary substantially among different cell types within the same organism. Biochemically, epigenetic changes that are measured at high-throughput belong to two categories: methylation of DNA cytosine residues (at CpG) and multiple kinds of modifications of specific histone proteins in the chromosomes (histone marks). RNA editing is another mechanism for epigenetic changes in gene expression, measured primarily by transcriptomic methods (Maas, 2010).

### Metabolomics

The metabolome is the complete set of small molecule metabolites found within a biological sample (including metabolic intermediates in carbohydrate, lipid, amino acid, nucleic acid, and other biochemical pathways, along with hormones and other signaling molecules, as well as exogenous substances such as drugs and their metabolites). The metabolome is dynamic, and can vary within a single organism and among organisms of the same species, due to many

factors such as changes in diet, stress, physical activity, pharmacological effects, and disease. The components of the metabolome can be measured with mass spectrometry (reviewed in Weckwerth, 2003) as well as by nuclear magnetic resonance spectroscopy (Zhang et al., 2011). This method also can be used to study the lipidome (reviewed in Seppanen-Laakso and Oresic, 2009), which is the complete set of lipids in a biological sample.

## EMERGING OMICS TECHNOLOGIES AND DATA ANALYSIS TECHNIQUES

Many emerging omics technologies are likely to influence the development of omics-based tests in the future, as both the types and numbers of molecular measurements continue to increase. Furthermore, advancing bioinformatics and computational approaches are enabling improved analyses of omics data, such as greater integration of different data types. Given the rapid pace of development in these fields, it is not possible to list all relevant emerging technologies or data analytic techniques. A few illustrative developments are briefly discussed.

Advances in RNA sequencing technology are making possible a higher resolution view of the transcriptome. These new approaches could facilitate the development of more novel molecular diagnostics. In the future it may be possible to develop omics-based tests on the basis of small non-coding RNAs, RNA editing events, or alternative splice variants that were not measured using previous hybridization-based technologies such as microarrays. For example, miRNA signatures (derived from RNA sequencing) show great promise for clinical diagnostics (Moussay et al., 2011; Sugatani et al., 2011; Tan et al., 2011; Yu et al., 2008).

Similarly, DNA sequencing is making it possible to identify rare or previously unmeasured mutations that may have important clinical implications. Next-generation sequencing technologies hold tremendous promise for identifying not only complete DNA and RNA sequences, but also high-throughput identification of epigenetic and posttranscriptional modifications to DNA or RNA, respectively. For instance, new sequencing technologies can monitor a wide variety of epigenetic changes at the genomic scale, in addition to sequencing information.

However, it is important to note that because next-generation RNA and DNA sequencing produces even more measurements per sample than do traditional approaches, these new technologies add to the challenge of extremely high data dimensionality and the risks of overfitting computational models to the available data (see the section on Computational Model Development and Cross-Validation for a discussion of overfitting). Large meta-analyses of sequencing datasets collected at multiple sites may prove useful for overcoming these risks and aid in developing clinically useful omics tests.

The field of proteomics has benefited from a number of recent advances. One example is the development of selected reaction monitoring (SRM) proteomics based on automated techniques (Picotti et al., 2010). During the past 2 years, multiple peptides distinctive for proteins from each of the 20,300 human protein-coding genes have been synthesized and their mass spectra determined. The resulting SRM Atlas is publicly available for the entire scientific community to use in choosing targets and purchasing peptides for quantitative analyses (Omenn et al., 2011). In addition, data from untargeted “shotgun” mass spectrometry-based proteomics have been collected and uniformly analyzed to generate peptide atlases for plasma, liver, and other organs and biofluids (Farrah et al., 2011).

Meanwhile, antibody-based protein identification and tissue expression studies have progressed remarkably (Ayoglu et al., 2011; Fagerberg et al., 2011); the Human Protein Atlas

has antibody findings for more than 12,000 of the 20,300 gene-coded proteins. The Protein Atlas is a tremendous resource for planning experiments, and will be enhanced by linkage with mass spectrometry findings through the emerging Human Proteome Project (Legrain et al., 2011).

As another example, recently developed protein capture-agent aptamer chips can be used to make quantitative measurements of approximately 1,000 proteins from the blood or other sources (Gold et al., 2010). For example, Ostroff et al. (2010) recently reported generation of a 12-protein panel from analysis of 1,100 plasma proteins that was shown to have promising clinical test characteristics for diagnosis of non-small cell lung cancers.

A major bottleneck in the successful deployment of large-scale proteomic approaches is the lack of high-affinity capture agents with high sensitivity and specificity for particular proteins (including variants due to posttranslational modifications, alternative splicing, and single-nucleotide polymorphisms or gene fusions). This challenge is exacerbated in highly complex mixtures such as blood, where the concentrations of different proteins vary by more than 10 orders of magnitude. One technology that holds tremendous promise in this regard is "click chemistry" (Service, 2008), which uses a highly specific chemical linkage (generally formed through the Huisgen reaction) to "click" together low-affinity capture agents to create a single capture agent with much higher affinity. It also is feasible to combine computational algorithms for modeling protein structures and conformation to infer functional differences among alternative splice isoforms of proteins, including those involved in key cancer pathways (Menon et al., 2011).

Improving technologies for measurements of small molecules (Drexler et al., 2011) also is enabling the use of metabolomics for the development of candidate omics-based tests with potential clinical utility (Lewis and Gerszten, 2010). Promising early examples include a metabolomic signature that identified a role for sarcosine, an N-methyl derivative of the amino acid glycine, in prostate cancer progression and metastasis (Sreekumar et al., 2009), metabolomic characterization of ovarian epithelial carcinomas (Ben Sellem et al., 2011), and an integrated metabolomic and proteomic approach to diagnosis, prediction, and therapy selection for heart failure (Arrell et al., 2011). Included within metabolomics is the emerging ability to more fully measure the lipids in a sample, a rich source of additional potential biomarkers (Masoodi et al., 2010). As with other omics data types, a lengthy, complex development path is necessary to establish a clinically relevant omics-based test from reports identifying metabolite concentration differences associated with a phenotype of interest (Koulman et al., 2009).

New technologies are emerging that will make it possible to obtain omics measurements (such as transcriptomics, proteomics) on single cells (Tang et al., 2011; Teague et al., 2010). Such detailed molecular measurements provide deep insight into the underlying biology of tissues, and potentially form a powerful basis for omics-based test development. However, as the resolution of these measurements increases, so too does the variability in the measurements due to the heterogeneity of cell states (Ma et al., 2011). Thus, while emerging omics technologies hold great potential for the development of omics-based tests, they also may exacerbate dangers of overfitting the computational model to the datasets.

Recent interest has focused on measuring multiple omics data types on a single set of samples, in order to integrate different types of molecular measurements into an omics-based test. Such multidimensional datasets have the potential to provide deep insight into biological mechanisms and networks, allowing for the development of more powerful clinical diagnostics. An encouraging example of simultaneous measurement of multiple types of omics data is the DNA-encoded antibody libraries approach (Bailey et al., 2007), which can measure

DNA, RNA, and protein from the same sample. Another example is the analysis of histone modifications to generate an epigenetic gene signature for prostate cancer prognosis (Bianco-Miotto et al., 2010).

Approaches that integrate multiple omics data types within the same clinical test are expected to grow in importance as the number of simultaneous measurements that can be made continues to increase. While it is relatively straightforward to increase the number of genomic and transcriptomic measurements (because DNA and RNA have complementary binding partners), increasing the number of protein measurements is more challenging due to the need for high-affinity capture agents, as discussed previously in this section.

Systems approaches that integrate multiple data types in functionally-based models can be advantageous for the development of omics-based tests. For instance, the analysis of omics measurements in the context of biomolecular networks or pathways can help to reduce the number of variables in the data by constraining the possible relationships between variables, ultimately leading to more robust and clinically useful molecular signatures. General approaches for using prior biological knowledge to enhance signal in omics data include removing measurements that are believed to be noise or for which there is no support in the published biological literature (filtering), using pathway databases or other sources to guide model construction, and aggregating individual measurements, often across data types, to integrate multiple sources of evidence to support conclusions (Ideker et al., 2011). For example, in a study of prion-mediated neurodegeneration, data from five mouse strains and three prion strains were used to identify the transcripts, pathways, and networks that were commonly perturbed across all genetic backgrounds (Hwang et al., 2009; Omenn, 2009).

Genome-wide association study datasets can be analyzed within the context of biological pathways in order to increase identification of disease-related mutations (Segre et al., 2010). The incorporation of evolutionarily conserved gene sets can lead to the identification of often surprising factors in disease (McGary et al., 2010). Large-scale mechanistic network models (for example, for metabolic, regulatory, or signaling networks) may be used to identify biomarkers grounded in disease mechanisms (Folger et al., 2011; Frezza et al., 2011; Gottlieb et al., 2011; Lewis et al., 2010; Shlomi et al., 2011). Genomics, transcriptomics, proteomics, and metabolomics data can be combined with structural protein analysis in order to predict drug targets or even drug off-target effects (Chang et al., 2010). While computational models of biomolecular networks for eventual clinical use are still in their infancy, their potential for providing stronger mechanistic underpinnings to omics-based test development is encouraging.

During the past 10 years, much of the effort to identify genes linked to disease and other conditions of biological interest has focused on genome-wide association studies, in which a set of cases and controls are sampled from a large population and genotyped, and each mutation identified is evaluated for association with the phenotype of interest. However, more recent work has successfully identified disease-causal genes using whole genome or exome sequencing (Ng et al., 2010; Roach et al., 2010). Such studies may prove enormously beneficial for the development of omics-based tests, and indeed such strategies are being used clinically today for the identification of the causal gene mutation resulting in unidentified and uncommon inherited disease states.

## STATISTICS AND BIOINFORMATICS DEVELOPMENT OF OMICS-BASED TESTS

In recent years, a large number of papers have reported new omics-based discoveries and development of new candidate omics-based tests: that is, computational procedures applied to omics-based measurements to produce a clinically actionable result. However, few of these candidate omics-based tests have progressed to clinical use (Ransohoff, 2008, 2009). Some of this discrepancy may be due to the inevitable time lapse of moving from initial identification of a candidate omics-based test to a precisely defined and validated test that can be used clinically.

However, more important are the many significant challenges in the formulation of appropriate research questions and in research design and conduct that confront the successful discovery of candidate omics-based tests, including the complexity of the data and the need for rigorous analyses, and the frequent lack of a plausible biological mechanism underpinning many of these discoveries. These challenges need to be addressed in order to realize the enormous clinical potential of omics research, taking into account issues specific to the field as well as broader principles of good scientific research.

Two primary scientific causes for failure of a candidate omics-based test to progress to clinical use are:

1. An omics-based test may not be adequately designed for answering a specific, well-defined, and relevant clinical question. This crucial point is addressed in Chapters 3 and 4.
2. Omics-based discovery studies may not be conducted with adequate statistical or bioinformatics rigor, making it unlikely or even impossible that the candidate test will prove to be clinically valid or useful. This critical problem is addressed in the remainder of this chapter.

Figure 2-1 highlights omics-based biomarker discovery and confirmation, the first component of the committee's recommended omics-based test development and evaluation process**. When candidate omics-based tests from the discovery phase are intended for further clinical development, several criteria should be satisfied and fully disclosed (for example, through publication or patent application) to enable independent verification of the findings (Recommendation 1),** as discussed below. For the purpose of this discussion, the committee assumed that a clearly defined and clinically relevant scientific or clinical question or questions have been identified, and an omics dataset from analyses of a set of patient samples, along with an associated clinical outcome for each patient, is available.

For example, an investigator may ask whether gene expression measurements could be used to predict recurrence in node-negative breast cancer samples in a way that is substantially more accurate than standard clinical prognostic factors, such as tumor size and grade. The investigator might have data consisting of gene expression measurements for breast cancer tissue samples obtained from patients with node-negative breast cancer, along with disease-free survival time for each patient following surgery. The goal would be to develop a defined assay method for data generation and a fully specified computational procedure[1] that can be used to reliably predict, on the basis of gene expression measurements on a new patient sample, whether a patient's cancer will recur.

[1] All component steps of the computational procedure-- namely, all data processing steps, normalization techniques, weights, parameters, and other aspects of the model, as well as the mathematical formula or formulas used to convert the data into a prediction of the phenotype of interest -- are completely formulated in writing.

Before embarking on omics-based discovery, it is worth considering whether or not the test that will eventually be developed has a reasonable chance of demonstrating clinical validity and utility. For example, the sensitivity and specificity needed, particularly in light of the prevalence of the condition in the population to be tested, should be considered (see also Appendix A, page 37, for a discussion of sensitivity and specificity needs for an ovarian cancer screening test).

Several steps need to be followed to achieve this goal: (1) data quality control; (2) computational model development and cross-validation; (3) confirmation of the computational model on an independent dataset; and (4) release of data, code, and the fully specified computational procedures to the scientific community. Each of these is discussed below.

### Step 1: Data Quality Control

As in most areas of science, data quality control is a crucial first step. Because omics datasets are typically composed of many thousands, if not millions, of measurements, data quality control is often performed computationally. For instance, an investigator might remove genes expressed across conditions near or below background levels on a microarray. The reproducibility of the measurements from run to run (the technical variance) also can be assessed. Furthermore, it may be useful to closely examine aspects of experimental design, including sample run date and other possible confounding factors such as the source of the tissue analyzed (including normal control tissue) and potential heterogeneities within the tissues, to determine if these have had an effect on the data. This is particularly important because factors such as run date or machine operator can in some instances have a much larger effect on omics measurements than the factors of biological interest (Leek et al., 2010), such as time to disease recurrence or cancer subtype.

It is essential that such quality assessment evaluations of the data be done in a blinded fashion, without knowledge of the clinical status or treatment outcomes of the patients whose specimens were tested. In addition, investigators should define expectations for successful confirmation of a computational model before proceeding to step 3.

### Step 2: Computational Model Development and Cross-Validation

Once investigators have determined in step 1 that the data are of adequate quality, a candidate omics-based test associated with a phenotype of interest, such as a biologic subgroup, preclinical responsiveness to a novel therapy, or a clinical outcome, can be developed on the basis of the omics measurements. An almost unlimited number of statistical tools can be used to perform this task; therefore, they are not enumerated here. However, some key characteristics and challenges are shared by nearly all of these methods and are discussed below.

In general, omics datasets consist of thousands to millions of molecular measurements. Typically, feature selection is first performed, which entails selecting a subset of the measurements that appear to be associated with the characteristic or outcome or that is thought to be biologically relevant based on prior knowledge. Using just this subset of measurements, a fully defined computational model can be developed to predict the clinical outcome on the basis of the omics measurements. This reduction of required measurements can be beneficial for avoiding the later possibility that an omics-based test involving a huge number of measurements is not clinically viable for financial or technical reasons. Note that if cross-validation will be performed in order to select tuning parameters or evaluate the computational model performance,

then feature selection must be carefully performed as part of the cross-validation process, as will be explained below.

Because omics datasets typically are composed of an extremely large number of molecular measurements, and because the sample size, in general, is quite limited relative to the number of molecular measurements, *overfitting*[2] the data is a major concern. In fact, in the absence of adherence to proper statistical procedures, it is likely that the data will be overfit. That is, given a typical omics dataset and an associated clinical outcome, it is nearly always possible to develop a computational model that fits the data perfectly, even in the absence of any true association between the omics measurements and the clinical outcome. However, such a model will be ineffective, as it will perform very poorly on an independent test dataset; a computational model's ultimate utility is measured by its performance on future patients rather than its performance on patients comprising the original dataset used to develop the computational model.

While it is always important to increase the "power" of the statistical analysis by using the largest sample size possible, overfitting will always still be a concern in any analysis of omics data due to the vast number of feature measurements. It is unrealistic that investigators could ever have such a large sample size that overfitting wouldn't be a concern.

The best way to avoid developing a computational model that overfits the data is to develop the model using a training set/test set approach, and to not use models with large numbers of parameters that are not justified by the sample sizes. When a limited availability of appropriate samples makes this approach infeasible, developers sometimes opt to use a less stringent process called cross-validation. First the training set/test set approach is described, followed by the cross-validation approach.

Ideally, an investigator will develop a computational model using two distinct data sets, referred to as a training set and a test set, each composed of independent samples that have been collected and processed by different sets of investigators at different institutions. Because any computational model contains a number of possible tuning parameters, and there are multiple ways to normalize the data, each choice of tuning parameter, normalization technique, and so forth can be considered a separate computational model.

First, investigators fit each computational model under consideration on the training samples. Note that if feature selection is performed before or as part of the model-fitting process, then those features must be selected on the basis of the training samples only. It is important that no information about the test set is used to fit the model on the training set in order to get a fair estimate of the error. Once the model is trained, investigators then evaluate its performance on the test sample set. Investigators should be aware that if multiple models were considered, then the best test performance observed may be an overestimate of the performance (or correspondingly an underestimate of the error rate) that will be observed on future samples. The tuning parameters and normalization techniques corresponding to the model that performed best on the test samples will be used one final time to build the model on the training and test set together—mirroring precisely what was done on the training set—since the final development of the candidate omics-based test should use all available data. The error estimate for the final

[2] Overfitting occurs when the model-fitting process unintentionally exploits characteristics of the data that are due to noise, experimental artifacts, or other chance effects that are not shared between data sets, rather than to the underlying biology that is shared between data sets. Overfitting leads to a statistical or computational model that exhibits very good performance on the particular data set on which it is fit, but poor performance on other data sets Although not unique to omics research, the chance of overfitting increases when the model has a large number of measurements relative to the number of samples.

model learned from all the data is what was established in the test set performance (and not the apparent accuracy of this final model, which would be overly optimistic).

At this point, the fully specified computational procedures are *locked down*, and the investigator proceeds into Step 3, in which the chosen model is evaluated on an independent data set. In other words, the precisely defined series of computational steps performed in processing the raw data, as well as the mathematical formula or formulas used to convert the data into a prediction of the phenotype of interest, are recorded and *no longer changed* (see also a more detailed discussion of computational procedure lock down on page 30).

In many studies, however, only a limited number of samples are available, and so developing separate training and test sets is not feasible. An alternative to having designated training and test sets is "cross-validation," a statistical method for preliminary confirmation of a model's performance using a single data set, by dividing the data into multiple segments, and iteratively fitting the model to all but one segment and then evaluating its performance on the remaining segment. (Cross-validation should not be confused with analytical and clinical/biological validation, as described in Chapter 3.) If performed properly, cross validation can be expected to mitigate over-fitting, but it does not necessarily eliminate it.

In greater detail, cross-validation involves splitting the samples that constitute a single data set into K sample sets. Then, K-1 of these sample sets are used to fit a number of computational models (with different tuning parameter values and normalization techniques) and the models are evaluated on the remaining, held-out sample set. This process is repeated until each of the K sample sets has been used to evaluate the models. Then, the investigators select a single computational model to fit on the entire set of samples, using the tuning parameter values and normalization technique that performed best, overall, on the held-out sample sets. This computational model is then locked down, and the investigator proceeds to Step 3.

Cross-validation provides a measure of the extent to which a given computational model can be expected to perform well on future observations, as opposed to having overfit the data used to train the model. Hence, a computational model that performs well in cross-validation is more likely to perform well on future patient samples than one that does not. In fact, a computational model that performs poorly in cross-validation has little chance of leading to a successful omics-based test. Though cross-validation is a simple approach, if not performed carefully, problems can arise.

For example, in some published studies, the subset of omics measurements that has the highest association with the clinical outcome of interest is identified on the basis of all of the samples in the data set. Then cross-validation is performed using that subset of measurements. Unfortunately, the resulting cross-validation error rates grossly underestimate the true error rates because the clinical outcome of interest on the held-out samples in each cross-validation fold has already been looked at to identify the subset of omics measurements. In other words, the held-out samples are not truly "held out." (For instance, this was the case with the cross-validation performed in the development of the MammaPrint test, which was described by Simon et al., [2003] as improperly performed "partial cross-validation," see Appendix A).

This can lead to substantial overfitting of the data and produces omics-based tests that appear to perform well in cross-validation, but are unlikely to perform well on future patient samples (Simon et al., 2003). To perform cross-validation correctly, all aspects of model-fitting, including feature selection and data processing, must be performed using only the K-1 sample sets used to train the computational model in each iteration of cross-validation.

Though both the training set/test set approach and the cross-validation approach provide error rates that estimate the accuracy of the computational model on independent test samples, these error rates can be highly optimistic. Cross-validation error rates tend to be overly optimistic because by randomly splitting the data it is guaranteed that the training and test sets within each cross-validation fold are drawn from the same population distribution. This means that a whole host of relevant sources of variance that affect clinical performance are ignored. If the training set/test set approach is used, then the resulting error rate may be overly optimistic because there may be similarities between the way that the training and test set samples were processed – for instance, in terms of the experimental protocol used – that may not be shared by future patient samples. Thus, neither cross-validation nor the training set / test set approach can be used in place of confirmation on an independent dataset, to be discussed in Step 3. Once a computational model has been selected using the training set/test set approach or using cross-validation, this model should be locked down before proceeding to Step 3.

Occasionally, an algorithm used to develop a computational model may contain some stochastic elements, such as k-means clustering. If this is the case, random seeds should be stored, and robustness of the results from multiple runs should be reported in the publications describing the discovery phase results. Before proceeding to Step 3, the computational model needs to be fully locked down so there is no longer any element of randomness – that is, a single random seed must be selected. The failure to select a random seed and lock down the model was a significant error in the development process for tests developed at Duke University (McShane, 2010)[3]. An investigator should be able to fully specify and publish the computational procedure developed in Step 2 that will be further investigated in Step 3. In addition, investigators should define expectations for successful confirmation of a computational model before proceeding to Step 3.

Once a single computational model is identified using the training set / test set or cross-validation approach, this model should be locked down and the investigators should proceed into Step 3. However, under certain circumstances, this computational model may not have shown adequate training set/test set performance or cross-validation performance. In this case, further improvement or refinement of the computational model may be necessary. After refinement, the modified computational model must once again be evaluated using the training set/test set approach or cross-validation.

The fact that cross-validation or the training set/test set process was performed repeatedly must be reported in detail in any publications describing the candidate omics-based test, as repeated refinements of the model can contribute to overfitting and can increase the probability that the model will show a deterioration in performance when applied to independent samples in Step 3. The process of refinement/improvement should not extend into Step 3; that is, before proceeding into Step 3, the computational model must be locked down.

### Step 3: Confirmation on an Independent Dataset

As mentioned in the previous step, an error estimation approach such as cross-validation should be applied while developing a candidate omics-based test in order to avoid overfitting of the data. However, such an approach will generally yield an *underestimate* of the error rate that will result from applying the computational model to future patient samples. This is because

[3] McShane, L. M. 2010. Notes from June 29 meeting with Duke.

when cross-validation is applied, the samples used to develop the model and the samples used to assess its performance typically share many characteristics in common, such as the patient population from which the samples are obtained or the lab in which the samples were processed. That is, the samples typically correspond to a relatively homogeneous set of patients, collected at one or a few institutions, which were run on the same machine at approximately the same point in time.

This does not correspond to the intended use case of an omics-based test, which will typically involve a much more heterogeneous patient population with samples run in different laboratories at different times. This is particularly important because the variability in omics data due to differences in time, laboratories, technicians, and patient populations often exceeds the variability that is due to differences that are of scientific interest, such as those that are associated with clinical outcome (Leek et al., 2010).

Furthermore, cross-validation estimates can suffer from bias due to the fact that multiple tuning parameters and models are typically considered and only the best is selected, as well as high variance. Therefore, cross-validation error rates provide insufficient evidence of a candidate test's performance. To avoid the wasted time, energy, cost, and resources associated with taking a test that has little chance of success into the later phases of the process, it is important to confirm the computational model on an independent test set in the discovery phase.

**Candidate omics-based tests should be confirmed using an independent set of samples not used in generation of the computational model and, when feasible, blinded to any outcome or phenotypic data until after the computational procedures have been locked down, and the candidate omics-based test has been applied to the samples (Recommendation 1a).** It is important that no samples used to develop the computational model in Step 2, and indeed no samples that were accessed or examined by investigators before the computational model was locked down at the end of Step 2, be included in this independent dataset or in the evaluation of clinical utility (see Chapter 4). In the cases that the committee reviewed, two tests used overlapping training and confirmation datasets at some point in their development processes: MammaPrint and AlloMap (discussed in Chapter 6 and Appendix A).

The independent specimen and clinical data set must be relevant to the intended use of the candidate omics-based test. Specifically, patients with the same type of disease, the same stage and same clinical setting for which the candidate test is intended to be used in the future must be used for the independent confirmation of the candidate omics-based test. Ideally, the specimens for independent confirmation will have been collected at a different point in time, at different institutions, from a different patient population, with samples processed in a different laboratory to demonstrate that the test has broad applicability and is not overfit to any particular situation.

If the independent set of specimens for confirmation are collected in the same laboratory, run on the same machine, processed by the same technician, etc., then peculiarities of that data, machine, lab, etc. will be shared between the samples used to develop the computational model and the samples used to evaluate the model. As a result, even if the computational model performs well on that independent set of samples, it might not perform well on samples from patients at a different hospital, processed by a different technician in a different lab, etc.

The failure of the Ovacheck diagnostic test (further discussed in Chapter 6 and Appendix A) is illustrative of the importance of independent datasets for confirmation. Most of the tissue specimens used to confirm the initial findings were obtained from the same institution that provided the specimens used to train the computational model (Petricoin et al., 2002). In some

cases it will not be possible to obtain independent sets of specimens and associated clinical data with all of these characteristics; however, it is important to keep in mind that the quality of evidence provided by good model performance on an independent specimen set depends critically on the characteristics of that set.

Hence, it is important that full descriptions of the independent specimen set are reported along with results of the computational model's performance in the discovery phase. Here two "levels of evidence" are presented for assessing omics-based computational model performance on an independent specimen set.

*Lower Level of Evidence: Independent sets of specimens and clinical data collected at a single institution using carefully controlled protocols, with samples from the same patient population.*

In this setting, good performance of the locked down computational procedure indicates that it works in this particular setting, with these protocols, with the patient profile at this institution, etc. However, this candidate omics-based test might not work well with a slightly different patient population or with samples processed in a different laboratory or using a slightly different protocol.

*Higher Level of Evidence: Independent sets of specimens and clinical data collected at multiple institutions.*

Success in this setting strongly suggests that the omics measurements and locked down computational procedure will work well on future patients. It provides evidence that the test is robust to the kinds of things that might change between locations: namely, aspects of the biology of the populations who tend to go to a particular hospital, sample collection and handling, and measurement techniques, etc. This is important because such differences can have large effects on the omics measurements obtained, often larger than the differences associated with the phenotypes of interest.

Once Step 3 has been initiated, further refinements to the computational model are strongly discouraged as they can lead to overfitting and consequently a very high risk that the model will not perform well in subsequent steps of the development process (unless investigators plan to get yet another independent data set to redo Step 3). In the event that further model refinement occurs once Step 3 has been initiated, this must be clearly documented in describing the computational model as well as its performance in Step 3.

### Step 4: Release of Data, Code, and the Fully Specified Computational Procedures to the Scientific Community

Once an omics-based measurement method and the locked down computational procedures have been shown to perform well on independent datasets (Step 3), the candidate omics-based test is ready to proceed to the test validation phase in which analytical and clinical/biological validity are assessed, as described in detail in Chapter 3. At this time, **data and metadata used for development of the candidate omics-based test should be made available in an independently managed database (e.g., the databases of Genotypes and Phenotypes [dbGaP]) in standard format (Recommendation 1b).** dbGaP is designed to archive and distribute data from genome wide association studies to examine genetic associations with phenotypic and disease characteristics (Mailman et al., 2007; NLM, 2006).

In addition, **computer code and fully specified computational procedures used for development of the candidate omics-based test should be made sustainably available (Recommendation 1c).** For publicly funded research this means that computer code and fully specified computational procedures should be made publicly available either at the time of publication or at the end of funding. For commercially developed tests, code and fully specified computational procedures will be submitted to the Food and Drug Administration (FDA) for review if the developers are seeking FDA approval or clearance. In the case of a Laboratory Developed Test (LDT, see Chapter 3), publication is essentially required because laboratories need to justify the claims for the test. As indicated in chapter 5, the committee also recommends that **journals require authors to make data, metatdata and prespecified analysis plan, computer code, and fully specified computational procedures publicly available (Recommendation 7a[ii])**.

Ideally, the computer code that is released will encompass all of the steps of computational analysis, including all data preprocessing steps, that have been described in this chapter. All aspects of the analysis need to be transparently reported. If some aspect of the code or data cannot be made available, then the reason for this omission should be documented.

Release of data, code, and the fully specified computational procedures is important for independent verification of results. This recommendation reinforces the call for transparency in reporting made in several National Research Council reports (NRC, 2003, 2005, 2006). Others have also recently called for wider access to the full data, protocols, and computer codes for published omics studies (Ince et al., 2012; Ioannidis and Khoury, 2011), despite the known challenges to broad disclosure (Box 2-1). In the omics setting, release of this information is particularly crucial because:

1. The complex nature of the data and analyses can make it difficult to replicate results, but rigorous verification of results by the scientific community is necessary to ensure that candidate omics-based tests are scientifically and statistically valid; and
2. If data are made available, subsequent investigators can continue to conduct additional analyses to obtain future scientific insights.

**BOX 2-1**
**Considerations in Data and Information Sharing**

*Intellectual Property Protections*

Test developers may be reluctant to share their data, code, and computational procedures out of concern that others may be able to capitalize on their research. Intellectual property protections provide one mechanism for test developers to protect their proprietary information. However, intellectual property protections have serious limitations and test developers may be cautious about relying on these laws as the principal basis for their competitive advantage. For example, clinical trial results, CLIA validation, customer service, price, and FDA approval are also elements of competitive advantage.

Copyright is a type of intellectual property that protects original works of authorship, including literary, dramatic, musical, and artistic works.[a] A 1991 U.S. Supreme Court case held that a compilation of factual information can only be copyrighted if it involves originality and creativity in its selection and arrangement, and not in the compiled facts as such.[b] In most cases subsequent to that decision in the United States, the copyright in collections of data has not offered much protection from re-use by not-for-profit users, such as academics and researchers, and even by commercial competitors. However, database compilers and rights holders have used other types of legal and technical protection, such as restrictive contracts coupled with digital rights management techniques, as well as various forms of online business practices (NRC, 1999).

A patent is a type of intellectual property right that gives the owner of the patent "the right to exclude others from making, using, offering for sale, or selling the invention throughout the United States or importing the invention into the United States"[c] in exchange for disclosing the invention. The inventor must disclose enough information to enable "a person of ordinary skill in the art"[d] to achieve the same result or make the same device when following the steps provided in the patent application; however, determining the level of detail needed to meet that goal is subjective. A patent can be obtained for inventions that are novel, useful, and nonobvious to fellow scientists in the same field.[e] Various aspects of an omics-based test could potentially be patentable, including the assay used to make molecular measurements, and the code and computational procedures used to analyze the results.

However, there are limitations to the protections offered by patents and court decisions can change the landscape of patent protections. For example, a recent federal circuit court case regarding gene patents may increase the uncertainties that developers have about the value of patents for gene-based tests. The American Civil Liberties Union (ACLU) sued Myriad Genetics, the exclusive holder of licenses for the BRCA genes, challenging the practice of patenting genes.[f] The court held that it is permissible to patent isolated DNA molecules. It applied the "machine or transformation" test[g] to assess the patentability of the methods of the BRCA tests, which states that a method is patentable if: "(1) it is tied to a particular machine or apparatus, or (2) it transforms a particular article into a different state or thing." The court held that the method of comparing an individual's BRCA sequence to a reference sequence in order to identify genetic mutations were not patentable because "they claim only an abstract mental process." However, Myriad's method for screening potential cancer therapeutics via changes in cell growth rates was patentable because it required the transformative step of manipulating the cells and their growth medium. The European Patent Office reached

a similar decision, upholding Myriad's patent for the BRCA test in a limited form (Siva, 2009). Applying the holding of these cases to omics-based tests, it is unclear whether the algorithms and methods for making molecular measurements in omics-based tests are patentable. A U.S. court would likely apply the machine-or-transformation test to the exact methods and computational procedures used in each test and make a decision on a case-by-case basis. However, as a 2006 NRC report noted, there is an overall trend in the courts towards an expansive interpretation of patentable subject matter (NRC, 2006).

There are several additional limitations to relying on patent protections to guard proprietary interests in omics-based tests. The patent holder must enforce his/her exclusive right to use and market the invention. For example in the Myriad case, Myriad sent a series of cease and desist letters to laboratories engaged in the commercial testing of the BRCA genes without a license. Myriad ultimately had to litigate the validity of its patent to maintain the exclusive right to conduct these test. It also is unclear how extensively others must change or modify an omics-based test before they create a new patentable invention that they can market. In addition, the Supreme Court is likely to weigh in on the patentability of biological tests in the near future[h,i] and the US Patent and Trademark Office is preparing a report on the patentability of genetic diagnostic testing, as required by the American Invests Act;[j] both actions have the potential to change this area of law.

*Infrastructure Requirements*

The infrastructure requirements for databases that store omics data are significant and costly. The databases must allow researchers to share DNA, amino acid sequences, and protein structure data in a manageable and searchable format; these datasets can be enormous (Quackenbush, 2009). Data curation is also necessary to ensure the quality of the information stored, but approaches to curation vary considerably. The databases also need to have sufficient security protections to guard the privacy of the information stored. Despite these obstacles, however, a number genomic and proteomic databases exist, such as the databases of Genotypes and Phenotypes (dbGaP) (NCBI, 2012), the European Molecular Biology Laboratory European Bioinformatics Institute (EBI, 2012), the National Library of Medicine's Gene Expression Omnibus (NLM, 2012), Compendia Biosciences (2012), UCSC Gene Browser (UCSC, 2012), and ProteomeXchange (2012).

Data repositories have developed varying policies and procedures to protect the privacy and security of the information. For example, dbGaP is a public repository for individual-level phenotype, exposure, genotype and sequence-based data and the associations between them. In order to protect research participant privacy, all studies in dbGaP have two levels of access: open and controlled. The open-access data, which can be browsed online or downloaded from dbGaP without prior permission or authorization, generally includes all the study documents, such as the protocol and questionnaires, as well as summary data for each measured phenotype variable and for genotype results. Preauthorization is required to gain access to the phenotype and genotype results for each individual and this individual-level data is coded to protect the identity of study participants (Mailman et al., 2007; NLM, 2006).

*Privacy of Health Information*

The laws protecting the privacy of individuals' health information are a potential obstacle to making omics data sustainably available to other investigators. Much of the data in omics research is from human subjects and potentially could be linked to a

specific individual, especially in the case of genetic data. In addition, most omics data used in the development of a clinical test need to be connected to individuals' clinical data to be useful in that development process.

The *Health Insurance Portability and Accountability Act* Privacy Rule protects the privacy of personally identifiable health information (called "protected health information [PHI]") created or received by health care professionals, health plans, or healthcare clearinghouses ("covered entities"). In general, the rule requires test developers to get authorization from research subjects in order to use and disclose their PHI in health research.[k] The rule does not require researchers to get authorization to use and disclose PHI that has been de-identified (as defined in the regulation). Until recently, there was considerable confusion about whether the Privacy Rule protected genetic information (IOM, 2009). However, the *Genetic Information Nondiscrimination Act* directed the U.S. Secretary of Health and Human Services to modify the Privacy Rule to explicitly recognize genetic information as PHI.

The Common Rule provides human subjects protections in omics research that is federally funded. It protects the safety, autonomy, privacy, and fair treatment of patient-participants in federally funded research conducted on humans, and the cultural groups from which they are recruited. The Common Rule requires researchers to get informed consent from a person to use his/her private identifiable information in research. Research that involves "anonymized data" (that is, information that is recorded in such a manner that subjects cannot be identified) is exempt from this requirement. However, an advanced notice of proposed rulemaking includes the proposal to revise this aspect of the Common Rule to match the Privacy Rule's more rigorous de-identification standards.[m] If this change becomes codified in the regulations, researchers may be required in many circumstances to obtain authorization and informed consent prior to sharing their research data, in order to comply with these laws, particularly as sequence-based data can now be considered identifiable.

[a] The Copyright Act of 1976, 17 U.S.C. §§ 101-810 (2008).
[b] Feist Publications v. Rural Telephone Service Co., 499 U.S. 360 (1991).
[c] Patent Act, 35 U.S.C. § 154 (2008).
[d] Id. at § 103(a).
[e] Id. at §§ 101-103.
[f] The Association for Molecular Pathology, et al. v. United States Patent and Trademark Office, et al., 653 F.3d 1329 (Fed. Cir. 2011).
[g] Bilski vs. Kappos, 130 U.S. 3218 (2010).
[h] Mayo Collaborative Services v. Prometheus Laboratories, Inc., 628 F.3d 1347, (Fed. Cir. 2010), *cert. granted*, (U.S. Dec. 7, 2011) (No. 10-1150).
[i] The Association for Molecular Pathology, et al. v. Myriad Genetics, Inc. et al., *petition for cert. filed* (December 7, 2011).
[j] Leahy-Smith America Invents Act, Public Law no. 112-29 § 27(2011).
[k] NOTE: The Secretary of Health and Human Services issued a notice of proposed rulemaking that includes potential modifications to the HIPAA Privacy Rule's authorization requirements in response to the statutory amendments under *Health Information Technology for Economic and Clinical Health Act* (the "HITECH Act"). See, Modifications to the HIPAA Privacy, Security, and Enforcement Rules Under the Health Information Technology for Economic and Clinical Health Act, 75 Fed. Reg. 40,868 (July 14, 2010).
[l] Genetic Information Nondiscrimination Act, Public Law No. 110-233 (2008).
[m] Human Subjects Research Protections: Enhancing Protections for Research Subjects

and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44,512 (July 26, 2011).

Listed below are characteristics of candidate omics-based test concepts that pose risks for development of a spurious test. Some of these issues have already been mentioned. Others are not specific to omics-based test development, but are nonetheless extremely important for the development of such tests. Characteristics are as follows:

1. **High dimensionality of the data:** Omics datasets are typically characterized by measurements (e.g., genes or gene products such as RNA or proteins) that are many orders of magnitude greater than the number of samples for which clinical data are available. This can lead to overfitting of the data unless proper measures, such as cross-validation combined with confirmation on an independent dataset, are performed. The initial MammaPrint studies were criticized for improper cross-validation and mixing training samples with the independent dataset used for confirmation (discussed in Appendix A).
2. **Biological plausibility:** While lack of a known plausible biological mechanism should not, in itself, prevent an investigator from moving forward with candidate omics-based, such tests are more likely to endure further investigation if there is a plausible biological mechanism behind them. For instance, an effect modifier for breast cancer recurrence that is based on a set of genes known to be involved in breast cancer is more likely to hold up in further study than an effect modifier based on some set of genes not already known to be implicated in the disease. Better results can often be obtained by beginning the omics-based test development process using a subset of the omics measurements for which a plausible biological mechanism is available. For instance, there was a plausible biological mechanism behind the HER2 tests and Oncotype DX to motivate their initial clinical trials, but less so for the Duke, MammaPrint, and Ova1 tests (discussed in Appendix A and B). Bioinformatics methods to link transcript or protein expression changes to relevant signaling pathways or biological networks need to be deployed appropriately.
3. **Data variability unrelated to clinical outcome of interest:** Often, a computational model developed on one dataset (Step 2) performs poorly on another independent dataset (Step 3). This can occur for a number of reasons, such as variability in patient population, sample preparation, time of sample collection, operator variability, etc. Hence, evidence of a computational model's performance based only on the dataset used to train the model, even if cross-validation is properly performed, provides little evidence of the model's suitability for future samples. A relevant example here is the Ovacheck case study, discussed in Appendix A, in which signals obtained on one dataset did not hold up when the test was applied to other independent sample sets.
4. **Need for multiple datasets:** For the reasons just described, computational models that are fit on multiple datasets in Step 2 will tend to perform better later. In other words, investigators are urged to develop a computational model on omics datasets derived from specimens and associated clinical outcomes collected at multiple laboratories at multiple institutions, rather than fitting a model on just a single dataset. For instance, the 21-Gene Recurrence Score (Oncotype DX) case study (Appendix A) was developed using multiple independent datasets (Paik et al., 2004). In that case data were analyzed by the same investigators, but different data sets were derived from different trials at multiple institutions.

5. **Study design and batch effects:** As in all areas of biomedical research, good study design is crucial. If the dataset used in Step 2 to develop the computational model resulted from poor experimental design (e.g., if the samples from patients whose cancers recurred were processed at a different time or by a different technician or in a different laboratory) then *batch effects* (Leek et al., 2010) can occur. This will lead to spurious signal, potentially resulting in a computational model that performs extremely well on the data on which it was developed (Step 2), but that will perform poorly on future patient samples (Step 3). A relevant example is the Ovacheck case study, discussed in Appendix A, in which changes to protocol mid-experiment and improper instrument calibration rendered an assay with results that were irreproducible across settings due to batch effects (Baggerly et al., 2004).
6. **Computational procedure lock down:** It is crucial that at the end of Step 2, the fully specified computational procedures be locked down before progressing into confirmation on an independent test set in Step 3. For instance, simply reporting the set of genes included in the computational model underlying a transcriptomics-based test is insufficient, as this does not constitute a fully specified computational procedure. In the original OncotypeDX study, the researchers locked down the computational model after Step 2 and reported the fully specified model in the paper (Paik et al., 2004). In the Corus CAD case study, lock down and the fully specified computational model were reported in the clinical validation paper (Rosenberg et al., 2010). The fully specified computational model for the AlloMap test is reported in Deng et al. (2006). In contrast, in the Duke studies, the genes used in the development of the computational model were reported, but the fully specified computational procedure was not; furthermore, it is likely that the computational procedures were not ever fully locked down before proceeding into Step 3 or further stages of omics-based test development, including clinical trials (see Appendix B for details).
7. **Role of biostatistics and bioinformatics experts:** In a relatively new and evolving field such as omics, it is not possible to predict all the possible pitfalls that investigators may face in the discovery phase. The involvement of properly trained biostatistical or bioinformatics collaborators who are fully integrated in all aspects of the discovery and evaluation process can serve as an additional safeguard. The type of biostatistical expertise that is required may vary depending on the stage or phase of test development. For example, experts in developing computational models for omics-based tests may not have sufficient expertise in clinical trial design, and vice versa. This is relevant to the Duke case study (as discussed in Appendix B, page 35), in which there was a lack of continuity in biostatistics personnel and numerous errors were identified in the statistical methodology and analyses.

## COMPLETION OF THE DISCOVERY PHASE OF OMICS-BASED TEST DEVELOPMENT

**A candidate omics-based test should be defined precisely**, **including the molecular measurements, the computational procedures, and the intended clinical use of the test, in anticipation of the test validation phase (Recommendation 1d).** There are enormous opportunities represented by the rapidly improving suite of omics technologies for measurements with potential clinical utility. However, there are significant challenges in moving from the identification of differences in omics measurements from clinical samples to validated and robust

clinical tests. Among these challenges are risks of overfitting the data in the development of the computational model and the enormous heterogeneity between different studies of ostensibly the same disease states (for both technical and biological reasons). Going forward, transparency in the reporting of all aspects of the development of an omics-based test, including the measurements made, preprocessing techniques used, and the fully specified computational procedure, are critical. The release of sufficient metadata with publication is also key to the identification of candidate omics-based tests that work across multiple sites, as needed to generate increasingly robust omics-based tests to enhance patient care.

In the next phase of test development (analytical and clinical/biological validation, described in Chapter 3), the methods used to obtain the omics measurements from patient samples may be changed in order to establish a clinically feasible, inexpensive, and robust assay for implementation in clinical practice. However, the fully specified computational procedures defined in the discovery stage must remain locked down and unchanged in all subsequent test development steps. At the end of the validation phase in chapter 3, the complete test method, including the methods for obtaining the omics measurements as well as the computational procedures, must be locked down before crossing the bright line to evaluate the test for clinical utility and use.

## SUMMARY AND RECOMMENDATION

This chapter has outlined best practices for the discovery phase for omics-based test development. Because omics-based tests rely on interpretation of high-dimensional datasets, it is important to guard against overfitting the data throughout the test development process. Overfitting due to lack of proper statistical methods can lead to a model that fits the training samples well, even though the model might perform poorly on independent samples not used in test development. The steps delineated in this chapter aim to prevent an overfit model from progressing to subsequent stages of test development. Cross-validation or a training set/test set approach can help reduce the risk of overfitting, but confirmation of all fully specified computational procedures and omics-based tests on a blinded independent sample set is the "gold standard" for assessing the validity of any test. The importance of independent confirmation is also emphasized in the committee's recommendations for funders (see Chapter 5), which urge funders to support this type of work. In addition, complex analyses of these large datasets highlight the need for availability of the data and code used for the discovery phase of omics-based test development. The result of the discovery process is a candidate omics-based test with locked down computational procedures that is then moved into the test validation phase to assess analytical and clinical/biological validation, as described in Chapter 3.

**RECOMMENDATION 1:**
**When candidate omics-based tests from the discovery phase are intended for further clinical development, the following criteria should be satisfied and fully disclosed (for example, through publication or patent application) to enable independent verification of the findings:**

**a. Candidate omics-based tests should be confirmed using an independent set of samples, not used in generation of the computational model and, when feasible, blinded to any outcome or other phenotypic data until after the computational**

**procedures have been locked down and the candidate omics-based test has been applied to the samples;**

**b. Data and metadata used for development of the candidate omics-based test should be made available in an independently managed database (such as dbGaP) in standard format;**

**c. Computer code and fully specified computational procedures used for development of the candidate omics-based test should be made sustainably available; and**

**d. A candidate omics-based test should be defined precisely, including the molecular measurements, the computational procedures, and the intended clinical use of the test, in anticipation of the test validation phase.**

## REFERENCES

Arrell, D. K., J. Zlatkovic Lindor, S. Yamada, and A. Terzic. 2011. K(ATP) channel-dependent metaboproteome decoded: Systems approaches to heart failure prediction, diagnosis, and therapy. *Cardiovascular Research* 90(2):258-266.

Ayoglu, B., A. Haggmark, M. Neiman, U. Igel, M. Uhlen, J. M. Schwenk, and P. Nilsson. 2011. Systematic antibody and antigen-based proteomic profiling with microarrays. *Expert Rev Mol Diagn* 11(2):219-234.

Baggerly, K. A., J. S. Morris, and K. R. Coombes. 2004. Reproducibility of SELDI-TOF protein patterns in serum: Comparing datasets from different experiments. *Bioinformatics* 20(5):777-785.

Bailey, R. C., G. A. Kwong, C. G. Radu, O. N. Witte, and J. R. Heath. 2007. DNA-encoded antibody libraries: A unified platform for multiplexed cell sorting and detection of genes and proteins. *Journal of the American Chemical Society* 129(7):1959-1967.

Ben Sellem, D., K. Elbayed, A. Neuville, F. M. Moussallieh, G. Lang-Averous, M. Piotto, J. P. Bellocq, and I. J. Namer. 2011. Metabolomic characterization of ovarian epithelial carcinomas by HRMAS-NMR spectroscopy. *J Oncol* 2011:174019.

Bianco-Miotto, T., K. Chiam, G. Buchanan, S. Jindal, T. K. Day, M. Thomas, M. A. Pickering, M. A. O'Loughlin, N. K. Ryan, W. A. Raymond, L. G. Horvath, J. G. Kench, P. D. Stricker, V. R. Marshall, R. L. Sutherland, S. M. Henshall, W. L. Gerald, H. I. Scher, G. P. Risbridger, J. A. Clements, L. M. Butler, W. D. Tilley, D. J. Horsfall, and C. Ricciardelli. 2010. Global levels of specific histone modifications and an epigenetic gene signature predict prostate cancer progression and development. *Cancer Epidemiol Biomarkers Prev* 19(10):2611-2622.

Chang, R. L., L. Xie, P. E. Bourne, and B. O. Palsson. 2010. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput Biol* 6(9):e1000938.

Compendia Bioscience, Inc. 2012. Compendia bioscience: Cure cancer with genomic data. http://www.compendiabio.com/ (accessed February 23, 2012).

Deng, M. C., H. J. Eisen, M. R. Mehra, M. Billingham, C. C. Marboe, G. Berry, J. Kobashigawa, F. L. Johnson, R. C. Starling, S. Murali, D. F. Pauly, H. Baron, J. G. Wohlgemuth, R. N. Woodward, T. M. Klingler, D. Walther, P. G. Lal, S. Rosenberg, S. Hunt, and for the CARGO Investigators. 2006. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *American Journal of Transplantation* 6(1):150-160.

Drexler, D. M., M. D. Reily, and P. A. Shipkova. 2011. Advances in mass spectrometry applied to pharmaceutical metabolomics. *Analytical and Bioanalytical Chemistry.* 399(8):2645-2653.

EBI (European Bioinformatics Institute). 2012. *Data Resources and Tools*. http://www.ebi.ac.uk/ (accessed February 23, 2012).

Fagerberg, L., S. Stromberg, A. El-Obeid, M. Gry, K. Nilsson, M. Uhlen, F. Ponten, and A. Asplund. 2011. Large-scale protein profiling in human cell lines using antibody-based proteomics. *J Proteome Res* 10(9):4066-4075.

Farrah, T., E.W. Deutsch, G.S. Omenn, D.S. Campbell, Z. Sun, J.A. Bletz, P. Mallick, J.E. Katz, J. Malmström, R. Ossola, J.D. Watts, B. Lin, H. Zhang, R.L. Moritz, R. Aebersold. 2011. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Molecular and Cellular Proteomics*. 10(9):M110.006353.

Folger, O., L. Jerby, C. Frezza, E. Gottlieb, E. Ruppin, and T. Shlomi. 2011. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol* 7:501-527.

Frezza, C., L. Zheng, O. Folger, K. N. Rajagopalan, E. D. Mackenzie, L. Jerby, M. Micaroni, B. Chaneton, J. Adam, A. Hedley, G. Kalna, I. P. Tomlinson, P. J. Pollard, D. G. Watson, R. J. Deberardinis, T. Shlomi, E. Ruppin, and E. Gottlieb. 2011. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*. 477(7363):225-228.

Gold, L., D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T. Fitzwater, D. Flather, A. Forbes, T. Foreman, C. Fowler, B. Gawande, M. Goss, M. Gunn, S. Gupta, D. Halladay, J. Heil, J. Heilig, B. Hicke, G. Husar, N. Janjic, T. Jarvis, S. Jennings, E. Katilius, T. R. Keeney, N. Kim, T. H. Koch, S. Kraemer, L. Kroiss, N. Le, D. Levine, W. Lindsey, B. Lollo, W. Mayfield, M. Mehan, R. Mehler, S. K. Nelson, M. Nelson, D. Nieuwlandt, M. Nikrad, U. Ochsner, R. M. Ostroff, M. Otis, T. Parker, S. Pietrasiewicz, D. I. Resnicow, J. Rohloff, G. Sanders, S. Sattin, D. Schneider, B. Singer, M. Stanton, A. Sterkel, A. Stewart, S. Stratford, J. D. Vaught, M. Vrkljan, J. J. Walker, M. Watrobka, S. Waugh, A. Weiss, S. K. Wilcox, A. Wolfson, S. K. Wolk, C. Zhang, and D. Zichi. 2010. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 5(12):e15004.

Gottlieb, A., G. Y. Stein, E. Ruppin, and R. Sharan. 2011. PREDICT: A method for inferring novel drug indications with application to personalized medicine. *Molecular Systems Biology* 7:496.

Honda, K. and D.R. Littman. 2011. The microbiome in infectious disease and inflammation. *Annual Review of Immunology*. 2011 Mar 24. [Epub ahead of print].

Hwang, D., I. Y. Lee, H. Yoo, N. Gehlenborg, J. H. Cho, B. Petritis, D. Baxter, R. Pitstick, R. Young, D. Spicer, N. D. Price, J. G. Hohmann, S. J. Dearmond, G. A. Carlson, and L. E. Hood. 2009. A systems approach to prion disease. *Molecular Systems Biology* 5:252.

Ideker, T., J. Dutkowski, and L. Hood. 2011. Boosting signal-to-noise in complex biology: Prior knowledge is power. *Cell* 144(6):860-863.

Ioannidis, J. P. A., and M. J. Khoury. 2011. Improving validation practices in "omics" research. *Science* 334(6060):1230-1232.

IOM (Institute of Medicine). 2009. *Beyond the HIPAA Privacy Rule: Enhancing privacy, improving health through research*. Washington, DC: The National Academies Press.

Ince, D. C., L. Hatton, and J. Graham-Cumming. 2012. The case for open computer programs. *Nature* 482:485-488.

Kinros, J. M., A. W. Darzi, and J. K. Nicholson. 2011. Gut microbiome-host interactions in health and disease. *Genomic Medicine* 3(3):14.

Koulman, A., G. A. Lane, S. J. Harrison, and D. A. Volmer. 2009. From differentiating metabolites to biomarkers. *Analytical and Bioanalytical Chemistry* 394(3):663-670.

Leek, J. T., R. B. Scharpf, H. C. Bravo, D. Simcha, B. Langmead, W. E. Johnson, D. Geman, K. Baggerly, and R. A. Irizarry. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 11(10):733-739.

Legrain, P., R. Aebersold, A. Archakov, A. Bairoch, K. Bala, L. Beretta, J. Bergeron, C. H. Borchers, G. L. Corthals, C. E. Costello, E. W. Deutsch, B. Domon, W. Hancock, F. He, D. Hochstrasser, G. Marko-Varga, G. H. Salekdeh, S. Sechi, M. Snyder, S. Srivastava, M. Uhlen, C. H. Wu, T. Yamamoto, Y. K. Paik, and G. S. Omenn. 2011. The human proteome project: Current state and future direction. *Mol Cell Proteomics* 10(7):M111. 009993.

Lewis, G. D., and R. E. Gerszten. 2010. Toward metabolomic signatures of cardiovascular disease. *Circ Cardiovasc Genet* 3(2):119-121.

Lewis, N. E., G. Schramm, A. Bordbar, J. Schellenberger, M. P. Andersen, J. K. Cheng, N. Patel, A. Yee, R. A. Lewis, R. Eils, R. Konig, and B. O. Palsson. 2010. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat Biotechnol* 28(12):1279-1285.

Ma, C., R. Fan, H. Ahmad, Q. Shi, B. Comin-Anduix, T. Chodon, R. C. Koya, C. C. Liu, G. A. Kwong, C. G. Radu, A. Ribas, and J. R. Heath. 2011. A clinical microchip for evaluation of single immune cells reveals high functional heterogeneity in phenotypically similar T cells. *Nat Med* 17(6):738-743.

Mailman M.D., M. Feolo, Y. Jin, M. Kimura, K. Tryka, R. Bagoutdinov, L. Hao, A. Kiang, J. Paschall, L. Phan, N. Popova, S. Pretel, L. Ziyabari, M. Lee, Y. Shao, Z. Y. Wang, K. Sirotkin, M. Ward, M. Kholodov, K. Zbicz, J. Beck, M. Kimelman, S. Shevelev, D. Preuss, E. Yaschenko, A. Graeff, J. Ostell, and & S. T. Sherry. 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nature Genetics* 39:1181–1186.

Masoodi, M., M. Eiden, A. Koulman, D. Spaner, and D. A. Volmer. 2010. Comprehensive lipidomics analysis of bioactive lipids in complex regulatory networks. *Analyticl Chemistry* 82(19):8176-8185.

Mass, S., 2010. Gene regulation through RNA editing. Discovery Medicine, 10(54):379-386.

McGary, K. L., T. J. Park, J. O. Woods, H. J. Cha, J. B. Wallingford, and E. M. Marcotte. 2010. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci U S A* 107(14):6544-6549.

McShane, L. M. 2010. *NCI Address to Institute of Medicine Committee Convened to Review Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials*. Presentation at Meeting 1: Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, December 20.

Menon, R., A. Roy, S. Mukerjee, S. Belkin, Y. Zhang, and G. S. Omenn. 2011. Functional implications of structural predictions for alternative splice proteins expressed in HER2/ neu-induced breast cancers. *Journal of Proteome Research.* [Epub ahead of print].

Moussay, E., K. Wang, J. H. Cho, K. van Moer, S. Pierson, J. Paggetti, P. V. Nazarov, V. Palissot, L. E. Hood, G. Berchem, and D. J. Galas. 2011. MicroRNA as biomarkers and regulators in B-cell chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 108(16):6573-6578.

NCBI (National Center for Biotechnology Information). 2012. dbGaP. http://www.ncbi.nlm.nih.gov/gap (accessed February 23, 2012).

Ng, S. B., A.W. Bigham , K.J. Buckingham , M. C Hannibal , M. J. McMillin, H. I. Gildersleeve, A. E. Beck , H. K. Tabor, G. M. Cooper, H. C. Mefford, C. Lee, E. H. Turner , J. D. Smith , M. J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D. A. Nickerson, M. J. Bamshad , J. Shendure. 2010. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genetics* 42(9):790-793.

NLM (National Library of Medicine). 2012. GEO: Gene expression omnibus. http://www.ncbi.nlm.nih.gov/geo/ (accessed February 23, 2012).

NLM. 2006. NIH launches dbGAP, a database of Genome Wide Association Studies. http://www.nlm.nih.gov/news/press_releases/dbgap_launchPR06.html (accessed December 12, 2006).

NRC (National Research Council). 1999. *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Washington, DC: The National Academies Press.

NRC. 2003. *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. Washington, DC: The National Academies Press.

NRC. 2005. *Catalyzing Inquiry at the Interface of Computing and Biology*. Washington, DC: The National Academies Press.

NRC. 2006. *Reaping the Benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health*. Washington, DC: The National Academies Press.

Omenn, G. S. 2009. A landmark systems analysis of prion disease of the brain. *Molecular Systems Biology* 5:254.

Omenn, G. S., M. S. Baker, and R. Aebersold. 2011. Recent workshops of the HUPO Human Plasma Proteome Project (HPPP): A bridge with the HUPO CardioVascular Initiative and the emergence of SRM targeted proteomics. *Proteomics* 11(17):3439-3443.

Ostroff, R. M., W. L. Bigbee, W. Franklin, L. Gold, M. Mehan, Y. E. Miller, H. I. Pass, W. N. Rom, J. M. Siegfried, A. Stewart, J. J. Walker, J. L. Weissfeld, S. Williams, D. Zichi, and E. N. Brody. 2010. Unlocking biomarker discovery: Large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS One* 5(12):e15003.

Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, and N. Wolmark. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine* 351(27):2817-2826.

Petricoin, E. F., A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta. 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359(9306):572-577.

Picotti, P., O. Rinner, R. Stallmach, F. Dautel, T. Farrah, B. Domon, H. Wenschuh, and R. Aebersold. 2010. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* 7(1):43-46.

ProteomeXchange. 2012. *Mission*. http://www.proteomexchange.org/ (accessed February 23, 2012).

Ransohoff, D. F. 2008. The process to discover and develop biomarkers for cancer: A work in progress. *Journal of the National Cancer Institute*. 100(20):1419-1420.

Ransohoff, D. F. 2009. Promises and limitations of biomarkers. *Recent Results in Cancer Research* 181:55-59.

Roach, J. C., G. Glusman, A. F. Smit, C. D. Huff, R. Hubley, P. T. Shannon, L. Rowen, K. P. Pant, N. Goodman, M. Bamshad, J. Shendure, R. Drmanac, L. B. Jorde, L. Hood, and D. J. Galas. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328(5978):636-639.

Rosenberg, S., M. R. Elashoff, P. Beineke, S. E. Daniels, J. A. Wingrove, W. G. Tingley, P. T. Sager, A. J. Sehnert, M. Yau, W. E. Kraus, K. Newby, R. S. Schwartz, S. Voros, S. G. Ellis, N. Tahirkhelli, R. Waksman, J. McPherson, A. Lansky, M. E. Winn, N. J. Schork, E. J. Topol, and for the PREDICT (Personalized Risk Evaluation and Diagnosis In the Coronary Tree) Investigators. 2010. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Annals of Internal Medicine* 153(7):425-434.

Segre, A. V., L. Groop, V. K. Mootha, M. J. Daly, and D. Altshuler. 2010. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 6(8). pii: e1001058.

Seppanen-Laakso, T., and M. Oresic. 2009. How to study lipidomes. *J Mol Endocrinol* 42(3):185-190.

Service, R. F. 2008. Chemistry. Click chemistry clicks along. *Science* 320(5878):868-869.

Shlomi, T., T. Benyamini, E. Gottlieb, R. Sharan, and E. Ruppin. 2011. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. *PLoS Comput Biol* 7(3):e1002018.

Simon, R., M. D. Radmacher, K. Dobbin, and L. M. McShane. 2003. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute* 95(1):14-18.

Siva, N. 2009. Myriad wins BRCA1 row. *Nature Biotechnology* 27:8.

Sreekumar, A., L. M. Poisson, T. M. Rajendiran, A. P. Khan, Q. Cao, J. Yu, B. Laxman, R. Mehra, R. J. Lonigro, Y. Li, M. K. Nyati, A. Ahsan, S. Kalyana-Sundaram, B. Han, X. Cao, J. Byun, G. S. Omenn, D. Ghosh, S. Pennathur, D. C. Alexander, A. Berger, J. R. Shuster, J. T. Wei, S. Varambally, C. Beecher, and A. M. Chinnaiyan. 2009. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457(7231):910-914.

Sugatani, T., J. Vacher, and K. A. Hruska. 2011. A microRNA expression signature of osteoclastogenesis. *Blood* 117(13):3648-3657.

Tan, X., W. Qin, L. Zhang, J. Hang, B. Li, C. Zhang, J. Wan, F. Zhou, K. Shao, Y. Sun, J. Wu, X. Zhang, B. Qiu, N. Li, S. Shi, X. Feng, S. Zhao, Z. Wang, X. Zhao, Z. Chen, K. Mitchelson, J. Cheng, Y. Guo, and J. He. 2011. A five-microRNA signature for squamous cell lung carcinoma (SCC) diagnosis and Hsa-miR-31 for SCC prognosis. *Clinical Cancer Research*. 17(21):6802-6811.

Tang, F., K. Lao, and M. A. Surani. 2011. Development and applications of single-cell transcriptome analysis. *Nature Methods* 8(4 Suppl):S6-S11.

Tilg, H. and A. Kaser. 2011. Gut microbiome, obesity, and metabolic dysfunction. *Journal of Clinical Investigations* 121(6):2126-2132.

Teague, B., M. S. Waterman, S. Goldstein, K. Potamousis, S. Zhou, S. Reslewic, D. Sarkar, A. Valouev, C. Churas, J. M. Kidd, S. Kohn, R. Runnheim, C. Lamers, D. Forrest, M. A. Newton, E. E. Eichler, M. Kent-First, U. Surti, M. Livny, and D. C. Schwartz. 2010. High-resolution human genome structure by single-molecule analysis. *Proc Natl Acad Sci U S A* 107(24):10848-10853.

UCSC (University of California Santa Cruz). 2012. *UCSC genome bioinformatics.* http://genome.ucsc.edu/ (accessed February 23, 2012).

Weckwerth, W. 2003. Metabolomics in systems biology. *Annu Rev Plant Biol* 54:669-689.

Wolf-Yadlin, A., M. Sevecka, and G. MacBeath. 2009. Dissecting protein function and signaling using protein microarrays. *Current Opinion in Chemical Biology* 13(4):398-405.

Yu, S. L., H. Y. Chen, G. C. Chang, C. Y. Chen, H. W. Chen, S. Singh, C. L. Cheng, C. J. Yu, Y. C. Lee, H. S. Chen, T. J. Su, C. C. Chiang, H. N. Li, Q. S. Hong, H. Y. Su, C. C. Chen, W. J. Chen, C. C. Liu, W. K. Chan, W. J. Chen, K. C. Li, J. J. Chen, and P. C. Yang. 2008. MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* 13(1):48-57.

Zhang, G. F., S. Sadhukhan, G. P. Tochtrop, and H. Brunengraber. 2011. Metabolomics, pathway regulation, and pathway discovery. *Journal of Biological Chemistry* 286(27):23631-23635.

# 3
# Best Practices for Omics-Based Test Validation Prior to Use for Patient Management Decisions in a Clinical Trial Setting

This chapter explains what steps are recommended after discovery and confirmation of a candidate omics-based test (as described in Chapter 2) and before conducting studies to assess the test for clinical utility and use (as described in Chapter 4). Figure 3-1 highlights omics-based test validation, the second component of the committee's recommended omics-based test development and evaluation process. **The committee defines an omics-based test as consisting of both the data-generating assay and the fully specified computational procedures used for analysis of the assay data. The committee recommends that both components of omics-based tests used to direct patient management in a clinical trial setting should be validated during the test validation phase using several steps (Recommendation 2).** The optimal validation process for translation of the candidate omics-based test from the discovery phase (see Recommendation 1, Chapter 2) into a validated test ready for use in a clinical trial setting is the same validation process used for new tests performed in clinical laboratories.

This chapter references the current standards used for regulation of clinical laboratories and clinical tests, specifically the Clinical Laboratory Standards defined under *Clinical Laboratory Improvement Amendments of 1988* (CLIA) and the Food and Drug Administration (FDA) regulatory standards for devices. Issues are noted for both of these standards. Under CLIA, which was designed, enacted, and implemented from 1988 to 1992, with updates over the years, the current regulations do not specify the mechanisms for validation of a test in a clinical laboratory (referred to as laboratory-developed tests or LDTs) and do not have specific requirements for omics-based tests. However, CLIA is a minimal standard for the operation of a clinical laboratory, which is more stringent and assures a baseline level of quality assurance compared to a research laboratory setting. A report by the Secretary's Advisory Committee on Genetics, Health, and Society in 2008 defines the issues with the current CLIA standards and oversight of clinical laboratories and makes recommendations for improving the oversight of clinical laboratories. The second standard referenced in this chapter is oversight by the FDA. The FDA has not enforced its authority for oversight of LDTs, which is one pathway for the translation of an omics discovery into a clinical test, and the FDA has not yet defined its regulatory approach to oversight of omics-based tests, especially when developed as an LDT. Implementation of the committee's recommendation could lead to an increased workload for the FDA, with an impact on the finite resources of the FDA. This issue is not addressed by this report. Finally, the FDA's responsibility for assuring the safety and effectiveness of any medical device, including omics-based tests, does not require demonstration of clinical utility, which is a standard required by payers. However, the FDA is the federal regulatory agency with oversight of testing and should be seen as a partner in the test validation phase. Therefore, the committee uses the CLIA and FDA regulatory standards as a minimum for assuring the quality of a test prior to use in a clinical trial to direct patient management.

In addition to federal oversight of clinical laboratories and clinical tests, several professional societies also develop guidelines and best practices for clinical laboratories. The College of American Pathologists (CAP) has laboratory accreditation standards that meet and exceed the CLIA standards, sets practice standards, and provides proficiency testing samples for molecular pathology clinical laboratories and other types of clinical laboratories. A second professional society that works to establish best practices for molecular laboratories is the Association for Molecular Pathology (founded in the 1995). A third society is the American College of Medical Genetics, which also develops and publishes guidelines and best practices in molecular genetic testing.

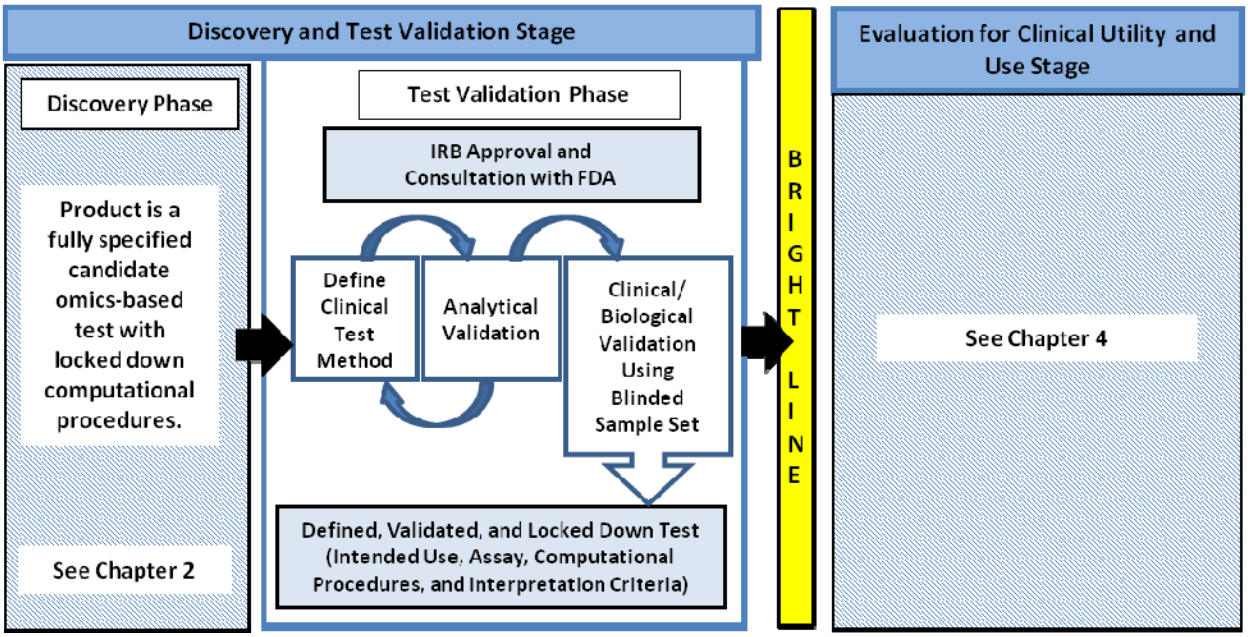


**FIGURE 3-1** Omics-based test development process, highlighting the test validation phase. In the first stage of omics-based test development, two phases take place: discovery and test validation. In the test validation phase, the omics-based test undergoes analytical and clinical/biological validation. Statistics and bioinformatics validation occurs throughout the discovery and test validation stage as well as the stage for evaluation of clinical utility and use. The bright line signifies the point in test development where a fully defined validated, and locked down clinical test method is necessary. Changes to the test after the bright line is crossed require a return to the test validation phase, approval by the Institutional Review Board (IRB), and possibly consultation with the Food and Drug Administration (FDA).

Two primary routes are used to develop omics-based tests intended for clinical use. The first is partnership with industry to develop the test for commercial production of test kits with clearance or approval as a device through the FDA or proprietary performance of the test by the company within a specific CLIA-certified clinical laboratory. The second is partnership with a commercial or local academic CLIA-certified laboratory to validate the discovery for performance as a laboratory-developed test (LDT) in the CLIA-certified laboratory. The committee's recommendations apply to both pathways. Regardless of the route pursued, however, the committee recommends that **the candidate omics-based test and its intended use should be discussed with the FDA prior to initiation of validation studies (Recommendation 2a).** As discussed in Chapter 4, if a test will be used to direct patient management in a clinical

trial, then an Investigational Device Exemption[1] (IDE) application must be filed with the FDA (Gutierrez, 2011) (see Recommendation 3b), the test should be validated in a CLIA-certified clinical laboratory, and the clinical trial that uses the test should be approved by an Institutional Review Board. If the clinical trial will not use the results of the omics-based test to direct patient management, then an IDE is not required and the testing could be performed in a research laboratory setting. However, if the ultimate goal of the clinical study of the new test is to produce evidence to support use of the omics-based test in medical care, then consulting regulatory authorities, such as the FDA, for guidance in the appropriate validation and clinical performance of the omics-based test is recommended to ensure complete assessment of the performance of the test before evaluation for clinical use. It is never too early to confer with the FDA about the omics-based test development process, and the agreements at a pre-IDE meeting may be crucial to the development process because they will provide a degree of certainty about subsequent test development steps and data collection.

An abbreviated description of the process for test validation in a clinical laboratory is provided in this chapter and serves as the committee's recommended best practices for translation of the candidate omics-based test into a test suitable for use in a clinical trial. Test validation includes steps to confirm the performance characteristics of the assay method, the fully specified and locked down computational model from the discovery phase (defined in chapter 2 as the series of computational steps performed in processing the raw data, as well as the mathematical formula or formulas used to convert the data into a prediction of the phenotype of interest), and all assay data handling methods required to reach a final test result.

## BACKGROUND

Translation of a candidate omics-based test from the discovery process into a clinical test requires confirmation of the research findings using a defined assay and fully specified computational model performed in the setting of a clinical laboratory certified under CLIA through the Centers for Medicare & Medicaid Services (CMS) or an accreditation organization with deeming authority under CLIA. This step is necessary whether the test will be developed into an in vitro diagnostic test kit with approval or clearance by the FDA, or performed as a LDT within a CLIA-certified laboratory. The CLIA regulations were established in 1988 to ensure that minimum personnel and quality requirements were followed in all clinical laboratories. Today, compliance with CLIA regulations is required for any laboratory performing testing on human specimens for clinical care, especially when seeking payment under the Medicare and Medicaid programs (CDC, 2004). If the laboratory does not report patient-specific results for the diagnosis, prevention, or treatment of a disease or for the assessment of the health of individual patients, then the laboratory does not need to comply with CLIA requirements when testing human specimens. However, if any test result, including an omics-based test result, is used for patient care, even in the setting of a clinical trial, the test is required to be performed in a CLIA-certified laboratory and must meet all the regulatory requirements for performance of a clinical test under CLIA regulations. These regulations state that a clinical test must be validated to define the test method and the performance characteristics of the test, and then the validated test must be performed and regularly assessed using specific quality standards. Validation of the omics-based test in a CLIA-certified laboratory setting before use for patient management decisions in a

[1] An FDA designation that allows an investigational device to be used in a clinical study to collect safety and effectiveness data supporting a pre-market approval application or a pre-market notification submission.

clinical trial is needed to ensure that clinical quality standards are used for the test validation and performance to minimize the potential for patient harm.

An assessment for clinical utility and use of an omics-based test can be accomplished through clinical trial designs that do not use the test results to direct patient management, which would allow the test to be performed in a research laboratory setting (see Chapter 4, Evaluation for Clinical Utility and Use Stage). However, if the goal is to generate data to support the clinical use of the omics-based test, performance of the test in compliance with CLIA regulations in a CLIA-certified laboratory would ensure proper definition of the test and its performance characteristics, and facilitate later transition of the test into clinical use following completion of the clinical trial. Therefore, the best practice for use of an omics-based test in a clinical trial setting, regardless of the trial design, is performance of the test in compliance with CLIA regulations in a CLIA-certified clinical laboratory.

This chapter will briefly describe the process of test validation and performance under CLIA regulations, with a focus on important overarching concepts underpinning the validation steps; the specific details of the process will vary considerably depending on the test method and the intended use. There are detailed guidelines, guidance documents, and published reviews providing extensive detail on validation in various scenarios, as cited in this chapter. Description of the CLIA-certified laboratory processes for test validation and performance will assist investigators at the discovery phase and clinical trial phase of omics-based test development research to understand the requirements for clinical test performance in preparation for assessment prior to use of the new omics-based test in a clinical trial setting (Chapter 4).

## TEST VALIDATION PHASE

### Considerations Prior to Analytical and Clinical/Biological Validation

**The committee recommends that test validation should be performed in a CLIA-certified clinical laboratory, beginning with a defined candidate omics-based test from the discovery phase (Recommendation 2b).** Ultimately, the goal of the investigator is to generate high quality evidence that warrants assessment of the test for its intended clinical use. Therefore, the candidate omics-based test from the discovery phase should be converted into an analytically validated and clinically practical test. Once an omics-based discovery is defined and confirmed using an independent, blinded set of samples in the research setting (see Chapter 2), investigators can work with the director and staff of a CLIA-certified laboratory to transition the test to the clinical laboratory setting in preparation for studies to evaluate the test for clinical use. The investigators should be able to provide the clinical laboratory with the defined assay method and the fully specified and locked down computational procedures used to interpret the omics measurements in the research setting. In addition, the investigators should work with the clinical laboratory director and staff to define the intended use and performance requirements of the test, including the purpose of the test (e.g., risk assessment, screening, diagnosis, prognosis, predicting response to therapy, or monitoring); the patient population to be tested (e.g., stage II colon cancer patients, low birth weight infants less than 6 months of age, etc.); the specimen type (e.g., blood, urine, fresh tumor tissue, paraffin-embedded formalin-fixed tumor tissue, etc.); the specimen handling requirements (e.g., immediately frozen tissue for RNA-based testing, separation of serum from blood cells within 4 hours of collection for viral load testing, etc.); required or ideal test performance characteristics (e.g., sensitivity, accuracy, specificity, etc.);

and the required turnaround time for test results based on clinical decision requirements (e.g., within 1 hour of specimen collection, same day, within one week).

Specimens comparable to those on which the omics-based test will eventually be applied in the clinic should be identified for the analytical and clinical/biological validation of the test. Availability of appropriate specimens is essential to the success of the validation process. Obtaining appropriate specimens can take a significant amount of time and can delay the validation process if not initiated early. Availability of appropriate specimens is essential for the successful analytical and clinical/biological validation of a test; strategies for sharing of available specimen sets are essential because this is a barrier identified by test developers. Specimens with defined characteristics that provide a range of test results (e.g., positive and negative results, or a distribution across a quantitative range of results, etc.) are needed for the analytical validation of the test. This validation requires a large volume of a small number of specimens representing the range of expected test results. The large volume of these specimens is needed for the repeated testing of the same specimens as the test method is optimized and defined, and for repeated testing to establish the technical variability of the test when performed on specimens handled in different ways prior to testing, and as the test is performed at different times, by different technologists, or on different instruments. Although clinical specimens may be optimal for analytical validation, in some cases commercial sources of control analytes may be spiked into negative specimens of the same type that will eventually be used for the test.

A set of clinical or biological specimens in sufficient numbers to provide statistical assessment of the test's performance characteristics is needed for the clinical/biological validation of the test. While a specific number of specimens cannot be generically defined for all tests, the New York State Department of Health has defined a minimum of 30 positive and 10 negative samples of each specimen type plus all controls used for infectious disease tests using amplification methods (Wadsworth Center, 2009). Clearly, the correct number of specimens depends on the complexity of the test itself, the range of expected results, the range of specimen types, and other factors. Specimens should come from patient populations similar to those on which the test will be used clinically, and should be available in sufficient volume to allow for several repetitions of the test. Sources of specimens include biobank specimens with related clinical information, residual specimens in clinical laboratories, or commercially available cell lines or pathogens with known characteristics that could be spiked into the relevant clinical specimen type. Ideally, a new set of specimens not used in the discovery phase should be used for the analytical and/or clinical/biological validation of the test in the CLIA-certified laboratory setting, although in some cases, that may not be possible.

Although a candidate omics-based test will have been developed and confirmed in the research setting using certain instruments and assay methods, those same methods may not be appropriate in the clinical laboratory setting based on the clinical requirements for the test. Clinical laboratories prefer to use existing instruments and methods familiar to the laboratory staff for financial and workflow reasons. When the clinical laboratory does not have appropriate instruments for a new omics-based test, either a new method can be developed and performed on instruments already available in the clinical laboratory, the appropriate instruments can be purchased for the clinical laboratory, or the research instruments can be brought under CLIA standards and used by the clinical laboratory for test validation and eventual performance of the test in a clinical trial setting. The financial support for the translation of an omics-based test into the clinical laboratory setting in preparation for clinical trials should be considered when seeking research support for either the omics discovery phase or the clinical trial phase of the translation

pathway, if the test will be validated in the academic setting. If the test is validated through an industry partnership, the discovery should be validated as a clinical test by the company prior to use of the test in a clinical trial setting.

**The committee recommends that the CLIA-certified laboratory should design, optimize, validate, and implement the omics-based test under current clinical laboratory standards (Recommendation 2c).** The clinical laboratory, in collaboration with the discovery investigators, should design the clinical test methods based on the intended use and clinical requirements for the test. While the use of a test may change once it reaches clinical use, the intended use of the candidate test should be defined prior to the test validation process to assure that the proper types of specimens are used for test validation. Using control analytes and clinical or biological specimens relevant to the intended use of the test, the initial test methods will be assessed for their performance characteristics, and adjusted as needed to achieve predefined required performance characteristics. The defined test method should include the fully specified computational procedures that were locked down at the end of the discovery phase, which will be used for analysis of the assay results to yield the final test result that is used as the interpretive criteria for the omics-based test. Once the test method, including the interpretive criteria, is established and documented, the methods can be assessed for their analytical and clinical performance characteristics in a test validation process.

### Analytical and Clinical/Biological Validation

The test validation process includes the steps used to establish the analytical and clinical performance characteristics and the test limitations of a test prior to clinical use (Jennings et al., 2009; Mattocks et al., 2010). Test validation is defined under ISO 9000:2005[2] to mean "confirmation, through the provision of objective evidence, that the requirements *for a specific intended use* or application have been fulfilled," which has been interpreted as "doing the correct test." This is in contrast to test verification, which is defined under ISO 9000:2005 to mean "confirmation, through the provision of objective evidence, that specified requirements have been fulfilled," which has been interpreted as "doing the test correctly" (Jennings et al., 2009; Mattocks et al., 2010). This distinction is critical in the validation of a new clinical test because the intended use of the test should be defined at a very early stage of test development to ensure that the proper specimen types from appropriate patients with specific clinical characteristics are used for the test validation before clinical use of the new test.

The test validation process has been defined for molecular tests (Jennings et al., 2009; Mattocks et al., 2010), but not for complex omics-based tests, which are based on large datasets that must be interpreted using a computational model. All aspects of the test validation process should be applied not only to the assay that generates the data, but also to the fully specified computational procedures used to analyze the data, assuring the analytical and clinical performance of both aspects of the complete test. For omics-based tests, the usual interpretive criteria used to produce a final test result will generally be replaced by these computational procedures, although the specific cutoff values and potentially definition of indeterminate results may need to be defined as the test method is developed within the CLIA-certified laboratory. Finally, for test validation processes that warrant publication due to the novel or new aspects of the test, the appropriate guidelines for reporting of biomarker study results should be considered at the planning phase of the test validation process and be used to ensure consistency in the

[2] These standards are published by the International Organization for Standardization (ISO). ISO 9000:2005 provides information on the fundamentals and vocabulary used in quality management systems.

reporting of the test validation process results (Surinova et al., 2011). As described in Appendix D, several reporting guidelines have been developed for different types of biomarker research, or different aspects of the test development process, including BRISQ (Moore et al., 2011), REMARK (Altman et al., 2012; McShane et al., 2005), and STARD (Bossuyt et al., 2003).

### *Analytical Validation*

The test method should be optimized to achieve an analytically validated test, including the assay and the fully specified computational procedures. Once the test methods have been defined and have achieved the predefined performance characteristics for the test, the next step should be rigorous testing of the analytical performance of the test. This analytical validation identifies and quantifies the technical variations of a test performed on patient specimens. The test parameters for analytical validation are extensive and vary depending on the purpose of the test, such as qualitative or quantitative, and include, for example, accuracy, precision, reproducibility, linearity, reportable range, analytical sensitivity and specificity, and limit of detection. Analytical validation is performed using specimens comparable to the patient specimens on which the test will eventually be used, with known or expected characteristics related to the test being validated. An alternative to using a limited supply of valuable clinical specimens with known or expected test results is the use of control materials that will provide known or expected test results and can be spiked into negative clinical specimens of the same type to be used for the test. Different amounts of the control material can be introduced into the negative specimens to assess the linearity, sensitivity, cut-off values, and other performance characteristics of the test during analytical validation of the test.

The different analytical performance characteristics to be established depend on the purpose of the test. The types of performance characteristics to be established during analytical validation of a test are described by Jennings et al. (2009). For an omics-based test, the precisely defined computational procedures used for interpretation of the assay results must be included in the validation process to ensure that it performs well on the assay results yielded by the test method in the clinical laboratory. Also, all data management steps and processes should be defined and tested to ensure that no errors are introduced at the data transfer or interpretation steps of the test.

### *Clinical/Biological Validation*

The optimized test should be evaluated for clinical/biological validity using an independent specimen set, consistent with the intended use for the test. The expected or previously determined test results for the validation set of specimens and clinical information about the patients from which the specimens were derived should not be available to the individuals performing and interpreting the test results to reduce the chance that results will be affected by conscious or unconscious bias. This process of blocking this information from the tester is called blinding of the specimens. Clinical/biological validation provides evidence that the test results are linked to disease states, and is guided by the intended use of the test. Clinical/biological validation includes but is not limited to reference range, clinical sensitivity, and clinical specificity (Jennings et al., 2009). Clinical/biological validation is performed using specimens comparable to the patient specimens on which the test will eventually be used, and a sufficient number of specimens to allow statistical assessment of the test performance in relation to the specimen characteristics. If clinical specimens are not available, then appropriate

biological specimens, such as cell lines approximating the eventual clinical specimens can be used, but this is not ideal because the clinical specimens will introduce variability not seen in the biological specimens. The number of specimens needed for the clinical/biological validation of a test depends on many factors, including whether a comparative test is available for comparison of results, the test performance characteristics required for clinical use (e.g., accuracy, sensitivity, etc.), the prevalence of the disease, and the confidence level required for good medical practice (Jennings et al., 2009).

As noted above, the identity and expected results of the specimens used for clinical/biological validation of the omics-based test should be blinded to the individuals performing and interpreting the test results during clinical/biological validation of the test. Generating the results in a blinded fashion will reduce bias during the clinical/biological validation and improve the odds of successful assessment of clinical use when the test is applied to specimens collected in prospective trials or is used to direct patient management in a clinical trial (see the Evaluation for Clinical Use Phase, Chapter 4). If specimens from the omics discovery phase are used for the clinical/biological validation of the test, then the specimens can be provided in tubes labeled numerically, while maintaining a master key to the identity of the specimens. Once results are completed, the key can be used to compare the test results to results obtained during the discovery research phase, or to the clinical outcomes of the patients from whom the specimens were derived. Again, the use of specimens from the discovery phase is not optimal, since it does not provide an independent assessment of the performance of the test.

## PREPARATION FOR INVESTIGATIONAL USE OF THE VALIDATED TEST

All requirements for performance of clinical tests in a CLIA-certified laboratory should be met for the defined and validated omics-based test. The final step of test validation is the implementation of the new test in the workflow and quality management system of the CLIA-certified laboratory. This final step in the test validation process is essential for assuring the test is performed under a quality management system designed to comply with best practices for clinical laboratories. Test implementation requires a written standard operating procedure (SOP) for the test that describes the specimens and specimen handling requirements, assay methods, data handling procedures, fully specified computational procedures used to interpret the assay results in order to yield clinically actionable results, and methods for reporting results. The SOP also should describe the procedures for quality control and quality assurance, including issues related to the preanalytical, analytical, and postanalytical aspects of testing; quality indicators to be monitored such as turnaround time, documentation, and investigation of test failures; and trends in test results, as well as ongoing calibration processes for the instruments and the test. The SOP also is used for training technologists who will perform the test, with documentation of the success of initial training and of continued competency of all personnel performing the test. Because proficiency testing programs for the new test almost certainly will not be available, procedures internal to the laboratory can be used for ongoing assessment of the quality of the test performance. If the test will be used in a clinical trial setting, mechanisms for reporting of the test results with delivery to the clinical team should be defined and tested.

**The committee recommends that if the omics-based test will be performed in more than one CLIA-certified laboratory for a clinical trial, analytical validation and CLIA requirements for the same omics-based test should be met by each laboratory, working with the primary laboratory (Recommendation 2d).** The test validation process described

above covers aspects of test validation in a single laboratory, but clinical trials often are performed at multiple clinical sites. Sometimes testing still can be performed in a single laboratory, even with multiple clinical trial sites, if the specimen handling requirements and the turnaround times required for clinical care can be accommodated with specimen transport to the single laboratory. If testing is to be performed in laboratories at each clinical site, then each laboratory must complete a test validation process, and a plan to compare findings from the different laboratories should be defined. However, if the test is validated as described above as a new test by one laboratory, then the other laboratories can use the test methods, specimens, and test results available from the first laboratory to complete a more abbreviated test validation process, and use the primary laboratory's SOP and quality control and assurance procedures with modifications as required by each laboratory site. If different test methods are used by the laboratories at different sites, then a full test validation process as described in this chapter should be completed by each laboratory. When different laboratories are used, comparison of the test results between sites should be included in the study design to assess bias introduced by laboratory differences.

## FUNDING FOR VALIDATION OF A CANDIDATE OMICS-BASED TEST IN A CLIA-CERTIFIED CLINICAL LABORATORY

Validation of a candidate omics-based test in a CLIA-certified clinical laboratory requires funding for the reagents, supplies, controls, and personnel time, as well as submission of the IRB protocol for access to and use of appropriate clinical specimens, and potentially for new instrumentation required for the selected test method. Two options are available for completing the test validation phase of the translational pathway; either partnering with industry through the technology transfer offices of academic institutions or continuing the development within the academic environment. While the discovery phase of omics-based test development (described in Chapter 2) is routinely funded by the National Institutes of Health (NIH), grant funding historically has not been available for validation of a candidate omics-based test in a CLIA-certified laboratory (described in this chapter) prior to assessment in a prospective clinical trial (described in Chapter 4). In a clinical setting, the costs of validation of a new clinical test for immediate clinical use are supported by the clinical funding for the laboratory in anticipation of clinical billing and reimbursement for the test; however, the costs of validation of a candidate omics-based test in preparation for assessment in a clinical trial will not be supported by the clinical funding. As an alternative to a CLIA-certified laboratory in an academic medical center, commercial clinical research organizations are available for development and validation of a candidate omics-based test in a CLIA-certified laboratory setting, but also will require financial support. Investigators need to include this phase of test development and validation in their grant proposals, and as recommended in chapter 5, funding agencies should provide appropriate support for this phase of test development that addresses analytical and clinical/biological validation, in anticipation of assessment of the test in a prospective clinical trial to direct patient management.

## SUMMARY AND RECOMMENDATION

For any test that is used for patient management decisions, whether in a clinical trial or patient care setting, testing should be performed in a clinical laboratory certified under CLIA through CMS or an accreditation organization with deeming authority under CLIA. CLIA

certification ensures that the clinical laboratory operates under standards for quality control, quality assurance, and quality management. Validation and performance of a test in a CLIA-certified laboratory provides an overall quality management system that ensures best clinical laboratory practices are used and reduces the risk for patient harm due to improper test validation or performance. At the end of the test development phase described in this chapter, the clinical test method should be locked down (fully defined and not changed in the next development stage), just as the computational procedures were locked down in the discovery phase in chapter 2.

**RECOMMENDATION 2:**
**An omics-based test consists of both the data-generating assay and the fully specified computational procedures used for analysis of the assay data. The committee recommends that both components of omics-based tests used to direct patient management in a clinical trial setting should be validated during the test validation phase using the following steps:**

- **a. The candidate omics-based test and its intended use should be discussed with the FDA prior to initiation of validation studies.**
- **b. Test development and validation should be performed in a CLIA-certified clinical laboratory, beginning with a defined candidate omics-based test concept from the discovery phase.**
- **c. The CLIA-certified laboratory should design, optimize, validate, and implement the omics-based test under current clinical laboratory standards.**
- **d. If the omics-based test will be performed in more than one CLIA-certified laboratory for a clinical trial, analytical validation and CLIA requirements for the same omics-based test should be met by each laboratory, working with the primary laboratory.**

## REFERENCES

Altman, D. G., L. M. McShane, W. Sauerbrei, and S. E. Taube. In press. Reporting recommendations for tumor marker prognostic studies (REMARK): Explanation and elaboration. *BMC Medicine* [Epub before print].

Bossuyt, P. M., J. B. Reitsma, D. E. Bruns, C. A. Gatsonis, P. P. Glasziou, L. M. Irwig, J. G. Lijmer, D. Moher, D. Rennie, H. C. de Vet, and Standards for Reporting of Diagnostic Accuracy. 2003. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clinical Chemistry* 49(1):1-6.

CDC (Centers for Disease Control and Prevention). 2004. *Subpart A-General provisions*. http://wwwn.cdc.gov/clia/regs/subpart_a.aspx#493.3 (accessed October 13, 2011).

Gutierrez, A. 2011. *Omics-Based Tests*. Discussion with the Science and Technology Working Group, Conference Call with Alberto Gutierrez, Washington, DC, August 19.

Jennings, L., V. M. Van Deerlin, and M. L. Gulley. 2009. Recommended principles and practices for validating clinical molecular pathology tests. *Archives of Pathology and Laboratory Medicine* 133(5):743-755.

Mattocks, C. J., M. A. Morris, G. Matthijs, E. Swinnen, A. Corveleyn, E. Dequeker, C. R. Muller, V. Pratt, A. Wallace, and EuroGentest Validation Group. 2010. A standardized framework for the validation and verification of clinical molecular genetic tests. *European Journal of Human Genetics* 18(12):1276-1288.

McShane, L. M., D. G. Altman, W. Sauerbrei, S. E. Taube, M. Gion, and G. M. Clark. 2005. Reporting recommendations for tumor marker prognostic studies (REMARK). *Journal of the National Cancer Institute* 97(16):1180-1184.

Moore, H. M., A. B. Kelly, S. D. Jewell, L. M. McShane, D. P. Clark, R. Greenspan, D. F. Hayes, P. Hainaut, P. Kim, E. A. Mansfield , O. Potapova, P. Riegman, Y. Rubinstein, E. Seijo, S. Somiari, P. Watson, H.-U. Weier, C. Zhu, J. Vaught. 2011. Biospecimen reporting for improved study quality (BRISQ). *Cancer Cytopathology*, 119(2): 92-102.

SACGHS (Secretary's Advisory Committee on Genetics, Health, and Society). 2008. *U.S. System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services; a Report of the Secretary's Advisory Committee on Genetics, Health, and Society*. http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf.

Surinova, S., R. Schiess, R. Huttenhain, G. Cerciello, B. Wollscheid, and R. Aebersold. 2011. On the development of plasma protein biomarkers. *Journal of Proteome Research* 10(1):5-16.

Wadsworth Center. 2009. Submission Guidelines for Nucleic Acid Amplification Tests for Infectious Agents, Rev 7/1/2009. http://www.wadsworth.org/labcert/TestApproval/forms/proposedguidelines.pdf (accessed February 2, 2012).

# 4
# Evaluation of Omics-Based Tests for Clinical Utility and Use

This chapter explains the recommended steps for assessing a validated omics-based test for clinical utility and use. In particular, this chapter will explore elements recommended by the committee to evaluate a test for clinical use in trials in which the omics-based test will be used to guide patient management decisions. It should be noted again that although the Institute of Medicine (IOM) committee was charged with recommending an evaluation process for omics-based tests, much of the proposed process, and the material presented in this chapter in particular, applies broadly to the development of biomarker-based tests. As described in chapter 1, omics-based tests can be considered a complex form of a biomarker, using a defined set of measurements combined with a computational model as a clinical test.

In parallel to the recommendations presented in Chapters 2 and 3, Figure 4-1 highlights the recommended steps in the evaluation for clinical utility and use of tests that are intended to guide patient management in a clinical care setting, the third component of the committee's recommended development and evaluation process for omics-based test. The end product of the test development phase described in chapter 3 in a fully defined and locked down clinical test that has undergone analytical and clinical biological validation.

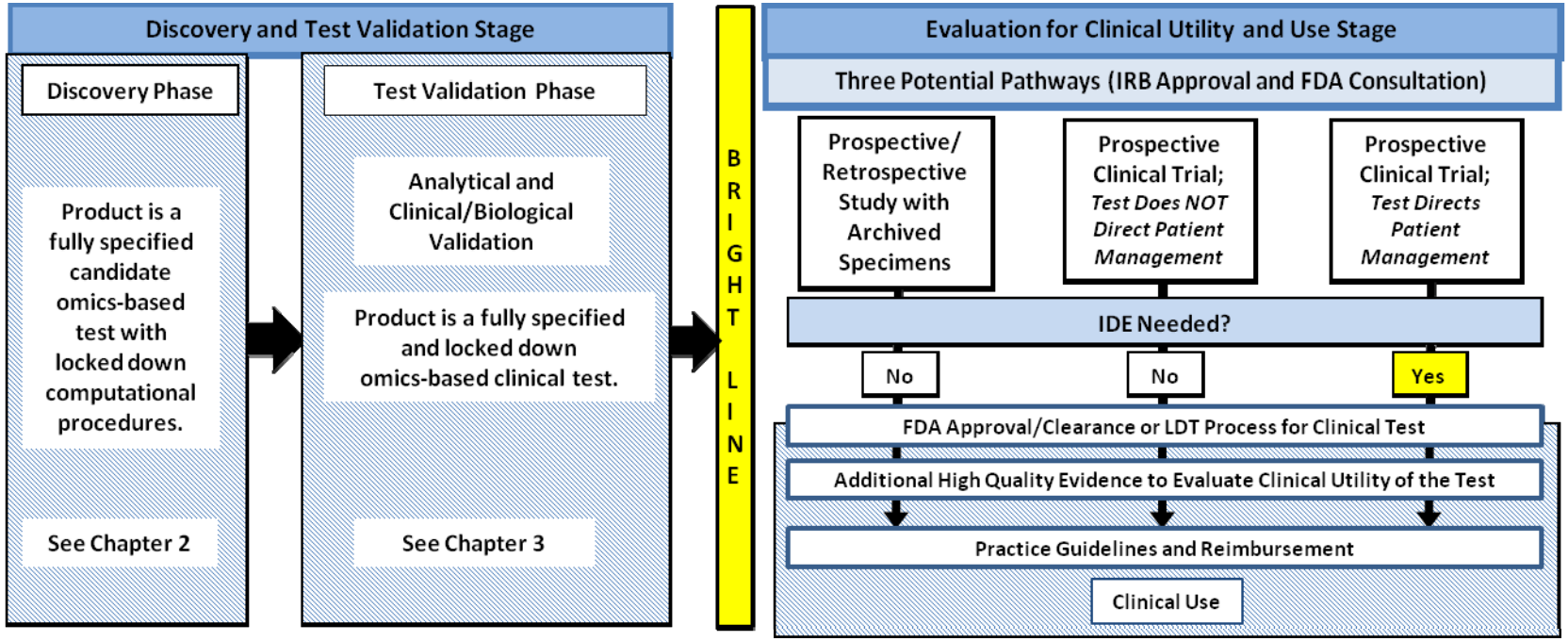


**FIGURE 4-1** Omics-based test development process, highlighting the evaluation for clinical utility and use stage. The bright line signifies the point in test development where a fully defined, validated, and locked down clinical test is necessary. Changes to the test after the bright line is crossed require going back to the test validation phase, approval by the IRB, and possibly consultation with the FDA. In the second stage of test development and evaluation, the fully defined and validated omics-based test undergoes evaluation for clinical use. Determination of clinical utility is a process that occurs over a longer time frame.
NOTE: FDA = Food and Drug Administration, IRB = Institutional Review Board, LDT = laboratory-developed test.

## BACKGROUND

As discussed in Chapter 1, the committee adopted a modified version of the terminology for biomarker validation developed by the Evaluation of Genomic Applications in Practice and Prevention Working Group. Clinical utility is defined as "evidence of improved measurable clinical outcomes, and [a test's] usefulness and added value to patient management decision-making compared with current management without [omics] testing" (Teutsch et al., 2009, p11). The Food and Drug Administration (FDA) does not require evidence of clinical utility for its evaluation of a clinical test; indeed, it can take many years after a test is introduced in the marketplace to demonstrate clinical utility of the test. Likewise, the lack of FDA review for a laboratory-developed test (LDT) does not mean the test does not have clinical utility.

Various methods have been proposed over the past two decades to systematically establish the clinical utility of biomarkers in oncology, similar to those used for new therapeutics (Altman and Lyman, 1998; Altman and Riley, 2005; Altman and Royston, 2000; Hayes et al., 1996; Simon, 2005a, 2005b; Simon and Altman, 1994; Simon et al., 2009). Demonstration of clinical utility for an omics-based test is fundamentally similar to that of biomarker tests in general, although attention to the challenges unique to omics-based test development is necessary in the discovery and test validation phases (see Chapters 2 and 3).

Traditionally, devices such as biomarker tests have been subject to a regulatory process different from that used for new drugs. However, just as an ineffective drug can be harmful to patients, an inaccurate test has the potential to harm patients by directing them to ineffective, toxic treatments or by directing them away from potentially life-saving drugs. As the use of biomarker tests, including omics-based tests as discussed in this report, become more closely tied to patient management decisions such as choice of therapies, it is worth drawing a comparison between the two processes. Just as it is important to fully standardize and understand the components, dose, and schedule of a new therapeutic before taking it into clinical trials to assess clinical efficacy, it is essential that analytical validity for an omics-based test be demonstrated (Chapters 2 and 3). Likewise, in a manner similar to the use of a Phase II clinical trial to demonstrate that a new therapeutic appears to have some level of activity that would justify the expense and costs of a large-scale definitive Phase III trial, it is essential to demonstrate clinical/biological validity of a new omics-based test prior to assessing its clinical utility. Without a demonstration of a test's "ability to accurately and reliably predict the clinically defined disorder or phenotype of interest" (Teutsch et al., 2009, p. 10), a claim that a clinical trial to evaluate potential clinical utility has a reasonable chance of success would not be justified. The goal of such a clinical trial in therapeutics is to demonstrate that the strategy is either superior to the current standard of care, or that it is equivalent to standard of care with some other advantage (lower cost, lower toxicity, more convenient administration, etc.), a concept that can be extended to biomarker tests, including omics-based tests.

In addition to regulatory oversight, the threat of legal liability for a defective product should also deter organizations from moving a defective omics-based test into a clinical trial or clinical use prematurely. Under tort law, a company developing and/or marketing a test is responsible for paying damages to an individual who has experienced a harm or loss caused by a defective product. Omics-based tests that go through the FDA approval process are protected

against certain torts.[1] However, all test developers are liable for losses and harms caused by negligence (i.e., the failure to exercise the care that a reasonably prudent person would exercise in like circumstances). The consequences of a products liability lawsuit may differ depending on the size of the test developer. For example, a small startup company has the net asset value of the firm to compensate for damages from defective products. Large academic medical centers have more resources to loose, and hence are more at risk from defective products.

## EVALUATION FOR CLINICAL UTILITY AND USE STAGE

Ultimate achievement of clinical utility, which is often assessed by organizations such as the US Preventive Services Task Force (USPSTF), the BlueCross BlueShield Association Technology Evaluation Center, the National Cancer Center Network, and the American Society of Clinical Oncology, for example, is a process that often continues long after a test is first introduced into clinical practice. Recent assessments by USPSTF of mammography and prostate-specific antigen screening highlight the length and dynamic nature of this process (USPSTF, 2009, 2011). Nonetheless, the process of gathering evidence pertaining to clinical utility begins before a test is introduced into clinical practice, and the level of evidence necessary prior to clinical use of a test is a subject of much scholarly work (AHRQ, 2011; Simon, 2010; Simon et al., 2009: Teutsch et al., 2009).

The purpose of this chapter is to briefly outline the types of clinical studies and clinical trials used to gather evidence prior to introducing tests into clinical practice and for assessing the clinical utility of tests, but only inasmuch as needed to explain the committee's recommendations related to the Evaluation for Clinical Utility and Use Stage of omics-based test development. As shown in Figure 4-1, investigators might follow one of three pathways to generate sufficient evidence for this evaluation stage. The design of the clinical study or clinical trial used to assess clinical utility depends on the intended use of the omics-based test as well as the availability of appropriate archived specimens. Regardless of the pathway chosen, investigators are strongly urged to communicate with the FDA, especially regarding the potential need for an investigational device exemption (IDE) if the test is to be used in a prospective clinical trial. Following the clinical studies or clinical trials, the test either must achieve FDA clearance or approval or must follow the LDT process for introducing a test into clinical practice, with subsequent evaluation and perhaps incorporation of the test into clinical practice guidelines and reimbursement by payers. These steps are shown for completeness, but are beyond the scope of this study and are not discussed in this chapter.

### FDA Regulation

Introduction of an omics-based test into standard clinical practice can occur in two ways: by seeking FDA approval or clearance using the premarket approval or 510(k) process, respectively, or by developing an LDT within a specific laboratory. Any clinical laboratory that reports tests for clinical management of patients falls under the purview of the *Clinical Laboratory Improvement Amendments of 1988* (CLIA). LDTs developed by a CLIA certified Laboratory do not require FDA approval or clearance. Although FDA has the authority to regulate these tests, it has chosen to exercise enforcement discretion in most cases. The question of whether the FDA should exercise greater authority over the regulation of LDTs has been the

[1] Murphree v. Pacesetter, No. CT-005429-00-3 (Cir. Ct. Tenn., 13th Judicial Dist.), and Horn v. Thoratec Corp., 376 F.3d 163 (3d Cir. 2004).

subject of much debate. The Secretary's Advisory Committee on Genetics, Health, and Society recently recommended that FDA should provide oversight of all LDT's, with priority given to higher risk tests (SACGHS, 2008). One challenge in implementing this approach is the need for additional resources and expertise at FDA to provide this level of oversight.

Of note, FDA review of a biomarker test has been focused principally on analytical and clinical/biological validity, but not on demonstration of clinical utility, as defined in this report. Therefore, FDA approval or clearance does not necessarily imply that the test improves clinical outcomes or should be used for patient management. LDTs performed in CLIA-certified laboratories also do not require evidence of clinical utility; only analytical and clinical validity of the test must be demonstrated prior to clinical use.[2]

This situation has led guidelines committees and third-party technical assessment panels to perform opinion or evidence-based analyses of available data to make recommendations regarding whether selected tumor biomarker tests should be used clinically, and whether they should be supported by reimbursement, independent of FDA approval or clearance (AHRQ, 2011; Allegra et al., 2009; Harris et al., 2007; Locker et al., 2006; NCCN, 2010; TEC, 2011; Teutsch et al., 2009). In general, these committees and panels have used similar definitions of analytical and clinical/biological validity and clinical utility adopted for this report.

Nonetheless, to address the fundamental charge of this Committee (determining when an omics-based test is fit for use as a basis for a clinical trial design), it is important to ensure that a test used to direct care in a prospective clinical trial is vetted very carefully. As in the case of therapeutics, the potential for conflicts of interest and the overall benefit of external, objective review highlight the importance of involving experts not associated with the design or conduct of the clinical trial in this vetting process. The committee believes that the FDA is the most competent and appropriate body for this role. Therefore, as discussed in Chapters 2, 3, and 6, the committee recommends that **investigators should communicate early with the FDA regarding the IDE process and validation requirements (Recommendation 3a).** When a test will be used to direct patient management in a clinical trial, an IDE is required according to current regulatory standards. This corresponds with the third pathway to clinical utility shown in Figure 4-1. For the first two pathways, while IDEs are often not necessary, consultation with the FDA at an early stage of test development still is recommended due to the complexity of the development process and regulations for eventual approval of the test for clinical use.

In a discussion with the committee, Dr. Alberto Gutierrez, director of the Office of In Vitro Diagnostics in the Center for Devices and Radiological Health, acknowledged that the FDA has not clearly defined what constitutes an LDT or when enforcement discretion would be applied, and noted that the FDA has received fewer IDE applications for in vitro diagnostic test devices than expected from academic centers (Gutierrez, 2011). He added that while academic medical centers are subject to the same regulations and policies as commercial test developers, FDA may not expect academic medical centers to comply with some quality systems regulations, such as good clinical practices (GCP) and good laboratory practices (GLP) (Gutierrez, 2011). Instead, a sufficient demonstration of analytical validation of the test would be required. Consistent with the committee's recommendation, he also advised that test developers should consult with the FDA to determine whether their devices qualify as LDTs or whether IDEs are needed (Gutierrez, 2011).

[2] For clinical laboratories within New York State and for any clinical laboratory performing testing on specimens from New York State patients, the New York State Department of Health must also review the validation and test performance characteristics of all LDTs prior to clinical use.

## Design of Clinical Studies and Clinical Trials

The objective of the current Institute of Medicine review is not to determine the criteria for clinical utility of an omics-based test. Rather it is to determine the criteria for considering use of an omics-based test to direct management of patient care in a clinical trial designed to assess the clinical utility of the test. However, a discussion of the criteria for generation of high quality evidence in a clinical trial to assess clinical utility is essential as a preamble for when such a test should be used to direct patient management in a clinical trial.

To continue the comparison between the regulatory processes for approval of a new therapeutic and the regulatory process for clearance of a new omics-based test, a new therapeutic must undergo rigorous definitive testing in one or more properly designed and conducted prospective clinical trials before it is accepted for routine clinical use. The same is true for omics-based tests that might be used to direct clinical management. In order for such a test to be incorporated into routine clinical care, investigators should understand and have articulated the precise intended use of the test, as described in Appendix C. It is also important that an estimate of the magnitude of difference in clinical outcomes associated with different results of the omics-based test be sufficiently large to warrant different patient management (Henry and Hayes, 2006).

For example, a test might have clinical/biological validity by virtue of distinguishing the outcome of two subgroups within a population with a high degree of statistical significance, but if this difference is not large, these two groups of patients might not be treated differently. Such a test has clinical/biological validity, but not clinical utility.

Ideally, an omics-based test would identify one group of patients for whom the benefits from a specific therapy are likely to outweigh the risks, and another group for whom the risks of the therapeutic strategy under consideration are likely to outweigh the benefits. A test with this characteristic would have utility as a clinically useful effect modifier.

Importantly, this estimate of the magnitude of difference in outcomes must be reliable. As discussed in Chapter 3, the committee's recommendations to establish analytical and clinical/biological validity of an omics-based test are critical, and the decision regarding whether to move forward into a clinical trial requires compelling clinical/biological validity. While most test validation studies may support the clinical/biological validity of a new test, they do not demonstrate clinical utility.

Determination of clinical utility requires generation of high quality evidence that supports the intended use of the test. In this regard, once the test performance characteristics are fully defined and the test is validated in a clinical laboratory, the investigator then needs to determine whether the test can and should be used to direct management of patient care. As illustrated in Figure 4-1, there are three pathways to generate such evidence:

- Prospective–retrospective studies using archived specimens from previously conducted clinical trials that address the intended use of the omics-based test, or
- Prospective clinical trials that directly address the utility of the omics-based test, where either
  - The test does not direct patient management, or
  - The test does direct patient management.

Each of these pathways will be discussed in the next sections of the chapter and examples of trial designs are summarized in Table 4-1. The trial designs discussed in these sections are examples

of different approaches investigators might take to "test the test"; different trial designs address different questions, so it is important to understand whether the selected trial design will provide the information needed to assess a given test for its intended use. More recently, adaptive trial designs, which have been used in therapeutic clinical trials (IOM, 2010), also have been suggested for trials that address the clinical utility of a biomarker test

**TABLE 4-1** Examples of Clinical Study Designs to Assess Omics-based Tests

| TYPE OF TRIAL DESIGN | DESCRIPTION | QUESTIONS STUDY DESIGN CAN ANSWER | ADVANTAGES | LIMITATIONS |
|---|---|---|---|---|
| Retrospective studies | Test is studied using archived specimens that were collected in the past, have been stored, and happen to be available. | Generally retrospective studies should be viewed as exploratory, and at best they can provide preliminary evidence of clinical validity by showing associations between the test result and outcome. | Availability of specimens and varying types of data, potentially including clinical and pathologic variables and clinical outcome. | • Not collected as part of a clinical trial, so treatment is not randomized and choices could have been confounded with markers underlying the test and other factors related to outcome.<br>• Patients may not be representative of intended use population for the test.<br>• Standard of care may have evolved over time.<br>• Archived specimens may not be available, might not have been collected, processed or stored in the manner required, or might have degraded over time rendering the test results unreliable.<br>• Clinical and pathologic data may not reflect current standard definitions, may have missing values, and inaccuracies. |
| Prospective-Retrospective studies | • Uses archived specimens from a previously conducted clinical trial with treatment(s) that are relevant to the intended clinical use of the test.<br>• This design requires a prospective written protocol describing study objectives, methods, and analytic plan and the test must be locked down and analytically validated. | • What is the best treatment in each test-defined subgroup?<br>• What is the best treatment in the overall study population?<br>• Is test-directed treatment better than standard of care in the overall study population? | • Study design may have evidentiary value close to a prospective study under certain conditions (Simon et al, 2009).<br>• Less resource- and time-intensive than a prospective clinical trial because specimens have already been collected and follow-up for observation of clinical endpoint has already occurred. | • Archived specimens may not be available, might not have been collected, processed or stored in the manner required, or might have degraded over time rendering the test results unreliable.<br>• Standard of care may have evolved from the time of the clinical trial.<br>• At least two prospective-retrospective studies showing consistent promising results are |

| | | | | |
|---|---|---|---|---|
| | | (indirect assessment only)<br>• Is the test prognostic?<br>• Is the test a treatment effect modifier?<br>(Note that the ability to answer the above questions depends on whether the required treatment arms are represented and statistical power is sufficient.) | • A possible study design if a prospective trial is not feasible for ethical or other reasons. | required<br>• Prospective clinical studies may still be needed to fully assess clinical utility.<br>• Statistical inference concerns relating to multiple testing, since the primary analysis of the original trial likely did not pre-specify a plan to study treatment effect in the test-defined subgroups. |
| Prospective studies, not test-directed | | | | |
| | Completely randomized design (Figure 4.2a) | • Patients are randomly assigned to therapy regardless of specimen availability and without knowledge of test result.<br>• The test is applied to patient specimens that are available. Primary analysis for treatment effect is not stratified by test result. | • What is the best treatment in the overall study population?<br>• Among patients for whom test results are available,<br>o is the test prognostic?<br>o Is the test a treatment effect modifier?<br>o what is the best treatment in each test-defined subgroup?<br>o is test-directed treatment better than standard of care? (indirect assessment | • Test result does not need to be known prior to randomization, so treatment need not be delayed for return of test result.<br>• Test can potentially be performed on banked specimens even years after the treatment study has completed. | • Unless availability of test result is required to be eligible for treatment randomization, the potential exists for bias to be introduced due to nonrandomly missing test results.<br>• If test is performed on specimens long after they are collected, there is a possibility that the specimens might have degraded over time rendering the test results unreliable.<br>• Statistical power to detect prognostic factors or treatment effect modifiers, or treatment effect in the test-positive subset may be limited unless effects are strong.<br>• If number of patients accrued to the trial is small or if the size of certain test-defined subgroups is very small, treatment assignments might not be well- |

| | | | | |
|---|---|---|---|---|
| | | only) | | balanced within the test result categories, resulting in a loss of efficiency. |
| Test-stratified design (randomized marker-block design) (Figure 4.2b) | • The test is applied to all patients to determine test-defined subgroups.<br>• Patients are randomly assigned to therapy within each test-defined subgroup to ensure balance of treatment arms within each subgroup.<br>• Statistical analyses are stratified by the test result. | • What is the best treatment in each test-defined subgroup?<br>• What is the best treatment in the overall study population?<br>• Is test-directed treatment better than standard of care in the overall study population? (Indirect assessment only)<br>• Is the test prognostic?<br>• Is the test a treatment effect modifier? | • Efficient design for assessment of relative treatment efficacy in each test-defined subgroup and in the whole group, and for assessment of predictive (interaction) effect.<br>• Up-front stratification ensures that test results are available for all patients who are accrued to the trial. | • Not feasible for evaluation of test strategies with a large number of treatment options |
| Prospective studies, test-directed (Pose highest risk to patients if test performance is poor) | | | | |
| Enrichment (Figure 4.3) | • The test is applied to all patients, but only test-positive patients are randomized and/or treated.<br>• Test-negative patients are either off study or followed prospectively in a registry. | • What is the best treatment in test-positive patients? | • Useful when clinical utility of some of the test-designated categories is already established or assumed and need not be re-evaluated, while other categories require prospective evaluation in a clinical trial.<br>• Efficient design for assessing efficacy of treatment in test-positive patients | • Cannot assess the treatment benefit in test-negative patients<br>• Cannot assess if the test is prognostic or a treatment effect modifier<br>• Provides insufficient information about the utility of the test if uncertainties exist about the new therapy's effect in test-negative patients; if the new therapy is beneficial in test-negative patients, a promising therapy may be missed for these patients |

| | | | | |
|---|---|---|---|---|
| | | | • Acceptable in circumstances where a certain subgroup of patients is thought so unlikely to experience an event with standard of care, or to benefit from the new therapy, that it would be unethical to randomize those patients | • Clinical utility of the test for therapy selection is not confirmed because treatment efficacy cannot be evaluated in test-negative patients and therefore cannot be compared with treatment efficacy in test-positive patients |
| Test-guided strategy versus standard of care (Figure 4.4) | • Patients are randomly assigned to the test-guided arm or the non-guided arm.<br>• Patients randomized to the test-guided arm are directed to therapy as dictated by the test (test-positive to new therapy, test-negative to standard of care).<br>• An alternate rule must be pre-specified to handle cases for which a test result is unavailable for a patient assigned to the test-guided arm.<br>• Patients randomized to the non-guided arm receive standard of care.<br>o Version 1: The test is applied to all patients (preferably at the same time in both arms).<br>o Version 2: The test is applied only to patients randomized to the test-guided arm. | • Is test-directed treatment better than standard of care in the overall study population? (direct assessment)<br>• What is the best treatment in test-positive patients? (Version 1 only)<br>• Is the test prognostic? (Version 1 only) | • Required number of tests to be performed using Version 2 will be half that required for Version 1 (assuming equal randomization) so it may be useful in settings of limited resources<br>• No issues associated with withholding test results from patients on non-guided arm because they receive standard of care<br>• Compliance with assigned treatment is not influenced by patient knowledge of test result in non-guided (standard of care) arm<br>• Can be used to evaluate complex test-directed strategies with a large number of treatment options or test categories | • Cannot establish the best treatment for test-negative patients<br>• Cannot establish whether the test is a treatment effect modifier<br>• Clinical utility as a treatment effect modifier (i.e., as a predictive test) is not confirmed because the new therapy might be better for all patients regardless of test result<br>• Inefficient due to the fact that test-negative patients would receive the same treatment regardless of whether assigned to the test-guided or non-guided arm |
| Test-guided strategy versus non-guided with randomization | • Patients are randomly assigned to the test-guided arm or the non-guided arm.<br>• Patients randomized to the | • Is test-directed treatment better than standard of care in the | • Required number of tests to be performed using Version 2 will be half that required for Version 1 (assuming | • Inefficient, in terms of number of patients required, relative to the test-stratified design<br>• Version 2 has the potential to |

| | | | | |
|---|---|---|---|---|
| (Figure 4.5) | test-guided arm are directed to therapy as dictated by the test (test-positive to new therapy, test-negative to standard of care).<br>• An alternate rule must be pre-specified to handle cases for which a test result is unavailable for a patient assigned to the test-guided arm.<br>• Patients randomized to the non-guided arm undergo a second randomization to receive either new therapy or standard of care.<br>  ○ Version 1: The test is applied to all patients (preferably at the same time in both arms).<br>  ○ Version 2: The test is applied only to patients who are randomized to the test-guided arm. | overall study population? (direct assessment)<br>• Is test-directed treatment better than the new therapy in the overall study population? (Direct assessment)<br>• What is the best treatment in each test-defined subgroup? (Version 1 only)<br>• Is the test prognostic? (Version 1 only)<br>• Is the test a treatment effect modifier? (version 1 only) | equal randomization) so it may be useful in settings of limited resources<br>• Can be used to evaluate complex test-directed treatment strategies using a large number of treatment options or test categories<br>• Required number of tests to be performed using Version 2 will be half that required for Version 1 (assuming equal randomization) and might be fewer than the completely randomized design or test-stratified design, depending on total sample size used in each design. However, not all questions can be addressed with version 2 of this design. | produce biased results. If the non-availability of a test result has prognostic importance or is related to outcome on the new therapy, then the rule specified for how to treat patients on the test-guided arm who do not have a test result will determine the nature of the bias. |

SPURCES: Freidlin et al., 2010; McShane, 2011; Sargent et al., 2005; Simon et al., 2009. Adapted from Freidlin et al., 2010.

Before a decision has been made to initiate a clinical study or clinical trial to assess the clinical utility of a new omics-based test, the test should be fully defined and validated as described in Chapter 3, and should not change during the clinical study or clinical trial. The committee recommends that **omics-based tests should not be changed during the clinical trial without a protocol amendment and discussion with the FDA. A substantive change to the omics-based test may require that the study be restarted (Recommendation 3b).** Both the Common Rule[3] and FDA regulations[4] require investigators to notify the IRB when a change to a trial protocol is made, and patient recruitment is paused until a decision can be made about whether the changes pose minimal risks to patients. The FDA should also be notified if the test is modified. A modest change, such as simplification of the assay method, that does not alter the performance or results of the test would likely not affect the outcome of the trial, but if the change could alter the accuracy or repeatability of the test results, the trial results may not be reliable.

*Prospective–Retrospective Studies*

Ideally, determination of clinical utility is derived from a prospective clinical trial. However, as stated by Simon, Paik, and Hayes, "In the case of tumor markers, practice guidelines and the availability of other diagnostic procedures can sometimes make it very difficult to perform new clinical trials because such trials may involve withholding of therapy that is considered standard of care. Even when they are considered ethical, such trials usually require many years to conduct and are quite expensive" (Simon et al., 2009, p 6). As with any randomized clinical trial, for randomization to be ethically acceptable, there must be equipoise between the two arms, as determined by appropriate review committees (IRB and scientific peer review). As noted by Freidlin et al., 2010, monitoring such trials can also be quite complex because there may be multiple potentially overlapping patient subgroups and/or multiple hypotheses under consideration. "To protect patient interests, it may be necessary to stop the trial (or some of its components) before all of the study objectives are definitively addressed. Conventional monitoring rules that are based on the observed treatment effect in the overall randomized population may often not be sensitive enough for timely stopping based on biomarker subgroup–specific trends in treatment effect" (Freidlin et al., 2010, p. 157). Nonetheless, such trials have been and are being conducted to test new genomic assays (such as the TailoRX, MINDACT, and RxPONDER trials in breast cancer).

However, given the difficulty of conducting prospective trials, it has been proposed that high quality[5] evidence to assess the clinical utility of a new omics-based test may be obtained by conducting prospective–retrospective studies using archived specimens from previously conducted prospective clinical trials or cohort studies that addressed the intended clinical use of the test (Pepe et al., 2008; Simon et al., 2009). However, even in this case, the investigators should have a prospective written protocol describing their objectives, methods, and analytical plan, and the test should be "locked down"− fully defined, validated and not changed during the study as described in Chapter 3.

[3] 45 CFR part 46 §46.103(4) (2009).
[4] 21 CFR part 56 § 56.108(a)(2011).
[5] Simon et al., (2009) define levels of evidence to assess tumor biomarkers. However, the approach to systematic reviews and guidelines development currently is moving away from "levels of evidence" rating systems and toward a more comprehensive approach to synthesizing the quality of a body of evidence (IOM, 2011).

The requirements to achieve high-quality evidence to assess the clinical utility of a new test using the prospective–retrospective pathway illustrated in the far left panel of Figure 4-1 are described elsewhere (Simon et al., 2009). Simon and colleagues explain the use of the prospective–retrospective study:

> Many biomarker studies are conducted with convenience samples of specimens, which just happen to be available and are assayed for the marker, with no prospectively determined subject eligibility, power calculations, marker cut-point specification, or analytical plans. Such studies are very likely to result in highly biased conclusions and truly deserve to be pejoratively labeled as *retrospective*. However, if a *retrospective* study is designed to use archived specimens from a previously conducted prospective trial, and if certain conditions are prospectively delineated in a written protocol before the marker study is performed, we argue that it might be considered a *prospective–retrospective* study. Such a study should carry considerably more weight toward determination of clinical utility of the marker than a simple study of convenience, in which specimens and an assay happen to be available. Having multiple studies of different candidate biomarkers based on archived tissues from the same prospective trial would, however, present a greater opportunity for false-positive conclusions than a single fully prospective trial focused on a specific biomarker. Consequently, independent confirmation of findings for specific biomarkers in multiple prospective–retrospective studies is important. (Simon et al., 2009, p. 3)

If appropriate archived specimens from previously conducted clinical trials are not available, investigators need to assess the clinical utility of an omics-based test in prospective clinical trials, as illustrated in the right side of Figure 4-1. Full descriptions of tumor biomarker trial designs have been published previously (Freidlin et al., 2010; Sargent et al., 2005; Simon, 2010) and are applicable to the evaluation of a new omics-based test. Prospective clinical trials in which an omics-based test either is or is not used for patient management decisions in the clinical trial are discussed in the next sections.

### *Prospective Clinical Trials Where the Omics-Based Test Is Not Used for Patient Management*

To establish clinical utility, the study must be properly designed and powered to address the intended clinical use of the omics-based test. One approach is to perform clinical trials in which the test is not used to direct therapy, but the primary objective of the clinical trial is to assess the clinical usefulness of the test for its intended use. Two designs can be used (Figure 4-2):

- Completely randomized design; or
- Test Stratified Design (Randomized marker-block design)

In the completely randomized design, the test result is not used in the randomization nor is patient accrual stratified according to the test results. In this case, test results can be generated in real-time or at the completion of the trial accrual. However, for such a trial to adequately assess clinical utility, investigators should follow the previous committee recommendations regarding test development, analytical validity, clinical/biological validity, and design and conduct of a properly powered study to address the specific intended use of the test.

In contrast, in the test stratified design, the test result needs to be available at the time of screening patients for accrual, and the result is used to stratify the randomization of patients to

arms of the trial. Of these two designs, the approach that stratifies the randomization would be preferred when the test result is strongly prognostic and there would be interest in reducing the risk of confounding effects in the overall analysis that does not adjust for the test result. However, stratification of the randomization is not necessary in settings where the numbers of test-positive and test-negative patients is large because in this situation the randomization should be sufficient to achieve approximate balance. In such a design, the test can be evaluated as a prognostic factor in the arm that does not receive the experimental therapy, and as an effect modifier when comparing outcomes in the arm that did receive the experimental therapy versus the standard of care (control).

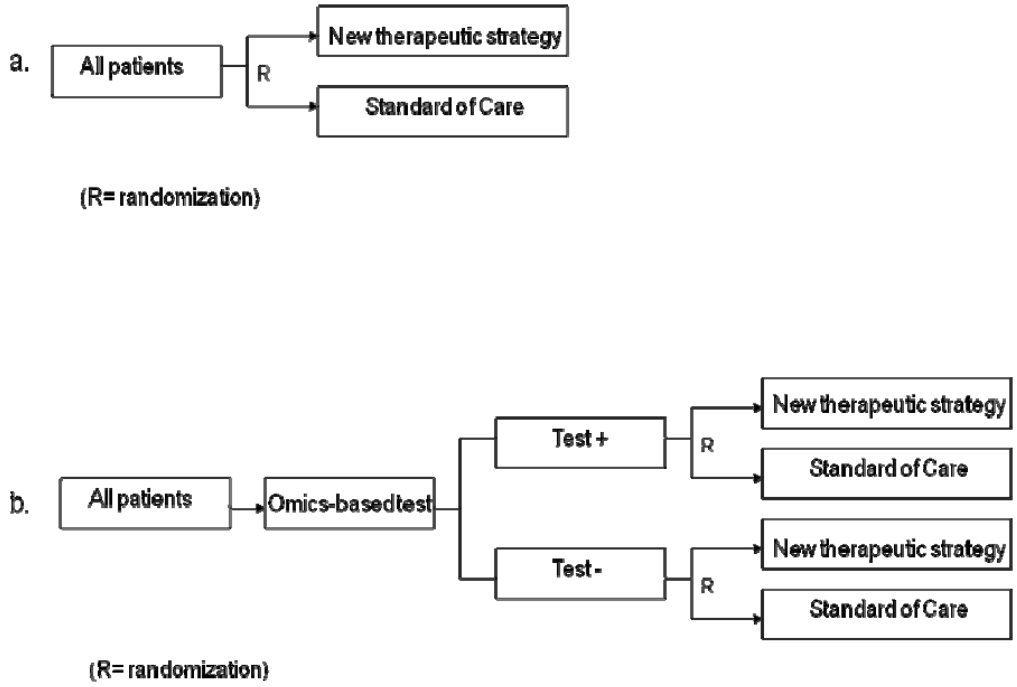


**FIGURE 4-2** Two clinical trial designs in which the test is not used to direct therapy, but the primary objective of the clinical trial is to assess the clinical usefulness of the test, given its intended use. a. Completely randomized design, where the omics-based test is used on all patients, but where the results are not used for randomization. b. Test-stratified design, where the omics-based test is used prerandomization, and patients are stratified based on the results and then randomized to a new treatment strategy or standard of care.
SOURCE: Adapted from McShane (2011).

Treatment effect modification is assessed by the evidence that the magnitude of treatment effect is dependent on the results of the omics-based test. More precisely, the clinical utility of treatment effect modifier biomarkers is established by evidence that an intervention provides a clinically meaningful improvement in the benefit-to-risk profile of an intervention only in certain subgroups identified by the test result. While such analyses need to be prespecified in the statistical analysis plan, stratifying the randomization has little impact on the power of such analyses. Hence, the two designs in Figure 4-2 are equally well suited to assess whether the biomarker is an effect modifier of treatment.

*Prospective Clinical Trials Where the Omics-Based Test Is Used for Patient Management*

The design that most directly assesses the clinical utility of a new omics-based test is a prospective clinical trial in which the omics-based test under study is used to guide patient management decisions. However, this design also poses the highest potential risk to patients due to the possibility that the test could be wrong. Use of a new test to direct patient management in a clinical trial requires the use of a defined and validated omics-based test performed in a CLIA-certified clinical laboratory (See Chapter 3).

Many prospective clinical trial designs have been proposed to evaluate potential treatment effect modifiers to guide the selection of patients to receive therapeutic interventions. In some instances, these effect modifier tests are studied in parallel with the development of a new therapeutic, an approach designated "co-development" by the FDA (Woosley and Cossman, 2007). In such settings, the version of the treatment effect modifier test that would be used in clinical practice ideally should be assessed in the same clinical trial designed to evaluate the new therapeutic. Draft guidance on the development of a therapeutic product that depends on an in vitro "companion" diagnostic device was recently distributed by FDA (FDA, 2011).

Several designs for evaluating treatment effect modifier tests have been proposed (Freidlin et al., 2010; Sargent et al., 2005). The optimal design depends on whether the test will be used to guide decisions about treatment selection for patients in the clinical trial (see Freidlin et al., 2010; specific advantages and disadvantages of these study designs discussed are provided in Table 1 of this reference). Given that a control group is needed to determine whether a test is a prognostic factor and whether it is a treatment effect modifier, designs of randomized trials are discussed rather than single-arm trials.[6]

**Enrichment design** In some cases, investigators may assume that the utility of one of the test-designated categories is established, or likely to be, while the others are uncertain and require prospective evaluation in a clinical trial. For example, investigators might determine that the clinical utility of the prognostic role of a biomarker test is established, but that the treatment effect modifier role for subsequent therapy in those patients who fall into the "poor" prognostic group still is investigational. In this case, patient groups designated as having a very favorable prognosis should not receive treatment, and therefore they are not eligible for a subsequent trial that addresses the benefit in those patients who are presumed to have a worse prognosis. The case study of Onco*type*DX, and the TailoRX study designed to assess its clinical utility, illustrates such a case (see Appendix A).

Likewise, investigators might conclude from prior preclinical or clinical studies that the negative predictive value of an effect modifier is established, but that the benefit of a targeted therapeutic agent still is investigational in those patients who have positive test results. In this case, patients who fall into a "negative" effect modifier/predictive factor subgroup might be considered to be so unlikely to respond that they are not eligible for a subsequent trial that addresses the utility of the investigational agent, and only patients with positive test results are enrolled.

[6] Note that single-arm trials are sometimes used for FDA decisions. For example, in the case of codevelopment of the crizotinib with ALK FISH, the drug accelerated approved on the basis of results from two single arm trials (FDA, 2011). While a Phase III study is underway, the FDA approved the drug because the FDA concluded that the clinical data was reasonably likely to predict a clinical benefit to patients. The objective response rate was 50-60 percent, with median response duration of 42-48 weeks. Approximately 3-5 percent of lung adenocarcinomas contain ALK fusions.

An enrichment trial design (Freidlin et al., 2010; Sargent et al., 2005) is illustrated in Figure 4-3. With this approach, the only patients entered into the clinical trial are those who test positive for the biomarker at screening. If there is reliable evidence that the new drug would not be effective in patients with negative test results, then this approach would be both efficient and ethical by limiting the assessment of efficacy of the new drug to the target population with an anticipated favorable benefit-to-risk profile.

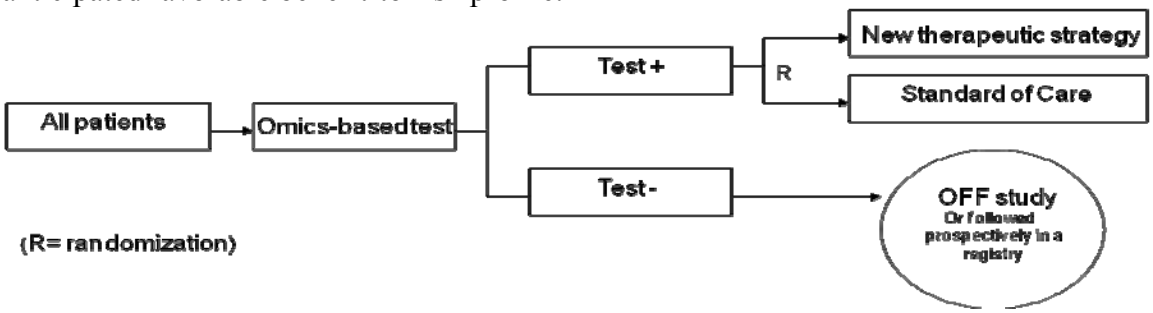


**FIGURE 4-3** Enrichment design, where only test-positive patients are randomized and treated. SOURCE: Adapted from McShane (2011).

Investigators should be aware that if there are uncertainties about whether the new drug might be effective in patients with negative test results, then the enrichment design is deficient because it provides no insight to address such uncertainties. Patients with negative test results could benefit from the new drug if the test does not reliably identify the target population, for example, by having the wrong threshold for defined positivity, or if the new drug has important mechanisms of action, not captured by the test, that result in favorable efficacy in patients with both positive and negative results. The development of trastuzumab (Herceptin), a humanized monoclonal antibody against HER2, for breast cancer provides such an example (see case study in Appendix A). Preclinical data strongly suggested that trastuzumab would only be effective against cancers that strongly overexpress HER2, and nearly all subsequent clinical trials have only enrolled patients with very high levels of HER2 protein and/or with *HER2* gene amplification (Wolff et al., 2007). However, post-hoc central analysis of HER2 in patients entered into two of the largest prospective randomized trials of adjuvant trastuzumab identified a small group of women whose breast cancers appear not to be HER2 positive by classic criteria, yet the hazard ratio for recurrence is equally favorable in these patients as it is for women with highly overexpressed or amplified HER2 (Paik et al., 2008; Perez et al., 2010). These results are the basis of an ongoing clinical trial testing the benefits of trastuzumab in patients who have lesser, but still detectable, HER2 expression (NSABP clinical trial B47).

**Test-guided strategy with control designs** Frequently, investigators cannot assume that certain test-designated categories are established, and in these cases an enrichment design is not appropriate. In this case, the trial is designed to "test the test," using appropriate control arms. There are at least two such trial designs.

Figure 4-4 presents the test-guided treatment strategy where patients are randomized between test-guided use of a new drug versus the application of a standard of care control regimen without regard to the test.

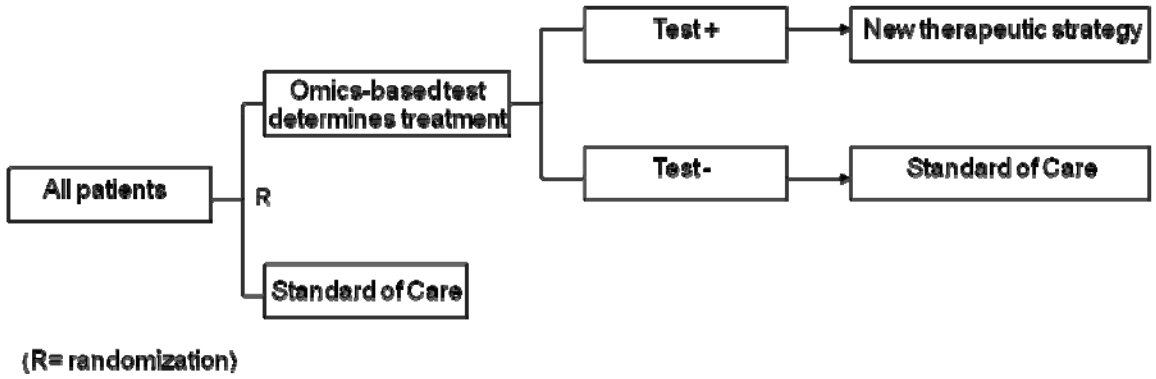


**FIGURE 4-4** Example of a test-guided strategy versus standard of care, where patients are randomly assigned to the test-guided arm or the non-guided arm.
SOURCE: Adapted from McShane (2011).

Figure 4-5 presents an alternative test-guided treatment strategy in which patients are randomized to an arm in which the test is used to guide selection of an investigational or standard of care management strategy (upper arm of Figure 4-5) vs. an arm in which patients are randomly assigned to investigational therapy vs. control standard of care management without knowledge of the test results. For example, this was the design used in the prospective Duke trial of preoperative chemotherapy for breast cancer included in the committee's statement of task (NCT00636441; see Appendix B). In contrast to the design in Figure 4-4, this design is less efficient if the control regimen were to remain as the standard of care, but would be useful if the new drug could emerge as the standard of care regimen for all subjects in this clinical setting. However, Freidlin and colleagues argue that both of these test-guided treatment strategies are inherently less efficient than either of the designs in Figure 4-2 (Freidlin et al., 2010), which estimate the treatment effect in all relevant biomarker populations.

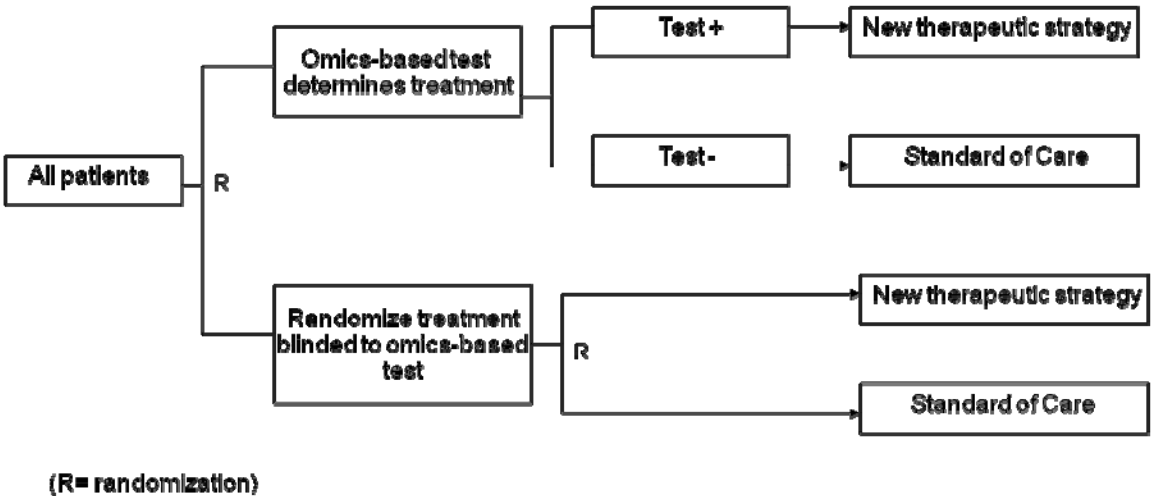


**FIGURE 4-5** Example of a test-guided strategy versus non-guided strategy with randomized treatment design. Patients are randomly assigned to use of the omics-based test to determine treatment versus randomized treatment selection (i.e., patients are tested but treatment is randomly assigned without consideration of the test result).

SOURCE: Adapted from McShane (2011).

## RECOMMENDATION

After analytical and clinical/biological validation of the candidate omics-based test, the test is ready for use in a clinical trial to assess clinical utility and use of the test in patient care. Figure 4-1 shows several options for clinical trial designs to assess clinical utility. Depending on the choice of trial design, investigators may need to obtain an IDE from the FDA. Regardless of choice, however, the committee strongly recommends consulting with the FDA prior to initiation of clinical trials. In the case of a trial in which patient management will be influenced by the omics-based test findings, obtaining an IDE from the FDA is a legal requirement. In other cases, the committee recommends consultation with the FDA because the requirement for an IDE based on the trial design is not always clear. In addition, if the test will later need clearance or approval from the FDA before marketing for clinical use, the study design and analysis will be subject to FDA review, and a pre-IDE consultation can assist both the test developer and the FDA in coming to an agreement on the data necessary for FDA clearance or approval. Critical considerations for moving a candidate omics-based test into clinical trials for assessing clinical utility are outlined in Figure 4-1 and Recommendation 3, below.

**RECOMMENDATION 3: Evaluation for clinical utility and use stage.**
**For investigators conducting a clinical trial to assess the clinical utility and use of an omics-based test that has been validated as described in Chapter 3, the committee recommends that:**

- **a. Investigators should communicate early with the FDA regarding the Investigational Device Exemption (IDE) process and validation requirements.**
- **b. Omics-based tests should not be changed during the clinical trial without a protocol amendment and discussion with the FDA. A substantive change to the omics-based test may require that the study be restarted.**

## REFERENCES

AHRQ (Agency for Healthcare Research and Quality). 2011. *Technology assessments*. http://www.ahrq.gov/clinic/techix.htm (accessed October 20, 2011).

Allegra, C. J., J. M. Jessup, M. R. Somerfield, S. R. Hamilton, E. H. Hammond, D. F. Hayes, P. K. McAllister, R. F. Morton, and R. L. Schilsky. 2009. American Society of Clinical Oncology provisional clinical opinion: Testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *Journal of Clinical Oncology* 27(12):2091-2096.

Altman, D. G., and G. H. Lyman. 1998. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Research and Treatment* 52(1-3):289-303.

Altman, D. G., and R. D. Riley. 2005. Primer: An evidence-based approach to prognostic markers. *Nature Clinical Practice Oncology* 2(9):466-472.

Altman, D. G., and P. Royston. 2000. What do we mean by validating a prognostic model? *Statistics in Medicine* 19(4):453-473.

FDA (Food and Drug Administration). 2011 a. News Release: FDA approves Xalkori with companion diagnostic for a type of late-stage lung cancer. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm269856.htm.

FDA. 2011 b. Draft Guidance for Industry and Food and Drug Administration Staff- In Vitro Companion Diagnostic Devices. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm262292.htm

Freidlin, B., L. M. McShane, and E. L. Korn. 2010. Randomized clinical trials with biomarkers: Design issues. *Journal of the National Cancer Institute* 102(3):152-160.

Gutierrez, A. 2011. *Discussion with the Committee on the Review of Omics-based Tests for Predicting Patient Outcomes in Clinical Trials*, Washington, DC, August 19.

Harris, L., H. Fritsche, R. Mennel, L. Norton, P. Ravdin, S. Taube, M. R. Somerfield, D. F. Hayes, and R. C. Bast, Jr. 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of Clinical Oncology* 25(33):5287-5312.

Hayes, D. F., R. C. Bast, C. E. Desch, H. Fritsche Jr, N. E. Kemeny, J. M. Jessup, G. Y. Locker, J. S. Macdonald, R. G. Mennel, L. Norton, P. Ravdin, S. Taube, R. J. Winn. 1996. Tumor marker utility grading system: A framework to evaluate clinical utility of tumor markers. *Journal of the National Cancer Institute* 88(20):1456-1466.

Henry, N. L. and D. F. Hayes. 2006. Uses and Abuses of Tumor Markers in the Diagnosis, Monitoring, and Treatment of Primary and Metastatic Breast Cancer. *The Oncologist* 11:541-552.

IOM (Institute of Medicine). 2011a. *Finding What Works in Health Care: Standards for systematic reviews*. Washington, DC: National Academies Press.

IOM. 2011b. *A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program*. Washington, DC: National Academies Press

Locker, G. Y., S. Hamilton, J. Harris, J. M. Jessup, N. Kemeny, J. S. Macdonald, M. R. Somerfield, D. F. Hayes, and R. C. Bast, Jr. 2006. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *Journal of Clinical Oncology* 24(33):5313-5327.

McShane, L. 2011. *Challenges in the Development and Validation of Biomarker-Based Tests for Personalized Therapeutic Decision Making in Oncology*. Presented at Accelerating Anticancer Agent Development and Validation Workshop, Bethesda, MD. May 19.

NCCN (National Comprehensive Cancer Network). 2010. *NCCN Clinical Practice Guidelines in Oncology*. http://www.nccn.org (accessed 2010).

Paik, S., C. Kim, and N. Wolmark. 2008. HER2 status and benefit from adjuvant trastuzumab in breast cancer. *New England Journal of Medicine* 358(13):1409-1411.

Pepe M. S., Ziding F., Janes H., Bossuyt P. M., and Potter J. D. 2008. Pivotal Evaluation of the Accuracy of a Biomarker Used for Classification or Prediction: Standards for Study Design. J *Journal of the National Cancer Institute*. 100 (20):1432 – 1438.

Perez, E. A., M. M. Reinholz, D. W. Hillman, K. S. Tenner, M. J. Schroeder, N. E. Davidson, S. Martino, G. W. Sledge, L. N. Harris, J. R. Gralow, A. C. Dueck, R. P. Ketterling, J. N. Ingle, W. L. Lingle, P. A. Kaufman, D. W. Visscher, and R. B. Jenkins. 2010. HER2 and chromosome 17 effect on patient outcome in the N9831 adjuvant trastuzumab trial. *J Clin Oncol* 28(28):4307-4315.

SACGHS (Secretary's Advisory Committee on Genetics, Health , and Society). 2008. *U.S. System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services; a Report of the Secretary's Advisory Committee on Genetics, Health, and Society*. http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf.

Sargent, D. J., B. A. Conley, C. Allegra, and L. Collette. 2005. Clinical trial designs for predictive marker validation in cancer treatment trials. *Journal of Clinical Oncology* 23(9):2020-2027.

Simon, R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Personalized Medicine* 7(1):33-47.

Simon, R. 2005a. Development and validation of therapeutically relevant multi-gene biomarker classifiers. *Journal of the National Cancer Institute* 97(12):866-867.

Simon, R. 2005b. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 23(29):7332-7341.

Simon, R., and D. G. Altman. 1994. Statistical aspects of prognostic factor studies in oncology. *British Journal of Cancer* 69(6):979-985.

Simon, R. M., S. Paik, and D. F. Hayes. 2009. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *Journal of the National Cancer Institute* 101(21):1446-1452.

TEC (Technology Evaluation Center). 2011. *Technology Evaluation Center Assessment Process*. http://www.bcbs.com/blueresources/tec/tec-assessment-process.html (accessed October 20, 2011).

Teutsch, S. M., L. A. Bradley, G. E. Palomaki, J. E. Haddow, M. Piper, N. Calonge, W. D. Dotson, M. P. Douglas, and A. O. Berg. 2009. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group. *Genetics in Medicine* 11(1):3-14.

USPSTF (U.S. Preventive Services Task Force). 2009. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Annuals of Internal Medicine*. 151(10):716-726, W-236.

USPSTF. 2011. *Screening for Prostate Cancer: U.S. Preventive Services Task Force Draft Recommendation Statement*. http://www.uspreventiveservicestaskforce.org/draftrec3.htm. (accessed December 5, 2011).

Wolff, A. C., M. E. Hammond, J. N. Schwartz, K. L. Hagerty, D. C. Allred, R. J. Cote, M. Dowsett, P. L. Fitzgibbons, W. M. Hanna, A. Langer, L. M. McShane, S. Paik, M. D. Pegram, E. A. Perez, M. F. Press, A. Rhodes, C. Sturgeon, S. E. Taube, R. Tubbs, G. H. Vance, M. van de Vijver, T. M. Wheeler, and D. F. Hayes. 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 25(1):118-145.

Woosley, R. L., and J. Cossman. 2007. Drug development and the FDA's Critical Path Initiative. *Clin Pharmacol Ther* 81(1):129-133.

# 5
# Responsible Parties

Omics research has ushered in a new era in research. Omics data are extremely complex and multidimensional, with a high risk of inaccuracies being introduced by inappropriate methods, human error, conflicts of interest, or acts of commission/omission. Omics research requires a multidisciplinary team with specialized expertise, which adds to the challenge of conducting scientifically rigorous research and makes overseeing and reviewing omics studies difficult. This multidimensionality introduces an inherent risk of over-fitting the data, making independent validation critical. While other fields such as high energy physics, astrophysics, and cosmology also require specialized expertise and multidisciplinary collaboration, and deal with data complexity and high dimensionality, the development of omics-based clinical tests is different in that the tests have potential commercial value and there is potential for developers to reap financial gains. In addition, patient safety is paramount for omics-based tests that are used to aid patient treatment decision. Although these characteristics are also true of drug development, that process has more uniform and more stringent oversight from the FDA; all new drugs must demonstrate clinical utility in well-designed clinical trials to gain FDA approval. Thus, those responsible for the integrity of omics research—investigators, institutions, funders, the U.S. Food and Drug Administration (FDA), and journals—should rethink the processes and protections designed to ensure that omics research is scientifically rigorous, transparent, and conducted ethically and with proper institutional and regulatory oversight.

The failures of the omics research at Duke University illustrate that current practices and safeguards can easily fall short (see Appendix B). The Duke events, thus, provide a watershed illustration—reminiscent of the Gelsinger gene therapy cases at the University of Pennsylvania, the Santillan mismatched heart transplant case at Duke University Hospital, the Johns Hopkins asthma trial death, and the viral link to chronic fatigue syndrome at the Whittemore Peterson Institute for Neuro-Immune Disease—of how such research can go awry even though institutions and other responsible parties have extensive systems in place to ensure research integrity, with roles and responsibilities delineated (Enserink, 2011; Kolata, 2001; Nelson and Weiss, 1999; Sloane, 2003; Yarborough and Sharp, 2009). These processes need to be rethought in the omics era. In short, the ability of healthcare decision makers to rely on the trustworthiness of omics-based tests to predict disease risk and treatment response will be limited unless renewed efforts are made by all parties responsible for the integrity of this research.

The committee makes four recommendations related to defining the roles and responsibilities of the key parties involved in the conduct and evaluation of omics research (Recommendations 4-7). These recommendations are directed toward investigators and institutions (i.e., intrainstitutional responsibilities), funders, the FDA, and biomedical journals. Recommendations 1 through 3, which are discussed in Chapters 2 to 4, refer to responsibilities of investigators, but focus on recommended best practices for the development, validation, and clinical utility assessment of candidate omics-based tests. Recommendations 4 to 7 are similarly critical because, without the participation of institutions, investigators, funders, the FDA, and journals, the committee's recommended evaluation processes for omics technologies intended for clinical use (Recommendations 1-3) cannot be implemented. The committee recognized that

the recommendations presented in this chapter may increase the oversight requirements for omics research in some cases, but decided that these potential costs were offset by the added safeguards to the integrity of this research. If an institution does not have the infrastructure or capability to follow the recommended Test Development and Evaluation Process defined in this report, then the committee believes that the institution should consider not engaging in the translation of omics-based discoveries into validated tests for use in clinical trials and potentially clinical practice.

The committee developed the recommendations discussed in this chapter by reviewing the available literature about the design, conduct, analysis, and reporting of omics research and by identifying lessons learned from case studies of the development of omics-based tests (see Appendixes A and B). This chapter emphasizes lessons learned from the Duke University case study (Appendix B), in particular, because the most publicly available information exists about this case study and because the Duke case was specifically highlighted in the committee's statement of task. The committee also relied heavily on the work of previous National Academies reports that have reviewed the roles and responsibilities of the parties involved in research. It is imperative that all responsible parties prepare for the omics research era, with its promise and its perils. This chapter discusses the details of how this preparation can be accomplished.

## INTRAINSTITUTIONAL PARTIES

The roles and responsibilities of investigators and institutions that are involved in omics-based research are discussed together because both parties contribute to the scientific research culture in which omics research is conducted. They are also the two most responsible and the most knowledgeable parties in the entire evaluation process. Investigators control the culture of individual laboratories embedded within the larger institution. Individual laboratories can have unique values and cultural norms that are separate from the broader institutional culture. These variables become more complex as the research becomes more interdisciplinary, with the lead investigators setting the "culture" for the investigational team. Institutions and the institutional leadership, on the other hand, have the primary responsibility for the policies and procedures, reward systems, and values that contribute to the overarching institutional culture as well as for the infrastructure of oversight and support for research. Institutions and their leaders also have the greatest responsibility for in-depth investigation of potential lapses in scientific integrity because they employ, promote, and supervise the investigators who conduct these studies.

The National Academies defined integrity in the research process as "the adherence by scientists and their institutions to honest and verifiable methods in proposing, performing, evaluating, and reporting research activities" (NAS, 1992, p. 27). The challenge is that science is a self-regulating community, with few comprehensive guidelines for responsible research practices (Steneck, 2006). The guidelines that do exist often contradict each other (Emanuel et al., 2000). There are inconsistencies in the rules governing the de-identification of personal health information, obtaining individual consent for future research, and the recruitment of research volunteers (IOM, 2009a). The 2011 Report of the Presidential Commission for the Study of Bioethics Issues recommended that the Common Rule be revised to include a section on investigators' responsibilities in order to bring it into harmony with the FDA regulations for clinical research and international standards (PCSBI, 2011). Moreover, when ethical standards and best practices are available to guide behavior, some investigators may be ignorant of these rules, or simply breach them. For example, Martinson and colleagues (2005) conducted a series of focus groups with investigators from top-tier research universities to identify the top 10

misbehaviors of greatest concern in science. They then surveyed more than 7,000 early- and mid-career U.S. investigators who have funding from the National Institutes of Health (NIH) and asked them to report on their own behavior. Thirty-three percent of the respondents reported engaging in at least one of the 10 misbehaviors during the previous 3 years. The three most common misbehaviors were: (1) overlooking other researchers' use of flawed data or questionable interpretations of data; (2) changing the design, methodology, or results of a study in response to pressure from a funding source; and (3) circumventing certain minor aspects of human-subjects research requirements (Martinson et al., 2005). This situation is problematic because the underlying science must be sound if patients are going to participate in clinical trials and, eventually, in consultation with their physicians, use research results for medical care decisions.

## Investigators

Responsible conduct in any research, including omics research, starts with the investigators. This includes both junior and senior investigators. This section of the chapter describes the roles and responsibilities of investigators who conduct biomedical omics research with the goal to improve patient care. These responsibilities include the most basic principles of science, such as a serious and in-depth consideration in a Discussion section of a journal article of "what might be wrong with the data and conclusions I have just reported" (Platt, 1964). The specific responsibilities discussed below include fostering a culture of scientific rigor and welcoming constructive criticism, comprehensively reporting the methods and results of a study and making data and code publicly available so that a third party could verify the data and result. Box 5-1 highlights themes extracted from several representative case studies for investigators to consider.

**BOX 5-1**
**Themes from the Case Studies for Investigators**

**The Duke Case Study**

Several questions have emerged regarding the degree to which key tenets of scientific rigor (for both laboratory-based research and clinical trials) were followed in the Nevins laboratory at Duke. First, there were numerous errors in the primary data (Baggerly and Coombes 2009; Coombes et al., 2007). Predictors derived from the training datasets were not locked down, leading to flaws in the validation process and the omics-based tests that were developed. Second, major results in the papers published by the Duke investigators were not reproducible. For example, figures in the Hsu et al. paper could not be reproduced with the data provided (McShane, 2010a). Third, it is unclear whether validation was based on blinded data. The *Lancet Oncology* paper stated that the validation was blinded, but the investigators had access to unblinded data. (Goldberg, 2009). Fourth, the Duke investigators did not provide the public with full access to their data and code (Baggerly and Coombes, 2009; Baron et al., 2010). They also failed to address the questions and challenges of external investigators who were trying to reproduce their work to the mutual satisfaction of all parties involved (Baggerly, 2011; McShane, 2010b). In response to the National Cancer Institute's investigation, the Duke investigators acknowledged that their tests were unreproducible and retracted the original papers (Bonnefoi et al., 2011; Hsu et al., 2010; Potti et al., 2011). Dr. Joseph Nevins, senior mentor of the investigators whose genomic predictors were used in the three clinical trials named in the statement of task, stated during discussions with the committee that "a critical flaw in the research effort was

one of data corruption" (Nevins, 2011). Throughout this process, the responsibilities of the coinvestigators on the research team and lines of accountability were apparently unclear.

**The Ovacheck Case Study**

The investigators of the Ovacheck test made their datasets publicly available. Independent investigators conducted analytical, statistical, and algorithmic validation studies and showed that the test was unreproducible (Baggerly et al., 2004). Thus, in this case, making the data and code publicly available helped prevent the routine clinical use of an unvalidated diagnostic tool.

**Commercially Developed Tests: Data and Code Availability**

A review of the six commercially available tests discussed in Appendix A illustrates that public availability of all omics-based test data has not been standard practice. The field of omics is early in its development and the standards for data sharing have been unclear and only now slowly evolving toward more transparency. Commercial interests and protection of proprietary information also may have limited the public availability of some data and information.

These six cases highlight several examples in which test developers explicitly note the availability of data. For example, Paik et al. (2004), Deng et al. (2006), and Rosenberg et al. (2010) reported the computational model for Onco*type* DX, AlloMap, and Corus CAD, respectively. Both tests developed as LDTs had published computational models (Oncotype DX and Corus CAD); only one FDA-cleared test has a published computational model (AlloMap). Discovery microarray data are available for MammaPrint, AlloMap, and Corus CAD (Deng et al., 2006; van't Veer et al., 2002).* Buyse et al. (2006) reports that raw microarray data and clinical data for the MammaPrint clinical validation study were deposited with the European Bioinformatics Institute ArrayExpress database. Although there are examples of developers reporting the availability of a test's computational model or data used in discovery or validation, often sufficient information is not publicly available for external investigators to fully reproduce a test.

NOTE: See Appendixes A and B on the case studies for more information.
* Microarray data from Corus CAD are available, but PCR data used in test development are unavailable. Personal communication, Steve Rosenberg, October 21, 2011.

*Culture*

All investigators have a responsibility to promote a culture of scientific rigor and to transmit ethical principles of science to future generations of investigators. Scientific rigor can be fostered by developing clear standards of behavior, disseminating those standards through education and mentoring, and reinforcing the standards through exemplary practice at all levels of the research community (Frankel, 1995). Investigators who do not adhere to these values are not fulfilling their ethical responsibilities. Although many cultural issues are not unique to omics research, taking steps to improve scientific culture is particularly important in omics research because of the nature of omics discoveries, which depend on large datasets, complex analyses, and a specialized multidisciplinary team.

A number of influential reports have recommended sets of values, traditions, and standards that investigators should embody to promote a culture of scientific rigor. The National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine (IOM) collaborated in producing the report, *Responsible Science, Volume I: Ensuring the Integrity of the Research Process* (NAS, 1992). This report highlighted the importance of

investigators upholding the highest standards of honesty, integrity, objectivity, and collegiality. The authoring committee directed individual investigators to accept formal responsibility for ensuring the integrity of the research process and creating an environment, a reward system, and a training system that encourage responsible research practices. A more recent National Academies report, *On Being a Scientist: A Guide to Responsible Conduct in Research* (NAS, 2009), identified three sets of obligations for investigators: (1) an obligation to merit the trust that their colleagues place in them (i.e., science is cumulative and investigators build on previous work); (2) an obligation to themselves (i.e., investigators should adhere to professional standards and develop personal integrity); and (3) an obligation to act in ways that serve the public (i.e., the public uses science to make policy decisions). The Office of Research Integrity (ORI) of the Department of Health and Human Services (HHS) also has outlined several values that investigators should share in promoting a culture of scientific rigor, including (1) honesty: conveying information truthfully and honoring commitments; (2) accuracy: reporting findings precisely and taking care to avoid errors; (3) efficiency: using resources wisely and avoiding waste; and (4) objectivity: letting the facts speak for themselves and avoiding improper bias (Steneck, 2006). These reports outline general guiding principles for investigators' behavior. However, identifying the values and obligations that investigators should possess does not directly inform investigators on how they should respond in specific situations and conflicts. Ultimately, investigators' actions need to be informed by good judgment and personal integrity.

Two of the major influences on the development of investigators' values and integrity are advisors and mentors (Bird, 2001; NAS, 1992, 2009), who define, explain, and exemplify scientific norms and ethics. All members of the research team, including biostatisticians and bioinformatics scientists, should have access to mentors with the appropriate expertise and credentials. Senior investigators' conduct can reinforce or weaken the importance of complying with these scientific norms and values. Sprague and colleagues (2001), for example, conducted a study to identify the methods by which ethical beliefs are passed on to students. They surveyed faculty and graduate students and asked respondents to rank methods of teaching about ethics; 1,451 surveys were distributed to faculty, and 627 were returned (45.2 percent return rate). An additional 6,000 surveys were sent to academic departments to be distributed to graduate students, and 1,152 were returned (19.2 percent return rate). A major weakness of this study is the low response rates. However, both faculty and students ranked courses dealing with ethical issues as most influential in teaching students ethical beliefs. Mentors in graduate school also were highly ranked, with graduate students ranking mentors as more important than faculty did. Other important influences included discussions in courses, laboratories, and seminars as well as interactions with other graduate students (Sprague et al., 2001). In other words, young investigators' interactions with other investigators shaped their beliefs and values.

Another important component of promoting a scientifically rigorous culture, which falls to investigators, is valuing teamwork and mutual respect and empowering people at lower levels in the hierarchy to speak up if they observe a problem or have a concern regarding research practices. The aviation and energy industries provide evidence for the pivotal importance of creating cultures that value these characteristics and consistently expect and laud persons who speak up to alert the group to problems and concerns. For example, the aviation industry has recognized for some time that errors are more likely to happen when there is suboptimal teamwork and communication (Helmreich, 2000). Thus, improvements in aviation safety have been attributed to training crews on how to address and prevent human error, the role of leadership, the need for monitoring and crosschecking decision-making processes, and the use of checklists. This same

approach has been applied successfully to the patient safety improvement movement to reduce the effects of human errors (Gawande, 2009; Hudson, 2003; Longo et al., 2005; Pronovost et al., 2003) and can be applied equally to the biomedical research enterprise.

*Full Reporting*

Fully reporting the methods and results of a study is essential for the reproducibility of research and for reviewers' and readers' evaluation of the validity of a study. Thus, investigators have a fundamental responsibility to provide a complete and accurate report of their methods and findings (NAS, 1992, 2009; Steneck, 2006). All publications- and omics publications in particular- should present a full and detailed description of the study methodology, the statistical analysis plan that was finalized before the validation data were analyzed, an accurate report of the results, and an honest assessment of the findings, including an explanation of limitations that may affect the conclusions (Platt, 1964; Steneck, 2006). This level of transparency should allow an independent third party to verify the data and results.

As discussed in Appendix D, reporting guidelines are tools to help investigators meet this obligation and report the essential information and elements of a study. All investigators who are coauthors on a report– and particularly a senior investigator or mentor – also are responsible for understanding the specific aims, methods, major findings, and implications of the interdisciplinary research. They are responsible for reading the complete manuscript, suggesting edits, and for being alert to misinterpretation, such as misrepresentation of findings and limitations, and discussing such observations with appropriate members of the team or oversight groups.

*Data and Code Availability and Transparency*

The scientific community widely agrees that investigators should make the research data and code supporting a manuscript, as well as the statistical analysis plan that had been finalized before data were unblinded and available for analysis, publicly available at the time of publication (see discussion in the journals section on the research protocol) (NAS, 1992, 2009; NRC, 1985, 2003). Transparency is essential for the interpretability and reproducibility of research and a tenet of any good scientific method. Indeed, the purpose of methods sections in journal publications is to provide enough detail so that other investigators can interpret the results and, if they wish, reproduce the study and obtain the same results. Thus, providing sufficient detail of methods allows independent investigators to verify published findings and conduct alternative analyses of the same data. It also discourages fraud and helps expedite the exchange of ideas (Peng et al., 2006). Investigators who refuse to share the evidentiary basis behind their conclusions, or the materials and analytical methods needed to replicate published experiments, fail to uphold transparency as a basic standard of science. In an era when much of the Methods section and/or elaborate data appear only in the Supplementary Materials section, more attention is needed to guide the reader through well-annotated supplementary material. This problem is perpetuated by the brevity of articles published in the higher impact journals.

The National Academies has issued numerous reports emphasizing the importance of data sharing. *Sharing Research Data* recommended that sharing research data at the time of publication should be a regular practice in science (NRC, 1985). A later report, *Sharing Publication-Related Data and Materials,* developed a uniform principle for sharing integral data and materials expeditiously (NRC, 2003). It recommended that authors include the code, algorithms, or other information that are central to verifying or replicating the claims in a publication. If the

data and code cannot be included in the actual publication (e.g., because the data files are too large), the report recommended that the data and code be made freely available through other means in a format that allows an independent investigator to manipulate, analyze, and combine the data with other scientific data. The report also stipulated that, if publicly accessible repositories for data have been developed and are in general use, the relevant data should be deposited in those repositories. Investigators are responsible for anticipating which materials are most likely to be requested and should include a statement on how to access the materials in the published paper.

In *On Being a Scientist*, the National Academies addressed the challenge of sharing research data in the current environment, where the quantity and complexity of data are increasing and the cost of sharing data is high (NAS, 2009). The complications and cost of sharing large datasets also were recently highlighted in an issue of *Science*, dedicated entirely to data collection, curation, and access issues (*Science* Staff, 2011). The National Academies concluded that, despite these challenges, investigators have a responsibility to develop methods to share their data and materials at the time of publication (NAS, 2009). Investigators may share data through centralized facilities or undertake collaborative efforts to form large databases, such as the databases of Genotypes and Phenotypes (dbGAP), the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL/EBI), the National Library of Medicine's Gene Expression Omnibus (NLM/GEO), Compendia Biosciences, UCSC Gene Browser, and ProteomeXchange. When data undergo extensive analysis as part of a scientific study, the requirements to share those data also include a requirement to share the software, code, and sometimes the hardware used in the analyses (NAS, 2009). Authors can facilitate the use of such information with graphical user interfaces introduced into the dataset, for example, as facilitated through nanoHUB (Klimeck, 2011).

Ultimately, many investigators are unwilling to comply with the requirement to share their data and code. For example, in an article for the *New York Times*, Andrew Vickers, a biostatistician at Memorial Sloan-Kettering Cancer Center, documented his lack of success in requesting cancer data from various investigators from numerous institutions (Vickers, 2008). Vickers also referenced a survey conducted by John Kirwan of the University of Bristol on investigators' attitudes toward sharing data from clinical trials. Three-quarters of the investigators surveyed stated that they were opposed to making original trial data available. They cited several reasons for refusing, such as the difficulty of putting together a dataset and the risk of their data being analyzed using invalid methods. Vickers concluded that investigators are often opposed to the potential use of their data by other independent investigators who may make influential discoveries, and often resist challenges to their conclusions that emerge from new analyses. Investigators may also be resistant to sharing their data and code because of the time and effort needed to curate and annotate a dataset and support other investigators' access to the material.

The obstacles to sharing data and code may seem particularly daunting in omics research. However, the fields of molecular biology and structural biology widely use web-based genomic and proteomic databases (e.g., GenBank and Protein Data Bank) (Brown, 2003). These databases allow investigators to share DNA, amino acid sequences, and protein structure data, and many journals mandate deposition of these data as a condition of publication. Microarray assays do produce an enormous quantity of data (Quackenbush, 2009); as many as 1 million variant positions on the genome across thousands of samples and next-generation RNA sequencing raises further challenges.

The scheme for *Minimum Information About a Microarray Experiment (MIAME)* was created and adopted by investigators in this field to improve the annotation of microarray data (Brazma et al., 2001). It established standard, comprehensive annotation requirements that have been adopted by most scientific journals. Data from more than 10,000 microarray studies have been deposited to date into public repositories designed to archive MIAME-compliant data (Brazma, 2009). MIAME also has stimulated the proteomics and metabolomics scientific communities to develop reporting standards and formats. In fact, the Minimum Information for Biological and Biomedical Investigations (MIBBI) project has cataloged more than 30 different reporting standards for biological and biomedical data (Taylor et al., 2008). Nevertheless, many investigators still fail to provide fully annotated data (Brazma, 2009; Quackenbush, 2009). Thus, further steps need to be taken to ensure investigators share their data and code. The committee's recommendations to journals and funders (discussed below) are intended to create additional incentives for investigators to comply with data- and code-sharing norms. Issues of proprietary information can be dealt with by depositing the materials with a responsible third party that can ensure confidentiality and protection of the material (e.g., the FDA). The patent system also protects private investments in omics research (SACGHS, 2010) (see Box 2-1 for a more detailed discussion on intellectual property law).

### Institutions and Institutional Leaders

This section of the chapter describes the roles and responsibilities for institutions that conduct biomedical omics research aimed at improving patient care, including: fostering a culture of scientific integrity, overseeing research, increasing awareness of reporting systems for lapses in research integrity, investigating credible concerns about scientific integrity, monitoring and managing financial and non-financial conflicts of interest, and supporting and protecting the intellectual independence of biostatisticians, bioinformatics scientists, pathologists, and other collaborators in omics research. These responsibilities lie ultimately with institutional leadership. Indeed, any institutional attempt to meet these responsibilities will fail without explicit and visible support and direction from institutional leadership (Schein, 2004). Some of these responsibilities are closely related to the responsibilities of the investigators.

Institutions, such as universities and companies, and the institutional leaders, in collaboration with their investigators, play an essential role in promoting a culture that encourages investigators to act ethically and conduct scientifically rigorous research. Institutions and their leadership bear direct responsibility for complying with existing rules and regulations governing research; overseeing and creating reward systems for investigators; providing training and education to investigators on relevant topics; and producing an environment of trust, openness, and honesty. The integrity of the research enterprise depends on investigators, collaborators, and observers feeling encouraged and supported when they identify and report either routine scientific disagreements or potential breaches of integrity, regardless of their position within the institution. Institutional leaders also have direct responsibility, when concerns are raised, for establishing and supervising a "process of evaluation" of specific research results and claims by their investigators.

In the Duke University case, inadequacies in the institutional oversight processes and a lack of sufficient checks and balances allowed invalid omics-based tests to progress to clinical trials (see Appendix B). Therefore, the committee believes that explicitly defining the roles and responsibilities of all of the parties involved in omics research is essential to ensuring that omics-based tests are credible and can be used to inform real-world clinical questions. Any overlap in

responsibilities can be an added layer of protection to ensure that omics research is scientifically rigorous, transparent, and conducted with proper oversight. Box 5-2 highlights relevant themes from several case studies for institutions to consider.

**BOX 5-2**
**Themes from the Case Studies for Institutions**

**The Duke Case Study**

Although the three clinical trials named in the statement of task involved cancer patients their operations were not being overseen by the Duke Cancer Center, which has a substantial infrastructure of biostatistics, bioinformatics, and data management support for all studies conducted within its jurisdiction. Rather, these trials were overseen by the Institute of Genomic Sciences and Policy (IGSP). According to Robert Califf, M.D., vice chancellor for clinical research and director of the Duke Translational Medicine Institute, "there was ambiguity" in the lines of authority and oversight in the IGSP during the conduct of the three clinical trials and there were "numerous missed signals" that there were problems with the research (Califf, 2011b). Moreover, there was discontinuity in the statistical team, which may have contributed to the research team's failure to follow proper data management practices (Kornbluth and Dzau, 2011). Junior investigators on the team either did not recognize what was wrong or did not feel comfortable expressing their concerns even though whistle-blowing systems were in place. Some members of the laboratory did ultimately come forward with concerns about the research, but only after the University began an investigation (Kornbluth, 2011).

Despite review of the clinical trials by a scientific review committee and approval by the IRB, the trials were initiated using omics-based tests that were not "locked-down" or properly validated and turned out to be unreliable. Three years later, the Duke Institutional Review Board (IRB) initiated an investigation of the three clinical trials based on the concerns of the National Cancer Institute and external statisticians. The IRB formed an external review committee composed of two statisticians to conduct an independent evaluation of the data but did not inform the external reviewers of the scientific questions raised by the MD Anderson biostatisticians and the NCI biostatisticians (Kornbluth and Dzau, 2011). The reviewers concluded that the omics-based tests were viable and likely to succeed based on the data provided to them by the Duke investigators. The university resumed the three clinical trials following this report (Kornbluth and Dzau, 2011). Sally Kornbluth, vice dean for research, stated in discussions with the Institute of Medicine in August 2011 that she wished, in retrospect, that they had directed the external reviewers to give more in-depth consideration to the specific concerns of the outside parties (Kornbluth, 2011). Califf explained to the IOM committee that the university administration exercised caution in investigating the work of Nevins, partly out of deference to a well-regarded, tenured professor (Califf, 2011a). This illustrates that institutions, in conducting these reviews, face non-financial conflicts (protecting reputations of the institution or of investigators) in addition to financial conflicts (of individuals; and of institutions in patents and spin-off companies).

There was also confusion about what constitutes individual and institutional financial conflicts of interest (COIs). The investigators had an intellectual property interest and financial stake in the omics-based tests they were developing and evaluating. After reviewing multiple versions of the trial protocols, the committee concluded that the consent forms for the clinical trials did not always include disclosure that some of the investigators held patents and had a financial stake in the omics-based tests being studied (which is recommended practice; OHSR, 2006; Kim et al., 2004). The institution itself was invested in the two spin off companies related to the omics studies; it divested itself of those interests after the misconduct investigation was initiated.

In response to this situation, the university formed the Translational Medicine Quality Framework committee to make recommendations to university leadership on appropriate oversight policies for omics research being tested in clinical trials (TMQF Committee, 2011).

**Commercially Available Tests**

Various types of institutions were involved in the development of the omics-based tests discussed in Appendix A, including universities (e.g., Mammaprint) and industry. However, the committee did not explore the institutional roles and responsibilities in these tests due to the lack of publicly available information and limited resources.

NOTE: See Appendixes A and B on the case studies for more information.

*Culture*

The NIH mission is clearly defined as accelerating translational research. Its mission statement is articulated as "seek[ing] fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce the burdens of illness and disability" (NIH, 2011). This emphasis creates stresses on the research oversight system and requires complex collaborations between clinical and basic investigators. Thus, it is timely, for institutional leaders to undertake a careful reappraisal of the research culture in their institution so that a culture of scientific integrity and transparency is promoted.

One challenge for institutions is expanding beyond a compliance-based culture, where the focus is on following the letter of the law, to a culture that emphasizes the spirit of the law and highlights the ethical principles underlying research-related behaviors (Geller et al., 2010; Yarborough et al., 2009). Institutions interests in preserving academic freedom can also make it difficult to impose rules that promote scientific rigor. Yarborough and colleagues (2009) conducted a workshop to identify strategies used by industries outside of biomedical research to promote ethically based cultures. The workshop participants emphasized the importance of self-regulation above and beyond what is required by the law. Also, the participants stated that when problems occur (from errors to misconduct), the institutions involved need to conduct root-cause analyses to understand how the system allowed the problems to occur and take steps to correct the systemic problems to avoid similar lapses in the future. These suggestions are particularly relevant to omics research because it is a quickly developing field, and merely complying with existing rules may be inadequate for the responsible conduct of research. Moreover, omics-based test development sweeps across basic research, translational research, clinical research, and regulatory requirements for clinical applications.

Although there is no guidebook detailing exactly what steps and actions institutions should implement to encourage good behavior, a recent NRC (2002) report identified a list of practices that institutions can engage in to promote responsible conduct in research, including:

- Providing leadership in support of responsible conduct in research;
- Encouraging respect for everyone involved in research;
- Promoting productive interactions between trainees and mentors;
- Advocating adherence to the rules regarding all aspects of the conduct of research;
- Anticipating, revealing, and managing individual and institutional conflicts of interest;
- Arranging timely and thorough inquiries and investigations of allegations of scientific misconduct and applying appropriate administrative sanctions;
- Offering education pertaining to integrity in the conduct of research; and

- Monitoring and evaluating the institutional environment supporting integrity in the conduct of research and using this knowledge for continuous quality improvement (NRC, 2002).

*Oversight of Research*

Biomedical research falls under the purview of multiple federal regulations. Institutions are responsible for ensuring that research conducted at their facilities or by their investigators complies with these regulations. Three major federal regulations governing human subjects research are: (1) the Common Rule, which protects the safety, autonomy, privacy, and fair treatment of patient-participants in federally funded research conducted on humans, and the cultural groups from which they are recruited; (2) the *Health Insurance Portability and Accountability Act* Privacy Rule, which protects the privacy of personally identifiable health information created or received by healthcare professionals, health plans, or healthcare clearinghouses; and (3) the FDA Protection of Human Research Subjects regulations, which protect the rights, safety, and welfare of human subjects involved in research on products that the FDA regulates, including drugs and medical devices (which include omics-based tests). The two main regulations governing animal research are: (1) *The Animal Welfare Act*, which sets standards for the transportation, care, and use of animals in research; and (2) *The Health Research Extension Act*, which delegates authority to the Secretary of HHS for animals used in biomedical research. Within an institution, multiple oversight bodies are involved in supervising research (see Box 5-3). The exact organization of the oversight bodies and policies is specific to each institution. Ultimately, every institution that conducts biomedical research should ensure that patient-participant safety and privacy are protected and that conflict of interest (COI) and good data management and analysis practices are followed.

**Institutions are responsible for establishing, supporting, and overseeing the infrastructure and research processes for omics-based test development and evaluation as well as best practices for clinical trials and observational research, including those incorporating omics technologies, and should assure that the evaluation process outlined in this report is followed for omics-based test development and evaluation at their institution (Recommendation 4a). Given the complexity of omics research and omics-based tests, the multidisciplinary nature of omics and research, and the potential for conflicts of interest in developing and evaluating tests for clinical use, institutional leaders should pay heightened attention to providing appropriate oversight and promoting a culture of scientific integrity and transparency (Recommendation 4b).** These recommendations aim to emphasize and enhance institutional awareness of existing responsibilities to ensure the integrity of the scientific process; this includes optimally organizing oversight bodies. Omics research may impose some novel research designs and situations for institutional oversight bodies. **To facilitate compliance with the FDA's regulation of new devices (including omics-based tests), the committee recommends that institutional leaders designate specific Institutional Review Board (IRB) member(s) to be responsible for considering Investigational Device Exemption (IDE) and Investigational New Drug (IND) requirements as a component of ensuring the proper conduct of omics-based clinical research (Recommendation 4[b][i]).** The committee makes specific recommendations to the FDA on how to improve this process (discussed below). IRBs are also required by law to include members "with varying backgrounds to promote complete and adequate review of research activities commonly conducted by the institution," such as ex-

pertise in clinical trial design or omics research.[1] In addition, institutions may need to develop infrastructure to protect data provenance in omics research (see Box 5-4 on Clinical Trial Management Systems).

Institutions also are responsible for overseeing COI policies. A recent IOM report defined a COI as "a set of circumstances that creates a risk that professional judgment or actions regarding a primary interest will be unduly influenced by a secondary interest" (IOM, 2009b, p. 46). The potential for bias and COI to compromise the scientific rigor of a study can be particularly great for industry-sponsored studies. A substantial body of evidence suggests that biomedical studies funded by industry are open to systematic bias (Bekelman et al., 2003; Blumenthal et al., 1996; Rennie, 1997; Stelfox et al., 1998). The new financial COI policy governing all research funded by the NIH or other U.S. Public Health Service (PHS) agencies states that institutional officials, rather than the investigator, should determine if payments from drug companies or other outside sources constitute a conflict (Kaiser, 2011).[2] This policy requires institutions receiving PHS funding to develop specific COI policies. Commonly accepted practices include requiring the disclosure of financial relationships, the prohibitions of certain relationships, and the management of COIs that have been identified. There is also a push in biomedical research to broaden the definition of COI to include secondary interest beyond financial interest, such as personal, professional, political, institutional, religious, or other associations (Drazen et al., 2009, 2010). At the same time, under the *Bayh-Dole Act*[3] and other laws and policies, there is a competing desire for universities to transfer technology to the private sector, which can result in financial profits for the institution and investigator(s). Thus, there is high interest in establishing spin-off companies, such as the two companies[4] formed by the relevant investigators at Duke (see Appendix B). This situation adds to the complexity of managing COIs effectively.

The IOM report on COI recommended that some basic policies should be implemented by all institutions, such as a presumption that investigators with COI should not conduct human subjects research, with exceptions permitted but managed (IOM, 2009b). Steps also need to be taken to control the COI of participants on institutional oversight bodies. Individuals on these bodies should not be associated in any way with the research they are supervising. In addition, these individuals should disclose their COIs to their colleagues, other committee members, and their trainees. The NIH policy for Data and Safety Monitoring, for example, requires that the individuals charged with monitoring a trial at an institution receiving NIH funding are not associated with the trials and recommends that institutions evaluate and manage any existing conflicts (NIH, 1998).

Institutional COIs are equally important as individual COIs and, thus, must be managed as part of the oversight of research. According to the IOM, an institutional COI arises "when an institution's own financial interests or those of its senior officials pose risks of undue influence on decisions involving the institution's primary interests" (IOM, 2009b, p. 218). These COIs are often due to the licensing of intellectual property owned by an institution itself, the institution's partial ownership of companies arising from its research, or COIs as the result of endowed chairs and scholarships. Institutional COIs also can occur when members of an institution's leadership have personal financial interests that may affect their decision making on behalf of the institution.

---

[1] Protection of Human Subjects, 45 CFR 46,107 (2009).
[2] *Responsibility of Applicants for Promoting Objectivity in Research for which Public Health Service Funding Is Sought and Responsible Prospective Contractors*, 76 Fed. Reg. 53256 (2011).
[3] *Patent and Trademark Act Amendments of 1980*. Public Law 96-517 (December 12, 1980).
[4] Expression Analysis and CancerGuideDX.

In addition, institutions can be influenced by secondary interest beyond financial interests, such as factors that impact an institution's reputation. In research, such reputational factors can be quite prominent and difficult to manage, including deference to esteemed and well-funded investigators and the importance to both investigators and institutions of faculty publications in high impact journals. Few federal laws and regulations oversee institutional COIs. The IOM report on COI recommended that an institution's board of trustees or an equivalent governing body be given authority to make judgments about institutional COIs (IOM, 2009b).

A key lesson learned from the Duke case study is that COI, though subject to multiple layers of oversight in most institutions, can still contaminate research integrity (see Appendix B). Thus, **the committee recommends that institutional leaders designate an institutional official who is responsible for comprehensive and timely documentation, disclosure, and management of financial and non-financial conflicts of interest, both individual and institutional (Recommendation 4(b)(ii)).** This official should have the power to act independently of other parts of the institution. Institutions should pay particular attention to non-financial interests such as loyalty to the institution, as well as promotion policies that incentivize publication in high-impact journals and grants, but not research integrity. In addition, the confirmation and replication of other investigators' work is frequently poorly rewarded. Incentives need to be established to assure appropriate oversight by knowledgeable professionals. Institutions need to be particularly sensitive to institutional COI when investigating scientific controversies (see discussion below). When there are substantial COI, including non-financial COI, it may not be possible for the institution to fairly conduct investigations into controversies. All of these considerations may be compounded in multi-institutional research studies, including the need for inter-institutional communication and disclosures.

**BOX 5-3**
**Examples of Institutional Oversight Bodies**

- **Institutional Review Boards (IRBs):** Protect human safety, privacy, and autonomy; ensure informed consent
- **Privacy Boards:** Protect the privacy of individuals involved in research (if not done by an IRB)
- **Scientific Review Boards:** Evaluate science proposed for testing in a clinical trial
- **Data and Safety Monitoring Boards (also called Data Monitoring Committees):** Independently monitor clinical trials to ensure the continuing safety of human subjects and the validity and integrity of data
- **Conflict of Interest Committees:** Review individual's and institution's possible conflict of interest

**BOX 5-4**
**Clinical Trial Management Systems (CTMSs)**

Clinical trials that provide the definitive evaluation and utility assessment of the omics-based test should use a clinical trial management system (CTMS), in which data are entered, edited, and stored in a controlled environment where audit trails can protect the data from inadvertent or intentional corruption. Such CTMS systems are now required by National Cancer Institute–funded cancer centers and National Institutes of Health–funded CTSA institutes, which are located at many leading academic research institutions. Although use of these systems is now widely accepted in clinical research, their use in omics-based research is not standard practice. Rather, various data management strategies may be developed locally by basic omics research groups where quality control and data integrity and security are not optimal. However, there are commercial software systems that can provide such data management security requirements, and institutions can readily purchase a system that is used by a central data management office for any omics-related research that is intended to lead to a product for clinical use, and, thus, should go through a rigorous clinical evaluation such as a randomized clinical trial. In addition, the omics data and algorithms that were used to develop the omics-based test should be available and auditable to allow external review and validation.

SOURCES: Chahal, 2011; Choi et al., 2005; Philip et al., 2003.

*Whistleblowing*

When credible questions arise regarding the reliability of scientific work, institutions have the primary responsibility for investigating the merits of such questions. This responsibility has two aspects: (1) creating a safe system for reporting potential lapses, and (2) conducting an investigation into credible concerns about scientific integrity. The National Academies has stated that whistleblowing can be valuable in preserving the integrity of the research process and should be supported by the entire research community (NAS, 1992). It recommended that institutions establish a central office for handling allegations of potential scientific irregularities in trials and develop clear policies for reviewing these allegations. Most institutions have implemented this recommendation and established central offices that are charged with promoting the ethical conduct of research and protecting good-faith whistleblowers from retaliation. However, there is variability across institutions with respect to which office has this responsibility. At The Johns Hopkins University School of Medicine, for example, the Office of Policy Coordination is charged with these responsibilities, whereas at the University of Virginia, the Office of the Vice President for Research is responsible. At the College of William and Mary, the Office of Sponsored Research is responsible. Furthermore, the degree to which investigators are aware of such whistleblowing systems at their institution is unclear and their use is uneven. This is problematic because individual investigators and trainees are in the best position to identify potential research irregularities (Frankel, 1995; NAS, 1992). At the same time, there may be concerns about unfounded allegations or personal grudges triggering troubling inquiries, as well as confusion about the acceptability of certain practices. Many institutions have designated ombudspersons whom investigators can consult when they need guidance about research integrity (IOM, 2002).

ORI's regulations state that each extramural entity that applies for a biomedical or behavioral research grant or cooperative agreement should establish policies and procedures that provide for "undertaking diligent efforts to protect the positions and reputations of those persons who, in good faith, make allegations" in accordance with the Code of Federal Regulations (42 C.F.R. Part 50.103[d][13]). This helps ensure that all individuals, regardless of rank, feel comfortable raising questions about the integrity of a research study (Geller et al., 2010). However, institutions inherently have a strong incentive to discourage whistleblowing or to not fully investigate allegations of research irregularities; institutions risk forfeiting industry support and grants, harming their reputation, garnering negative publicity, and being subject to retaliatory litigation when there are breaches in research ethics (Rhodes and Strain, 2004). These negative incentives illustrate again the influence of non-financial COI. Many investigators perceive the reluctance of institutions to investigate potential misconduct and fear negative consequences. Thus, a major barrier to whistleblowing is the perceived threat of institutional recriminations against whistleblowers (NAS, 2009). In addition, whistleblowers might harbor guilt about coming forward too late if they were aware of inappropriate activity for some time but did not report it. They might fear being accused of covering something up or of being held legally liable.

A recent survey of 4,298 investigators at 605 institutions with NIH extramural research funding asked respondents about observed instances of likely misconduct over a 3-year period from 2002 to 2005 (Titus et al., 2008). Eight percent of the respondents indicated that they had observed or suspected investigators in their own department of committing research misconduct. Only 58 percent of the incidents, however, were reported to officials at their institutions (either by the respondent or someone else); 37 percent of the incidents were not reported by anyone; and in 5 percent of the cases, the respondent did not know if the incidents were reported. An earlier survey conducted by the International Society for Clinical Biostatistics also found that a majority of respondents were aware of instances of medical fraud, but many respondents did not know whether their organization had a formal system for reporting suspected fraud (note: the response rate in this survey was only 37 percent) (Ranstam et al., 2000). In a multimodal needs assessment conducted at The Johns Hopkins University to inform the development of research ethics education and services, a major theme that emerged was fear of punishment for reporting research breaches and the difficulties of the power differential between the levels of the organizational hierarchy (Geller et al., 2010). Fifteen percent of the respondents in this assessment indicated they would not feel comfortable reporting suspected breaches in research integrity out of fear of professional repercussions.

**Thus, the committee recommends that institutional leaders designate an institutional official who is responsible for establishing and managing a safe system for preventing, reporting, and adjudicating lapses in scientific integrity, to enhance patient safety (Recommendation 4[b][iii]).** The lines of reporting within the whistleblowing system should be independent from the institutional offices responsible for developing intellectual property in order to minimize the pressure of any financial COIs. Academic medical centers may be able to draw upon their growing experience with "blame-free reporting" of hazards to improve patient safety and quality care to implement corresponding systems-type oversight and inquiries for basic, translational, and clinical research (Pronovost et al., 2003).

*Investigating Scientific Controversies*

Little guidance is available on how much or how strong the allegations and their basis should be before triggering an institution's investigation of a scientific controversy. There also

are no existing criteria to inform an institution's decision to initiate an internal versus external review of a potential scientific controversy. As discussed in the IOM report on COI, "no decision maker in an institution is fully free of conflict in the case of institutional conflicts of interest." The Duke leadership recognized this problem and stated that in some instances an institution's COI may be too substantial to conduct an effective internal investigation. In such cases an external institution should conduct the investigation (see Appendix B) (Kornbluth, 2011). Ultimately, these decisions depend on the circumstances of each case and require careful attention and judgment by the institutional leadership.

**The committee recommends that institutional leaders designate an institutional official who is responsible for establishing clear procedures for response to inquiries and/or serious criticism about the science being conducted at the institution (Recommendation 4[b][iv]).** For example, this individual would be the responsible official for journals to contact with a serious concern about a manuscript, ensure that relevant information is provided to external investigators to help resolve issues of transparency of methods and data, and inform funders when an investigation of potential scientific misconduct is initiated. NIH's ORI only will become involved in an institution's investigation if the scientific controversy rises to the level of misconduct (ORI, 2011). All institutions receiving federal grants are required to have assurances on file with ORI stating that they have developed and will comply with an administrative process for responding to allegations of research misconduct. When problems do arise, ORI monitors the institution's investigation of the misconduct, but ultimately the institution is responsible for addressing and resolving any controversies (NIH, 2010).

These investigations require funding, and sometimes, outside expertise. Adequate resources and time should be made available for an investigation to be done thoroughly and fairly. The review should be conducted by individuals with the necessary expertise, and these experts should be completely independent in their review. Access to all relevant information from within and outside the institution, such as information from external investigators who have expressed concerns and from funders and journal editors, is essential to the success of the review. If an institution has a COI impacting the case, special protections should be in place to guarantee the lack of bias in the review. One example is to use only external experts who have free access to all data relevant to the case. In addition, some institutional COIs, such as substantial financial investment in the research or the potential for a high-impact breakthrough that can greatly enhance the institution's stature, should be more carefully managed and acknowledged during investigations and may indicate the need for the investigation to be conducted completely independent of the institution.

### *Biostatisticians, Bioinformatics Scientists, Pathologists, and Other Collaborators*

Omics research is multidisciplinary and requires effective teamwork. Institutions play a pivotal role in promoting teamwork, training faculty in effective team-based practices, and in rewarding collaborative accomplishments (Altshuler and Altshuler, 2004). Institutions also have the responsibility to ensure that the research team includes individuals with all of the required expertise. This section emphasizes the important role that biostatisticians, bioinformatics scientists, and pathologists play in omics research. However, many of the issues discussed below apply more broadly to the various collaborators who are involved in omics research and test development, including experts in omics technology and clinical trials.

The complementary disciplines of biostatistics and bioinformatics are both required in order to analyze and interpret the large multidimensional datasets used in omics research. Alt-

hough there is overlap in the principles and methods of these two disciplines, they are distinct. Biostatisticians are trained in experimental design and data analysis. Bioinformatics faculty focus on developing fast, efficient algorithms for data reduction, data mining, and literature search techniques, and formulating biologically informative annotations relating to DNA or RNA sequence, gene or protein expression, and the interaction of pathways, networks, phenotypes, and druggable targets. Biostatisticians and bioinformatics scientists publish in distinct sets of journals. In recent years, biostatisticians have tended to focus their careers either on classical clinical research, including clinical trials, or on the newer fields of genomics and statistical genetics. Most biostatisticians do not possess expertise and experience in both realms. Given the nature of omics research and omics-based clinical trials in particular, it is important that biostatisticians with expertise in both statistical genomics and clinical trials be involved, as well as individuals with bioinformatics expertise.

The shortage of these quantitative scientists is well known and the gap between supply and demand has been growing since the genomics era began (DeMets, 2009; DeMets et al., 1998). Reasons for this shortage are numerous, but include the fact that the supply of Ph.D.-trained experts in these fields has remained relatively constant for the past two or three decades while the demand has skyrocketed. Unfortunately, the NIH, which funds most of the doctoral training programs in this country, does not have a unified approach to training biostatisticians and bioinformatics scientists. Rather, training grants for these fields are scattered across the disease-oriented institutes, and the review of these training grants do not always include peers who are quantitative scientists. Further compounding the professional staffing crisis is the new set of challenges for the design, conduct, and analysis of research in the era of genomics, much like that experienced in the field of clinical trials four decades ago. There are special needs for biostatisticians and bioinformatics scientists to develop new experimental designs and methods for analysis because existing methods are not optimal or even adequate for current challenges (Apweiler et al., 2009; Mischak et al., 2007; Simon, 2008, 2010). Trained bioinformatics scientists are also needed to perform the complex analyses required by omics research, a collaborative task that may not promote career advancement.

Investigators developing new biomarkers for clinical use often do not include in their collaboration teams the pathologists and scientists with clinical laboratory expertise in proper methods for tissue diagnosis and selection for testing, test development, test validation, and ongoing test performance in compliance with clinical laboratory standards. In the worst of circumstances, investigators are not aware of the benefits of collaboration with pathology experts and the contributions such faculty can make to the translational process of omics-based test development, validation, and implementation in clinical use. Alternatively, pathologists are viewed as technicians who simply perform tests, rather than physicians with knowledge and experience who can facilitate the translational aspects of an omics-based discovery into a clinical test. The inclusion of pathologists or clinical laboratory scientists in the proper validation of a new omics-based test prior to use in a clinical trial to direct patient care enhances patient safety and the quality of the testing during the clinical trial. **Thus, the committee recommends that institutions that conduct biomedical omics research, including test development and clinical trials, should train, recognize, and support the faculty-level careers of individuals from the multiple collaborating disciplines, including biostatistics, bioinformatics, pathology, omics technologies, and clinical trialist (Recommendation 4c).**

The critical roles for these disciplines in omics research should compel institutions to develop analytical units, sections, or departments for biostatistics, bioinformatics, and pathology

faculty, staff, and trainees. Involving these faculty and staff only at selected steps in the omics-based test development is inadequate. Rather, they need to be viewed as equal partners on the research team. **The committee recommends that biostatisticians, bioinformatics scientists, and pathologists, as well as other collaborators in omics research, be treated as equal co-investigators and co-owners of responsibility (Recommendation 4[c][i]).** In the NRC report, *Catalyzing Inquiry at the Interface of Computing and Biology* (NRC, 2005), devaluation of the contributions of collaborators from different fields was discussed as an important cultural issue for research taking place at the so-called BioComp (biomedical–computational) interface. This concept can be extended to biostatisticians, bioinformatics scientists, pathologists, and other collaborators who participate in omics research. As an example, the National Cancer Institute- (NCI-) supported Cancer Centers and NIH-supported Clinical Translational Science Award (CTSA) units at many of the leading academic research institutions provide support for both biostatistics and bioinformatics cores (Berry, 2012; DeMets, 2009). In doing so, they have created an expectation and tradition of having these faculty members as collaborators, starting with the experimental design.

**The committee also recommends that institutions ensure that biostatisticians, bioinformatics scientists, pathologists, and individuals from the other multiple disciplines that collaborate in omics research are represented on all relevant review and oversight bodies within the institutions (Recommendation 4[c][ii]).** Omics-based test development with serious design or analysis flaws will ultimately fail the clinical validation process and will waste investigator time and resources. Minimizing false-positive leads in omics-based tests is in the best interests of the investigators, institutions, and, most importantly, patients. The same is true for the test validation steps. Trials with serious design flaws or with classifiers/tests that have not yet been defined should not be approved or implemented. If conducted, they may produce data that erroneously lead to an omics-based test being used to guide clinical decision making, with potential adverse consequences for patients. As omics-based grant applications and clinical protocols are being prepared, biostatistics, bioinformatics, and pathology faculty should be part of both the research team and the review team. This will ensure that only the most appropriate and most rigorously designed and analytically sound plans and testing processes are being proposed. Funders and journals that are involved in omics research also need to ensure that biostatistics, bioinformatics, pathology, and other faculty collaborating in omics research are involved in reviewing grant proposals and submitted manuscripts. Biostatistics, bioinformatics, pathology, and other faculty collaborating in omics research ideally should be part of a larger unit, such as a section or department, where they can get mentoring. Being part of a larger unit also allows them to get support from a more senior leader so they can contribute all of their expertise to the omics research effort without feeling pressured to deviate from best practices. Faculty, and especially staff, working in isolation may not know how to defend themselves when they are asked to conduct incomplete or flawed analyses that would create biased or misleading results and interpretations. Furthermore, such isolation does not foster academic development or promotions and should be avoided. **The committee recommends that institutions ensure that individuals from the multiple disciplines that collaborate on omics research and test development are intellectually independent, preferably reporting to an independent mentor and/or department chair as well as to the project leader (Recommendation 4[c][iii]).** The key concept here is to ensure that collaborators in omics research can act as independent scientists in applying their specific analytical expertise. However, they should be heavily integrated in their scientific

research team in order to be effective. This arrangement enhances independence and reduces risks for inappropriate pressure and COI.

## FUNDERS

Multiple types of organizations fund omics research, including government agencies, for-profit institutions, private foundations, public non-profit organizations, and international organizations. The NIH is by far the largest funder of research at academic and independent research institutions in the United States. The principal roles and responsibilities of funders of omics research are the same regardless of the type of entity. However, international organizations are outside the scope of this report.

Funders have influence over the conduct of research because they determine which projects are funded, and thus, which projects ultimately are conducted. They have a responsibility to sponsor scientifically rigorous research and to develop policies that promote the responsible conduct of research among their grantees. Funders also can use their relationship with investigators and institutions to encourage these parties to adopt and adhere to standards and best practices, such as sharing data and code. For example, the NIH and National Science Foundation now require data-sharing plans for large grants (NIH, 2010; NSF, 2001). The challenge is to balance the funder's interest in promoting innovative science and advancing a field of study with the need for oversight. Funders can find fulfilling an oversight role to be particularly difficult when financial support for a specific project comes from multiple funders, and it is unclear which aspect of a project any given funder is supporting.

This section of the chapter presents the roles and responsibilities of funders. It focuses primarily on the NIH, and specifically the NCI of the NIH for the Duke case, because more information is known about the practices of the NIH than for-profit institutions, private foundations, and public non-profit organizations. The committee makes several recommendations for funders of omics research that address the availability of data and code, the support for data repositories and test validation, and the role of funders in responding to scientific controversies. Box 5-5 highlights themes from the case studies for funders to use.

**BOX 5-5**
**Themes from the Case Studies for Funders**

**The Duke Case Study**

This case highlights significant barriers that funders face with regard to effective oversight and communication. For example, the National Cancer Institute (NCI) has a policy that requires investigators to make their data and code publicly available at the time of publication (NIH 2010) but that policy was not followed by Duke investigators. It is unclear whether the other funders of the clinical trials had similar policies in place, but it is clear that, where such policies exist, funders face a challenge in overseeing compliance with them. In 2009, the NCI contacted Duke regarding its concerns about the validity of the omics-based tests being used in the three clinical trials named in the statement of task. NCI initially relied on Duke's investigation of the allegations. However, NCI was unable to review the university's charge to that committee, or the draft report of the external reviewers to ensure that it was responsive to NCI's concerns. Duke had told NCI they would notify the sponsors of the trials about the actions they were taking when they initially suspended the 3 statement of task trials. NCI's review of its trials databases and ClinicallTrials.gov did not reveal NCI sponsorship for any of the three trials (McShane, 2010a,b), but when the NCI staff determined in April 2010 that it was providing partial funding through an R01 grant to Potti for the tests for sensitivity to cisplatin and pemetrexed, it requested the resulting data and computer code necessary to reproduce results in the paper cited in the grant as providing validations for the cisplatin and pemetrexed predictors (Hsu et al., 2007). NCI staff evaluated the cisplatin test and were unable to reproduce the results. (McShane, 2010). The NCI then asked Duke to produce the original raw data that would reproduce the findings in the papers. On October 22, 2010, Duke notified the NCI that multiple validation datasets associated with the cisplatin predictor were corrupted (McShane, 2010). Thus, the trials were closed and retraction of the paper was initiated.

**Commercially Available Tests**

Various types of funders supported the omics-based tests described in Appendix A, including government funders, private non-profit organizations, and industry. However, the committee did not explore the roles and responsibilities of funders in the development process of these tests due to the lack of publicly available information and limited resources.

NOTE: See Appendixes A and B on the case studies for more information.

## Role and Responsibility in Research

Funders, including those who support omics research, are responsible for screening prospective research projects and for monitoring funded projects. The peer review processes that funders use to select the projects they want to support generally rely on committees made up of scientific peers. The NIH, for example, uses a dual-level peer review process. The first level review is conducted by a Scientific Review Group (called a study section), which is composed of non-federal scientists and lay members, and focuses on the scientific merit of the proposal. The reviewers are directed to consider the following criteria in evaluating a proposal: significance, investigator expertise, innovation, approach, and research environment. The second level of review is performed by NIH staff within the specific Institute of the NIH that is considering the

proposal, and assesses whether the proposal is consistent with the Institute's programmatic and funding priorities. The Institute directors make the final funding decision based on the advice of reviewers (NIH, 2010). A similar process is used by many non-federal funding agencies. However, non-scientific reviewers also may be included. At the American Cancer Society, for example, proposal review is conducted by a peer review committee made up of 12 to 25 cancer investigators and non-scientists. Each application for funding is assigned to at least two committee members who consider the application's scientific merit, originality, feasibility; the qualifications and expertise of the investigative team; the facilities and resources available for the project; and the potential of the research to improve cancer treatment (ACS, 2011).

In omics research, whether it is funded by a federal or non-federal entity, it is important that the peer review process involve biostatisticians and bioinformatics scientists who can assess the quality of complex biomarker trial design and the proposed data collection and analysis plans (see discussion on biostatistics/ bioinformatics in the institutions section above). In addition, some very large grants, such as program projects or Center grants, have subprotocols embedded within them or developed during the life of the grant. These subprotocols also should be reviewed by biostatisticians and/or bioinformatics scientists, as appropriate, if they involve omics research.

Funders also should have a method to track a study once it is funded. This is important because it allows funders to oversee the research process and ensure that investigators and institutions are applying best research practices. There are several mechanisms for tracking research studies that are mandated by law. The *Federal Funding Accountability and Transparency Act* (FFATA) created the website www.USASpending.gov, which provides information on federal grants and contracts over $25,000. Similarly, www.clinicaltrials.gov is a repository of most clinical trials involving a drug, biological product, or device (see discussion on trial registration below). However, FFATA is limited to federal contracts and grants, and clinicaltrials.gov only includes clinical trials (not other study designs). Thus, privately funded omics studies that are not clinical trials are not included in either of these repositories.

Most funders use more active methods of monitoring their funded studies to address these gaps and to provide an additional level of oversight. For example, the NIH conducts active monitoring by reviewing progress reports and correspondence from grantees, requiring audits, and conducting site visits (NIH, 2010). Funders with more limited budgets are likely to rely heavily on progress reports as their mechanisms of oversight (ACS, 2011; PCF, 2011; PhRMA Foundation, 2011). Some also may require meetings with the grantees to monitor the research (PCF, 2011).

### Data and Code Availability

Many funders have policies requiring grantees to make their data and code publicly available prior to publication (Sherpa, 2011). Requiring investigators to share their data and code can maximize the societal benefit resulting from a funder's support and contributions to a project. The policies of the Wellcome Trust, for example, specifically state that sharing data and code leads to: (1) faster progress in translating research results into practices and products that improve human health, (2) better value for the money, and (3) higher-quality science (Wellcome Trust, 2011). The requirement to share data and code is also consistent with the 2003 NRC report on data sharing, which recommended that sponsors of research "clearly and prominently state their policies for the distribution of publication-related materials and data" (NRC, 2003, p. 11). It

also recommended that sponsors provide the recipients of research grants with the financial resources needed to support the dissemination of data and code.

Funders' existing policies on data and code sharing vary in stringency and level of detail. NIH policy endorses the "timely release" of final research data (i.e., at the time of publication) (NIH, 2010). Unfortunately, the "final research data" phrase is quite ambiguous because a whole series of publications over many years may be based on the ongoing analyses of the data until the research data are called "final" by the investigators. The NIH policy also requires investigators applying for projects with direct costs of $500,000 or more in a given year to address data sharing in their applications. NSF's data-sharing policy specifically addresses the availability of algorithms and code and requires investigators to share any corresponding software and materials that are necessary to interpret the data (NSF, 2001). When investigators believe the data arising from their studies are not amenable to sharing, the Medical Research Council's policy states that investigators should provide an explicit explanation in their proposal for not making the data available (Lowrance, 2006).

However, there are still many funders who do not require grantees to share their data and code. A group of 33 research universities with an interest in open access have created a website that tracks research funders' policies on data sharing (Sherpa, 2011). Of the 80 funders' policies assessed on this website, only 18 have data-archiving policies. In addition, many funders with data-sharing policies do not enforce them (Piwowar, 2011). The NRC report on data sharing recognized this problem and recommended that funding organizations have published procedures for resolving problems of non-compliance with data sharing (NRC, 2003).

**The committee recommends that funders require investigators to make all data, metadata, prespecified analysis plans, code, and fully specified computational models publicly available and readily interpretable either at the time of publication or, if not published, at the end of funding, and funders should financially support this requirement (Recommendation 5[a][i]).** If the investigators make this information available at the end of funding, it should be held in escrow for 2 years to allow the investigators an opportunity to publish their research. Issues of proprietary information can be dealt with by depositing the materials with a responsible third party that can ensure confidentiality and protection of the material. **Funders also should provide continuing support for independent repositories to guarantee ongoing access to relevant omics and clinical data (Recommendation 5[a][ii]).** Although the methodology and funding for making data publicly available from sponsored research is still under discussion, such efforts should be made so the field of omics research can move forward. Transparency always is healthy in research, and the more individuals who can examine the available data, the more robust the conclusions will be. NIH Director Francis Collins has declared as an NIH priority that the genomic data generated be accessed and harvested (Collins, 2010). Omics-based tests, and the data on which they are based, clearly fall in this realm.

## Funding of Test Validation

A crucial step in developing an omics-based test to guide patient management in a clinical trial setting is appropriate validation in a CLIA-certified laboratory (described in Chapter 3). A candidate omics-based test may be applied to patient samples from a completed trial or even from an ongoing trial as part of the validation process as long as the testing does not interfere with the conduct of the clinical trial or impose undue hazards to patients. Before an omics-based test is considered ready to direct patient management in a clinical trial, the investigators from the discovery phase should identify a CLIA-certified laboratory, either a commercial or an academic

medical center clinical laboratory, to confirm that the candidate omics-based test is stable, reproducible, and validated appropriately for the intended study design for assessment of the clinical utility of the test (see Chapter 4). Investigators are responsible for arranging for the independent validation and, very importantly, sharing the evidence and methods necessary for a CLIA-certified laboratory to validate the candidate omics-based test in preparation for use in a clinical trial to direct patient management. However, the cost for validation of the candidate omics-based test in a CLIA-certified laboratory must be funded. In addition to the test validation described in Chapter 3, the committee also discussed that confirmation of the discovery phase (Chapter 2) findings which are the basis of the candidate omics-based test may be worthy of independent replication by another research laboratory. Thus, **the committee recommends that funders should support test validation in a CLIA-certified laboratory (as described in Chapter 3) and consider the usefulness of an independent confirmation of a candidate omics-based test prior to evaluation for clinical use (Recommendation 5[a][iii]).**

If an independent confirmation is funded, the validation study should be conducted using fully independent specimens or datasets, to provide a fully independent test of the omics-based discovery. As for the original research, all data, metadata, prespecified analysis plans, code, and fully-specified computational models of the independent study should be made publicly available either at the time of publication or at the end of funding, and funders should financially support this requirement (recommendation 5[a] above). Confirmation of a candidate omics-based test either by a CLIA-certified laboratory in preparation for use in a clinical trial or by an independent research laboratory is particularly important for complex omics-based tests because of the great potential for investigators' bias to influence the results, the complexity and quantity of the data, and the high likelihood of overfitting. The adage "trust but verify" is appropriate for this setting.

### Responding to Scientific Controversies

**The committee recommends that funders should designate an official to alert the institutional leadership when serious allegations or questions have been raised that may warrant an institutional investigation; if the funder has initiated that question, then the funder and institution should communicate during the investigation (Recommendation 5(a)(iv)).** As stated above, the committee recommends that all institutions that conduct omics research identify an administrator or office that outside parties, such as funders, can approach with serious concerns about the validity of work conducted by investigators within the institution, including problems that do not rise to the level of misconduct. The committee also recommends that institutions should inform the funder(s) when investigations are initiated on a study they have funded based on other parties' concerns regarding the integrity of that research. When the funding agency requests the review, the agency may request that the institution conduct either an internal or external review. The research institution and funding institution should communicate effectively to ensure that the funding institution's specific concerns are fairly heard and considered. The funders should be prepared to evaluate the research institution's review to decide if they believe it is thorough and convincing. In some cases, the funding institution may ask for their own appointed external reviewers to at least review the institution's report. Funding institutions may need to set aside a small fund to support investigations of serious allegations.

In the case of the Duke clinical trials evaluating omics-based tests, funding came from multiple sources including the NCI and the Department of Defense (DOD). Yet the committee could find no evidence of communication between NCI and DOD. To address this problem, **the**

**committee recommends that funding agencies should establish lines of communication with other funders to be used when serious problems appear to involve interdependent research sponsored by another funder along the omics-based test development process (Recommendation 5[a][v]).** Establishing such communication channels should help to alleviate confusion when multiple funders support various stages of omics-based test development research. The committee recognizes that it will be easier to establish lines of communication between federal funders of omics research than it will be for the many private funders of omics research. However, all funders of omics-based research have a responsibility to communicate with each other. The Interagency Oncology Task Force (IOTF) is an example of federal agencies communicating with each other and could serve as a model for communication among funders of omics research. Through the IOTF, the NCI and FDA are jointly sponsoring fellowship programs to train scientists in both preclinical and clinical research and the FDA's policies and regulations that govern research (IOTF, 2011). In addition, **the committee recommends that federal funders of omics-based translational research have the authority to exercise the option of investigating any research being conducted by a funding recipient after requesting an investigation by the institution (recommendation 5[b]).** The investigation by NCI is what lead to the discovery of the underlying problems with the data in the Duke University case study (see Appendix B).

## FDA

Two federal agencies have regulatory authority relevant to omics-based tests: the FDA and the Centers for Medicare & Medicaid Services (CMS). The FDA oversees the marketing of devices, including in vitro diagnostics (which encompass most omics-based tests).[5] However, the FDA has exercised enforcement discretion with regard to laboratory-developed tests (LDTs), meaning it does not oversee the development of tests that fall into this category. Laboratories that provide LDT services are regulated by CMS under the *Clinical Laboratory Improvement Amendments*[6] (CLIA) to ensure the quality of the laboratory testing services.

It is challenging for all parties involved in the development of omics-based tests (e.g., investigators, institutions, and IRB committees) to understand and correctly navigate the FDA's current oversight system. This is primarily due to the rapidly changing technological landscape and the longstanding and unclear practice of the FDA in its use of enforcement discretion as described in Chapter 4. In recent years, FDA has taken some initial steps to clarify regulatory policy for these tests, but more could be done to guide investigators. For example, FDA developed draft guidance on In Vitro Diagnostic Multivariate Assays (FDA, 2007), but that guidance was never finalized, and FDA is now moving away from that terminology. Box 5-6 highlights the great variability in mechanisms that omics-based test developers have used to bring a test to the market. However, test developers' uncertainty about the FDA's enforcement discretion does not excuse their mistakes and failures to submit a test that directs therapy and is of significant risk to the health, safety, or welfare of a subject to the FDA for discussion and consideration for an IDE.

**In order to enable investigators and institutions to have a clear understanding of their regulatory responsibilities, the committee recommends that the FDA develop and finalize a risk-based guidance or regulation on bringing omics-based tests to the FDA for review and on the oversight of laboratory developed tests (Recommendation 6a).** Specific

[5] *The Medical Device Amendments of 1976*. Public Law 94-295 (May 28, 1976).
[6] *The Clinical Laboratory Improvement Amendments of 1988*. Public Law 100-578 (October 31, 1988).

areas that need clarification include the circumstances when: (1) an omics-based test qualifies as an LDT, (2) an omics-based test and algorithm qualifies as a device, (3) devices are exempt from submission and review by the FDA, and (4) devices are considered to pose significant risk (e.g., does this determination take into consideration the state of health of the intended patients?). Clarification of enforcement discretion in the LDT arena, particularly for highly complex omics-based tests, is of paramount importance.

**BOX 5-6**
**Themes from the Case Studies for the FDA**

**The Duke Case Study**

In 2009, the FDA sent a letter to the investigators stating that the omics-based tests being studied in the three clinical trials named in the statement of task needed to go through the Investigational Device Exemption (IDE) process (Chan, 2009). In response, the investigators made some changes to the protocol of the studies and contacted the FDA for further clarification about whether an IDE was still required (FDA, 2011b; Potti, 2009). When the FDA failed to respond to these letters, the Duke Institutional Review Board (IRB) determined that an IDE was not needed (FDA, 2011b). However, in retrospect, the Duke IRB recognized that an IDE should have been obtained for the omics-based tests because the tests were used to direct patient management in the clinical trials (FDA 2011b).

**Commercially Available Tests**

A review of the six commercially available tests discussed in Appendix A demonstrates that companies have pursued both laboratory-developed tests (LDTs) and FDA pathways for translation of an omics-based test. The availability of multiple pathways indicates a lack of clarity and consistency on the regulatory requirements for omics-based tests. Five of the commercially available tests that the committee examined are performed exclusively by each company's proprietary laboratory that has certification under *The Clinical Laboratory Improvement Amendments of 1988* (CLIA) as an LDT.[a] Two companies did not seek FDA clearance and market their tests as LDTs: Genomic Health (Onco*type* DX) and CardioDx (Corus CAD). Four tests received FDA 510(k) clearance of their tests: Agendia (MammaPrint), Pathwork Diagnostics (Tissue of Origin), Vermillion (OVA1), and XDx (AlloMap).

In several of the case studies, the company and the FDA held a pre-IDE meeting to determine whether an IDE would be required for the test under development and validation. The FDA determined that an IDE was not needed for the AlloMap and Tissue of Origin tests because the test was not directing patient therapy in the studies proposed to assess the test. Physicians can now use these tests for that purpose, however.[b] Agendia reported that they received an IDE for MammaPrint that helped clarify the process and requirements for the de novo 510(k),[c] and Vermillion reported that they received an IDE for OVA1.[d] Two ongoing prospective studies direct patient management on the basis of Onco*type* DX Recurrence Score. For both trials, information required for approval of investigational use of Onco*type* DX in the trial was submitted as part of an investigational new drug application to FDA.[e] Regardless of which pathway is taken to market, consultation with the FDA can be beneficial. For example, the developers of the OVA1 test sought FDA input, and this early dialogue with the FDA prompted Vermillion to include two different cut-off values for the test, depending on a patient's menopausal status (Fung, 2010).

[a] OVA1 is performed exclusively by Quest Diagnostics, which is subject to CLIA certification (Quest Diagnostics, 2011). Currently Pathwork Diagnostics offers Tissue of Origin exclusively through its CLIA-certified laboratory, but is developing an in vitro diagnostics test kit for other laboratories (Pathwork Diagnostics, 2010).
[b] Personal communication, Mitch Nelles, XDx, October 12, 2011; personal communication, Ed Stevens, Pathwork Diagnostics, October 18, 2011.
[c] Personal communication, Laura van' t Veer, November 28, 2011.
[d] Personal communication, Scott Henderson, Vermillion, November 1, 2011.
[e] Personal communication, Lisa McShane, National Cancer Institute, February 9, 2012.

NOTE: See Appendixes A and B on the case studies for more information.

## FDA Operations

Two specific areas where FDA operations could be more transparent include (1) the pre-IDE process and (2) the quality reporting system. The FDA's pre-IDE process often is very helpful to test developers by providing clarity about the regulatory requirements for marketing an omics-based test. For example, the evidence supporting an omics-based test resulting from a prospective–retrospective trial design is very strong and should eliminate the requirement for an IDE if not used to direct choice of therapy. A pre-IDE meeting resulting in an agreement that this design is sufficient for market (or an agreement that this design would be part of a package of data) would be extremely beneficial to test developers. The FDA could improve its transparency by continuing to use the pre-IDE process, making it as widely available as possible within reasonable resource constraints, and publicly advertising its willingness to hold pre-IDE meetings. The committee recognizes that this type of pre-IDE agreement may be challenging in omics research because the science is changing rapidly.

The FDA also could improve its transparency by clarifying when its quality system requirements and manufacturers' quality management systems (QMSs) are required. These systems provide a high level of assurance about a product's production integrity and safety. However, the requirements are quite demanding, and many academic laboratories do not meet the requirements. In addition, many aspects of test development (e.g., analytical validation) are needed regardless of whether manufacturers go through the FDA process (and use a QMS) or whether the manufacturers develop a test in a CLIA-certified laboratory as an LDT.

## Communication of IDE Requirements

**The committee recommends that the FDA communicate the IDE requirements for use of omics-based tests in clinical trials to the Office of Human Research Protections (OHRP), IRBs, and other relevant institutional leadership (Recommendation 6b).** The committee determined that, in general, IRBs lack knowledge of the IDE requirements compared to their understanding of the IND requirements, and that clarification and education by the FDA about IDE requirements are necessary. This communication could be conducted online and via technologies such as webcasting in order to reduce the FDA's cost and time requirements. However, this educational outreach effort should be adequately resourced and as up to date as possible.

### Emerging Technologies

Although omics technologies are being developed in a rapidly changing environment, the FDA should try to develop guidance about emerging technologies in advance of the technologies coming to the market. The committee recognizes that this will be challenging. When it is impossible to have FDA guidance keep pace with technological advances, the committee encourages the FDA to organize forums with members of the scientific community and have an open and publicly accessible dialogue, as FDA has done on other matters. This will provide test developers with some insight into the FDA's thinking and potential next steps.

The challenges faced by the FDA with emerging technologies are particularly salient with respect to companion diagnostic tests. The committee applauds the FDA for its recently issued guidance on companion diagnostics (FDA, 2011a). However, where possible, further clarification of the relationship between IND and IDE requirements in the presence of a combination product or a companion diagnostic test would be of assistance to the scientific community.

## JOURNALS

Data compiled for *The Wall Street Journal* by Thomson Reuters suggest that the number and percentage of papers that journals are retracting has increased significantly in recent years, from 22 retractions in 2001 to 339 in 2010 (Naik, 2011a). A bigger percentage of these retractions have occurred in high-impact journal than in low-impact journals (Cokol et al., 2007). It is unclear whether this increase in retractions is due to an increase in mistakes and inappropriate methodologies by investigators, the increasing number of articles published in the increasing number of journals, or to increased vigilance by journals and the scientific community. Regardless, journal editors should play a key role in overseeing the quality of published research, including omics research.

Journal editors have a responsibility to use due diligence to ensure that the information reported in an omics study is consistent with what the investigators actually performed and that the conclusions are supported by the evidence. Journal articles of omics studies should accurately document the steps in the omics-based test development process in enough detail that the methods and results could be reproduced. The challenge for journal editors is ensuring that this standard is met and that omics studies are conducted in a transparent and scientifically rigorous manner. In addition, journals play a significant role in overseeing and minimizing the effects of bias and COI in published research. The *Manual of Style: A Guide for Authors and Editors* (MoS), for example, outlines several requirements that journals could implement to prevent bias and COI from undermining the credibility of reports containing original data (Fontanarosa et al., 2011; MoS, 2007).

If patients and clinicians are to rely on omics studies and tests to guide treatment decisions, journals need to ensure that the omics studies that are published adhere to best practices. The editorial policies of journals can institute quality control measures for assessing the merit of articles submitted for publication. The instructions for authors also can direct authors to abide by certain standards that advance science as a condition for publication (CSE, 2009). Specific policies may include requiring registration of clinical trials involving omics-based tests in clinicaltrials.gov; ensuring data and code availability; protecting the scientific integrity of published research; and developing a process to respond to significant scientific concerns. These requirements for publication are not unique to omics research. However, the importance of journals instituting policies that promote quality and transparency is magnified in omics research,

where the methodologies are highly complex and rapidly advancing. Box 5-7 highlights themes from the case studies for journals. The challenges of the reproducibility of science are increasingly discussed in high-visibility professional literature and the lay press (Ioannidis and Khoury, 2011; Naik, 2011b; Peng et al., 2006).

**BOX 5-7**
**Themes from the Case Studies for Journals**

**The Duke Case Study**

Keith Baggerly and Kevin Coombes, two MD Anderson biostatisticians who tried to reproduce the research results of Potti and Nevins, submitted letters to the journal editors of *Nature Medicine*, the *Journal of Clinical Oncology* (JCO), and *Lancet Oncology* with concerns about the omics-based tests being studied in the three clinical trials named in the statement of task. In general, correspondence with journals and "letters to the editor" did not provide resolution of questions about reproducibility because information contained in Potti and Nevins' responses did not enable investigators to reproduce the results and journals declined to pursue the issues further following additional inquiries by Baggerly and Coombes. *Nature Medicine* published their letter along with the authors' reply (Coombes et al., 2007; Potti and Nevins, 2007); JCO published one of their letter and the authors' reply (Baggerly et al., 2008; Dressman et al., 2008), but declined to publish their letter regarding the Hsu et al. (2007) article; *Lancet Oncology* rejected their letter (Baggerly, 2011). The Duke investigators maintained in their replies that, with only a few exceptions, the errors were clerical errors that had no impact on the actual tests developed or the reported test performance results (Dressman et al., 2008; Potti and Nevins, 2007). Meanwhile, their papers were used and cited by hundreds of other investigators.*

After deciding that the originating journals would not help to address and resolve remaining questions, Baggerly and Coombes published their alternative analysis and detailed critique of each of the papers in a specialty statistics journal (Baggerly and Coombes, 2009). Based on simultaneous inquiries from the National Cancer Institute (NCI), Duke decided to undertake independent analyses of the omics-based tests. The original papers were eventually retracted after NCI identified problems with the data in the Hsu et al., 2007 JCO paper, and then directed Duke to find the original raw data underlying that paper and to check for potential data corruption in that paper and others. The original papers were retracted after NCI analyses uncovered the data corruption (Bonnefoi et al., 2011; Hsu et al., 2010; Potti et al., 2011). Duke University also took steps to retract multiple additional papers where Potti and Nevins were co-authors (Califf, 2011b).

**Commercially Available Tests**

Many of the case studies described in Appendix A document important steps in the tests' development processes in the peer review literature. However, the committee did not explore the roles and responsibilities of journals in these case studies due to the lack of publicly available information and limited resources.

NOTE: See Appendixes A and B on the case studies for more information.
* The Potti et al. (2006) article was cited 306 times, the Hsu et al. (2007) article was cited 60 times, the Dressman et al. (2007a) article was cited 111 times, and the Bonnefoi et al. (2007) article was cited 95 times in *Scopus* (all as of October 28, 2011).

## Trial Registration

*The FDA Modernization Act of 1997* created clinicaltrials.gov to increase the transparency of clinical trials. It requires registration of trials of drug effectiveness for "serious and life

threatening diseases and conditions."[7] *The FDA Amendments Act*, Section 801, broadened the scope of clinicaltrials.gov to include a results database. All clinical investigations involving a drug, biological product, or device (other than Phase I trials), regardless of sponsor, are required to register results in this database.[8] Many journal editors have accommodated this requirement by stipulating that posting summary results will not interfere with publication if the results are presented as an abstract or table (Laine et al., 2007). To date, 108,000 trials have been registered in clinialtrials.gov and results for 3,600 trials have been reported (Marshall, 2011). There is also a recent push to create registries for new types of studies, such as tumor biomarker studies (Andre et al., 2011).

The argument in support of trial registries is strong. Individuals agree to participate in clinical trials based on the understanding that trials will improve medical knowledge and potentially lead to improved health for others. This only can happen if the public is knowledgeable about ongoing trials and the results are disseminated (Zarin and Tse, 2008). Currently, the reporting of biomedical research findings is often incomplete and biased (IOM, 2011). Investigators are most likely to publish positive findings, often report only a subset of the relevant data and outcomes, and may fail to report relevant adverse events (Chan and Altman, 2005; Chan et al., 2004a,b; Curfman et al., 2006; Dickersin and Chalmers, 2010; Dwan et al., 2008; Song et al., 2009; Turner et al., 2008; Vedula et al., 2009). Trial registries have the potential to address reporting bias by creating a public record of ongoing and completed trials (DeAngelis et al., 2005).

The impact of clinicaltrials.gov in addressing reporting bias was initially limited because the information submitted to the database was often inaccurate and incomplete, many investigators failed to comply with the registration mandate, and the government instituted few quality control mechanisms (Marshall, 2011; Zarin and Tse, 2008). The International Committee of Medical Journal Editors (ICMJE) increased registration by requiring all clinical trials to register at clinicaltrials.gov or an appropriate trial registry at the onset of patient enrollment to be considered for publication (DeAngelis et al., 2004, 2005; Laine et al., 2007). The ICMJE defined a clinical trial broadly to include "any research project that prospectively assigns human subjects to intervention and comparison groups to study the cause-and-effect relationship between a medical intervention and health outcome" (DeAngelis et al., 2004, p. 2436). This policy led to a 73 percent increase in trial registrations at clinicaltrials.gov for all intervention types (Zarin, 2005). More recently, every protocol registered at clinicaltrials.gov is required to undergo an automated review to identify missing information and a quality review to assess whether the experiment is presented accurately (Zarin et al., 2011). These practices are likely to improve the quality of the entries.

Despite the recent increase in trial registration, not all trials are registered at clinicaltrials.gov, and many journals still publish studies that have not been posted in the database. Those entries that are posted often lack essential information about the trial. Also, some trials fail to register prior to patient enrollment, as required by law and the ICMJE policy (Meldrum and DeCherney, 2011; Zarin et al., 2011). Although the FDA is trying to encourage sponsors to present more information on the website, "the usefulness of clinicaltrials.gov ultimately depends on whether responsible investigators and sponsors make diligent efforts to submit complete, timely, accurate, and informative data about their studies" (Zarin et al., 2011, p. 860). Journal policies are one mechanism to encourage comprehensive trial registration, including trials of omics studies. **Thus, the committee recommends that journal editors require**

[7] *Food and Drug Administration Modernization Act of 1997*, Public Law No. 105-115 § 113 (1997).
[8] *Food and Drug Administration Amendments Act of 2007*, Public Law No. 110-85 § 801 (2007).

**authors submitting manuscripts describing clinical evaluations of omics-based tests to register all clinical trials at clinicaltrials.gov or another clinical trial registry acceptable to the journal (Recommendation 7[a][i]).** The peer review process should confirm that authors have registered their trials and that any data posted in the registry is consistent with the data submitted for publication (Meldrum and DeCherney, 2011).

## Data and Code Availability

Baggerly and Coombes, two statisticians from MD Anderson who wanted to reproduce the omics-based tests being used in clinical trials at Duke University, reported spending more than 1,500 person-hours trying without success to replicate the statistical analyses. If the data used to develop the omics-based tests had been transparent and publicly available, checking the validity of the results would have been much faster and easier. To facilitate the reproducibility of omics research, Baggerly and Coombes recommended that journals require authors to make the following five items available prior to the publication of an omics study: (1) raw data, (2) the code used to derive the results from the raw data, (3) evidence of the provenance of the raw data so that data labels can be checked, (4) written descriptions of any nonscriptable analysis steps, and (5) the prespecified analysis plans (Baggerly and Coombes, 2011). The ability of independent investigators to access data and code for omics-based tests is particularly important because of the complexity of the data and analyses. As Baggerly's and Coombes' experience suggests, without access to the data and code, it is very difficult to judge the scientific integrity of the data and conclusions drawn. A recent edition of the journal *Science* was dedicated to data reproducibility (Jasny et al., 2011). Several of the articles also emphasized the importance of journals demanding that authors make their data and code available to improve the reproducibility of published research (Ioannidis and Khoury, 2011; Peng, 2011). Other investigators are currently working to identify methods of making computational research data readily available to the public (Stodden, 2010).

Journals have widely divergent policies on data sharing. Piwowar and Chapman (2008) investigated data sharing policies at 70 journals that published more than 15 articles on gene expression in 2006. Eighteen (26 percent) of the journals did not mention data sharing in the instructions to the authors, 11 (16 perfect) included requirements for sharing non-microarray data but no requirement for data in general, and 42 (60 percent) included data-sharing policies applicable to microarrays. Oncology journals often lacked any microarray data-sharing policies. In another study, Piwowar examined the percentage of gene expression microarray journal articles reporting an associated dataset published in a data repository from 2000 to 2009 (Piwowar, 2011). Of the 11,603 articles identified over the entire time period, only 2,901 articles (25 percent) indicated that the data were deposited in a data repository. However, the percentage of microarray journal articles reporting that data were submitted to a data repository increased each subsequent year, with less than 5 percent of articles submitting data in 2001, but 30-35 percent submitting data in 2007-2009.

The strength of data-sharing policies among journals that include instructions to authors on data and code availability also vary greatly. For example, *Science*'s and *Nature*'s instructions to the authors state that all data, materials, and associated codes and protocols should be available to the reader as a condition of publication. After publication, the authors should fulfill all reasonable requests for data and materials necessary for independent investigators to replicate the findings (*Nature*, 2011; *Science*, 2011). By contrast, the *Annals of Internal Medicine*'s policy on reproducible research only requires authors to publish a statement of their willingness to share

data and code and to specify any conditions to sharing (*Annals of Internal Medicine*, 2010). There is no actual requirement to share the data and code. This difference in policy is likely due to the different nature of the articles that these journals publish (basic science vs. clinical research). The challenges associated with data and information sharing vary across different fields of research. The journal *Biostatistics* offers authors the opportunity to request a "reproducibility review," in which the journal runs the data and code and confirms that the results can be reproduced. Articles are rated as R (reproducible), D (data provided), C (code provided), or none of the above (Peng, 2009). See Box 5-8 for lessons on data and code sharing from the banking industry.

The preferred method of many journals for sharing datasets is for authors to deposit the data in an approved database, such as clinicaltrials.gov, dbGaP, or GEO, and include instructions on accessing the datasets in the published paper (*Nature*, 2011; *Science*, 2011). Some journals also have websites that can host supplementary materials, including data and code (MoS, 2007). When necessary, journal editors may use their influence to encourage authors to more fully share their data and code and respond to queries from individuals who have a legitimate interest in understanding the methods of a study (MoS, 2007). The challenge for journal editors with limited resources and limited access to the necessary review expertise is overseeing authors' compliance with this policy and ensuring that the data and code deposited in repositories are accurate and complete. **The committee recommends that journal editors require authors submitting manuscripts describing clinical evaluations of omics-based tests to make their data, metadata, prespecified analysis plans, code, and fully specified computational models publicly available in an independently managed database (e.g., dbGaP) in standard format (Recommendation 7[a][ii]).** Journals should not accept papers on complex biomarkers for publication, if the corresponding data and software are not independently certified as available. This requirement is particularly important in omics research where the datasets are enormous and the code is complex. It may be useful for journals to contact not only individual investigators when this policy is not effectively implemented, but also deans of investigators' institutions. A system that journals could use to verify that the code is reproducible from the starting data has also been developed recently (Segal et al., 2012).

**BOX 5-8**
**Lessons from the Banking Industry on Data and Code Sharing**

The *Journal of Money, Credit, and Banking* instituted a policy in 1982-1984 requiring authors to submit the data and code supporting their manuscripts to the journal. Despite this policy, investigators showed that the majority of the published studies could not be replicated with the data and code provided. In response to this study the journal changed its policy and in 1996 it mandated that authors deposit their data and code into an archive. From 1996-2003 the journal published 193 empirical articles in which the authors should have deposited their data and code into the archive. However, an analysis of the archive showed that the authors only deposited information in 69 of the cases, often including incomplete information. Replication could only be achieved in 14 cases.

SOURCES: Dewald et al., 1986; McCullough, 2007.

## Safeguards for Scientific Integrity

**The committee recommends that journal editors require authors submitting manuscripts describing clinical evaluations of omics-based tests to provide the journal with the sections of the research protocol relevant to their manuscript (Recommendation 7(a)(iii)).** The research protocol and statistical analysis plan provide a detailed description of the objectives and methods of a study that should be developed at the outset of a study. These include the prespecified primary and secondary outcome measures as well as the prespecified primary analyses of each of these measures. These documents are likely to be more detailed for prospective clinical trials of omics-based tests than for observational studies in omics. However, at a minimum, these documents should specify the research questions being addressed, primary outcomes of interest, and the data analysis strategy.

Previous statements made by organizations about the benefits of journal editors requiring authors to submit their research protocols for clinical trials with their manuscripts have had little impact of journals' practices (Korn and Ehringhaus, 2006). However, requiring authors to share their research protocol and statistical analysis plan with journal editors is an important mechanism to ensure the integrity of an omics study. This allows journal editors and the referees to compare the prespecified outcome measure and analyses plans to the manuscript to make sure they are the same. For trials, a comparison between the sections of the protocol that the authors submit to a registry and what is included in the manuscript also is useful. Often amendments to the protocol and statistical analysis plan are necessary during the conduct of the trial. However, authors should document any amendments and provide an explanation for these changes. Access should be provided to the version of these documents that was in place at the time the outcome data were unblinded to allow data analysis. This prevents investigators from retrospectively accessing the results of the study to modify the principal findings. Both acts of omission (e.g., incomplete reporting of primary and secondary outcome measures) and acts of commission (e.g., unacknowledged changes to prespecified outcome measures) may bias the study (Fleming, 2010; MoS, 2007; Zarin et al., 2011).

Journal editors also can institute policies that enforce the authors' responsibility for the scientific integrity of the manuscript. The committee was informed by the MoS in formulating its recommendation on this topic. The MoS requires journals to obtain a statement from at least one author declaring that he/she "had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis" (MoS, 2007, p. 29). When a study is industry sponsored, this statement should come from an independent investigator (preferably the Principal Investigator), who is not employed by any commercial funding sources. This policy does not excuse all other authors of responsibility for the integrity of the data. Best practices in science require that data be checked repeatedly for quality by multiple members of a research team. An individual author's ability to guarantee the validity of the data is based on his or her trust in the internal data-checking process used in an omics study.

The MoS also states that journals should not publish any industry-sponsored studies when the data analysis was solely conducted by statisticians employed by the company sponsoring the research. Industry-employed statisticians can be listed as authors. To be eligible for publication, the study should include an independent analysis of the data conducted by a statistician at an academic institution or government research institute. The independent statistician should have access to the entire raw data and the study protocol, verify the appropriateness of the analytic plan, and conduct an independent analysis of the data. The manuscript should include the results of this independent analyses (Fontanarosa et al., 2011; MoS, 2007).

The committee also reviewed the policies of the ICMJE in formulating its recommendations. The ICJME policy states that "an author must take responsibility for at least one component of the work, should be able to identify who is responsible for each other component, and should ideally be confident in their co-authors' ability and integrity (ICMJE, 2009a)." Some of the ICMJE journals require that one or more authors guarantee the integrity of the work as a whole, from inception to publication (i.e., the guarantors). The ICMJE also recognizes the importance of reporting guidelines in preparing a manuscript for publication and documenting important information from a study. It encourages authors to consult the reporting guidelines relevant to their specific research design (ICMJE, 2009b). Appendix D discusses reporting guidelines in more detail.

**Based on this review of existing journal policies, the committee recommends that journals require every author to identify their role in the development, conduct, analysis, writing, and editing of the manuscript. Journals also should require the lead and senior authors to attest to the integrity of the study and the co-authors to confirm shared responsibility for study integrity (Recommendation 7[a] [iv]). In addition, the committee recommends that journal editors should require authors submitting manuscripts describing clinical evaluations of omics-based tests to use appropriate reporting guidelines (e.g., the Consolidated Standards of Reporting Trials [CONSORT] (Moher, et al., 2010) and the Reporting recommendations for tumor MARKer prognostic studies [REMARK] (Altman et al., in press; McShane et al., 2005)) and submit checklists to certify guideline use (Recommendation 7[1][v])**.

### Responding to Credible Concerns About Published Manuscripts

Evidence suggests that many biomarker studies inadequately document important aspects of the scientific process (Brundage et al., 2002; Burton and Altman, 2004; Riley et al., 2003). Omics studies also may fail to meet the requirements for transparency and scientific rigor. For example, it has been reported that many tumor biomarker studies are poorly designed; fail to standardize the omics-based test; conduct inappropriate or misleading statistical analyses; and are based on inadequate study sample sizes (Burke and Henson, 1993; Concato et al., 1993; Fielding et al., 1992; Gasparini et al., 1993; Hall and Going, 1999; McGuire, 1991; McShane et al., 2005; Ransohoff, 2002; Ransohoff and Feinstein, 1978; Simon and Altman, 1994). In extreme cases, journals may respond to the technical shortcomings by issuing corrections, retractions, or expressions of concerns (ICMJE, 2009c; MoS, 2007). Journals have a responsibility to respond to credible questions about the scientific integrity and accuracy of the research they publish. However, a recent study found that of 122 leading biomedical journals, only 21 had a retraction policy, 76 had no policy, and the remainder did not post a policy or respond to the researcher's inquiry (Atlas, 2004).

Many existing journals' policies recognize that work demonstrating problems in reproducing the analysis in a paper is fundamentally different from disagreements about interpretation of the results of a study − and should be treated differently. For example, the MoS states that if there is an error in a published paper that can be proved with data, a correction should be published and attached to the original article in PubMed. If there is merely a difference of opinion, a reader can submit a letter through the normal peer review process (MoS, 2007). The challenge for journal editors is determining the correct response and identifying a dispute as an error or a difference of an opinion.

The committee identified a lack of a consistent and clear route to publication of reader-initiated comments and corrections to published papers. Significant scientific disputes rarely have obviously right or wrong solutions and journal editors often have to become very involved to adjudicate disagreements. This may be beyond the resources of a journal. It may be useful for journals to utilize the peer-review process to assess the nature and seriousness of challenges raised by readers. In general, letters to the editor have serious limitations. Letters are not linked to the original PubMed articles and the original author is normally given the opportunity to respond last. Sometimes there is no provision for further correspondence or inquiry regarding disputes after initial letters are exchanged. This can result in the authors delaying or stonewalling the publication of the letter, as journals are reluctant to publish a letter without an author response. **Thus, the committee recommends that journal editors develop mechanisms to resolve possible serious errors in published data, metadata, code, and/or computational models and establish clear procedures for management of error reports (Recommendation 7b).** Potential solutions include creating a mechanism by which investigators who identify a substantive issue in a published paper can submit an erratum notice for peer review and possible publication that links the referred erratum and the original publication in PubMed and on the journal's website. Alternatively, journal editors could ask the original peer reviewers to consider a particularly cogent criticism or proposed correction, if the authors are unresponsive to a letter. Finding a way to link the letter to the original PubMed article is important. Journal editors also could invite peer-reviewed commentaries or editorials that are linked to the key primary articles in PubMed. **The committee also recommends that journals alert the institutional leadership and all authors when a serious question of accuracy or integrity has been raised (Recommendation 7c).** The data and other information needed to investigate an allegation are under the domain of the author's institution (see discussion above on institutional investigations of scientific controversies).

## RECOMMENDATIONS

Both investigators and institutions contribute to the scientific research culture in which omics research is conducted; investigators control the culture of individual laboratories, while institutions put policies and procedures in place that support scientific integrity. While Recommendation 4 focuses on the necessary institutional policies and procedures, these policies and procedures will guide all members of the institution, including investigators.

Institutions that conduct omics research to improve patient care have responsibilities for supporting the integrity of the omics research and the test development process. Although the committee does not intend for these recommendations to create barriers to innovation in this promising technology, it is clear that in the era of omics-based research with its multidisciplinary highly specialized teams and complex data, standard procedures in some institutions do not currently assure the integrity of the scientific process, at either the discovery or test validation phases of omics-based test development. If an institution does not feel it has the infrastructure or capability to follow these recommendations, then the committee believes that such an institution should consider not engaging in research aimed at the development of omics-based tests for use in medical practice, including clinical trials. While this may reduce the number of clinically-oriented studies and publications in omics-based research, if the end result is higher quality publications, this would be a positive change, given the limited resources for research.

**RECOMMENDATION 4: Institutions**

**4a: Institutions are responsible for establishing, supporting, and overseeing the infrastructure and research processes for omics-based test development and evaluation as well as best practices for clinical trials and observational research, including those incorporating omics technologies, and should assure that the evaluation process outlined in this report is followed for omics-based test development and evaluation at their institution.**

**4b: Given the complexity of research and omics-based tests, the multidisciplinary nature of omics and research, and the potential for conflicts of interest in developing and evaluating tests for clinical use, institutional leaders should pay heightened attention to providing appropriate oversight and promoting a culture of scientific integrity and transparency. They should designate:**

- **i. A specific IRB member(s) to be responsible for considering IDE and IND requirements As a component of ensuring the proper conduct of omics-based clinical research.**
- **ii. An institutional official who is responsible for comprehensive and timely documentation, disclosure, and management of financial and non-financial conflicts of interest, both individual and institutional.**
- **iii. An institutional official who is responsible for establishing and managing a safe system for preventing, reporting, and adjudicating lapses in scientific integrity, to enhance patient safety.**
- **iv. An institutional official who is responsible for establishing clear procedures for response to inquiries and/or serious criticism about the science being conducted at the institution. (For example, this individual would be the responsible official for journals to contact with a serious concern about a manuscript; ensure that relevant information to external scientists to help resolve issues of transparency of methods and data, and inform funders when an investigation of potential scientific misconduct is initiated).**

**4c: Institutions that conduct omics research, including clinical trials, should train, recognize, and support the faculty-level careers of individuals from the multiple disciplines that collaborate on omics research and test development including, among others, omics technology, biostatistics, bioinformatics, pathology, and clinical trials and ensure that they are:**

- **i. Treated as equal co-investigators and co-owners of responsibility.**
- **ii. Represented on all relevant review and oversight bodies within the institutions.**
- **iii. Intellectually independent, preferably reporting to an independent mentor and/or department chair as well as to the project leaders.**

The committee also addressed responsibilities of funders, the FDA, and journals in ensuring rigorous development of omics-based tests. Funders play a leadership role in encouraging a culture of integrity and transparency in science, while they seek to accelerate progress through discovery, translation, and clinical applications. The committee highlighted the responsibilities

of funders; funding of independent verification and validation of these tests is particularly important because funders have generally not supported such work, and they do not consider it to be original, innovative science. Without this support, replication and validation will be difficult, and the field will be left with promising ideas published in journals that may be used in clinical practice prematurely or not at all. The FDA should take steps to improve understanding of regulatory requirements for omics-based tests, by directly communicating with investigators and academic institutions and by developing a guidance or regulation that spells out the relevant requirements in this dynamic field. Finally, the responsibilities of journal editors with respect to the adoption and adherence to the omics-based test development and evaluation process are complicated by the wide spectrum of policies adopted and resources available to individual journals.

**RECOMMENDATION 5: Funders**
**5a: All funders of omics-based translational research should:**

i. **Require investigators to make all data, metadata, prespecified analysis plans, code, and fully specified computational models publicly available and readily interpretable either at the time of publication or, if not published, at the end of funding, and funders should financially support this requirement.**
ii. **Provide continuing support for independent repositories to guarantee ongoing access to relevant omics and clinical data.**
iii. **Support test validation in a CLIA-certified laboratory and consider the usefulness of an independent confirmation of a candidate omics-based test prior to evaluation for clinical use.**
iv. **Designate an official to alert the institutional leadership when serious allegations or questions have been raised that may warrant an institutional investigation; if the funder (e.g., the NIH) has initiated that question, then the funder and institution should communicate during the investigation;**
v. **Establish lines of communication with other funders to be used when serious problems appear to involve interdependent research sponsored by another funder along the omics-based test development process.**

**5b: Federal funders of omics-based translational research should have authority to exercise the option of investigating any research being conducted by a funding recipient after requesting an investigation by the institution.**

**RECOMMENDATION 6: FDA**
**6a: In order to enable investigators and institutions to have a clear understanding of their regulatory responsibilities, the FDA should develop and finalize a risk-based guidance or a regulation on:**

vi. **Bringing omics-based tests to the FDA for review.**
vii. **Oversight of LDTs.**

**6b: The FDA should communicate the IDE requirements for use of omics-based tests in clinical trials to the OHRP, IRBs, and other relevant institutional leadership.**

**RECOMMENDATION 7: Journals**
**7: Journal editors should:**
**7a: Require authors submitting manuscripts describing clinical evaluations of omics-based tests to:**

i. **Register all clinical trials at clinicaltrials.gov or another trial registry acceptable to the journal.**
ii. **Make data, metadata, prespecified analysis plans, code, and fully specified computational models publicly available in an independently managed database (e.g., dbGAP) in standard format.**
iii. **Provide the journal with the sections of the research protocol relevant to their manuscript.**
iv. **Identify each author's role in the development, conduct, analysis, writing, and editing of the manuscript. Require the lead and senior authors to attest to the integrity of the study and the co-authors to confirm shared responsibility for study integrity,**
v. **Use appropriate guidelines (e.g., CONSORT, REMARK) and submitting checklists to certify guideline use.**

**7b: Develop mechanisms to resolve possible serious errors in published data, metadata, code, and/or computational models and establish clear procedures for management of error reports.**

**7c: Alert the institutional leadership and all authors when a serious question of accuracy or integrity has been raised.**

## REFERENCES

ACS (American Cancer Society). 2011. *Pilot and Exploratory Projects in Palliative Care of Cancer Patients and their Families.* http://www.cancer.org/acs/groups/content/@researchadministration/documents/document/acspc-023897.pdf (accessed August 10, 2011).

Altman, D. G., L. M. McShane, W. Sauerbrei, and S. E. Taube. In press. Reporting recommendations for tumor marker prognostic studies (REMARK): Explanation and elaboration. *BMC Medicine*.

Altshuler, J. S., and D. Altshuler. 2004. Organizational challenges in clinical genomic research. *Nature* 429 (6990):478-481.

Andre, F., L. M. McShane, S. Michiels, D. F. Ransohoff, D. G. Altman, J. S. Reis-Filho, D. F. Hayes, and L. Pusztai. 2011. Biomarker studies: A call for a comprehensive biomarker study registry. *Nat Rev Clin Oncol* 8(3):171-176.

*Annals of Internal Medicine.* 2010. *Information for Authors.* http://www.annals.org/site/misc/ifora.xhtml (accessed August 22, 2011).

Apweiler, R., C. Aslanidis, T. Deufel, A. Gerstner, J. Hansen, D. Hochstrasser, R. Kellner, M. Kubicek, F. Lottspeich, E. Maser, H. W. Mewes, H. E. Meyer, S. Müllner, W. Mutter, M. Neumaier, P. Nollau, H. G. Nothwang, F. Ponten, A. Radbruch, K. Reinert, G. Rothe, H. Stockinger, A. Tárnok, M. J. Taussig, A. Thiel, J. Thiery, M. Ueffing, G. Valet, J. Vandekerckhove, C. Wagener, O. Wagner, and G. Schmitz. 2009. Approaching clinical proteomics: Current state and future fields of application in cellular proteomics. *Cytometry, Part A* 75(10):816-832.

Atlas, M. C. 2004. Retraction policies of high-impact biomedical journals. *J Med Libr Assoc* 92(2):242-250.

Baggerly, K. A. 2011. *Forensics Bioinformatics*. Presented at the Workshop of the IOM Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30-31.

Baggerly, K. A., and K. R. Coombes. 2009. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Annals of Applied Statistics* 3(4):1309-1334.

Baggerly, K. A., and K. R. Coombes. 2011. What information should be required to support clinical "omics" publications? *Clin Chem* 57(5):688-690.

Baggerly, K. A., J. S. Morris, and K. R. Coombes. 2004. Reproducibility of SELDI-TOF protein patterns in serum: Comparing datasets from different experiments. *Bioinformatics* 20(5):777-785.

Baggerly, K. A., K. R. Coombes, and E. S. Neeley. 2008. Run batch effects potentially compromise the usefulness of genomic signatures of ovarian cancer. *Journal of Clinical Oncology* 26(7):1186-1187.

Baron, A. E., K. Bandeen-Roche, D. A. Berry, J. Bryan, V. J. Carey, K. Chaloner, M. Delorenzi, B. Efron, R. C. Elston, D. Ghosh, J. D. Goldberg, S. Goodman, F. E. Harrell, S. Galloway Hilsenbeck, W. Huber, R. A. Irizarry, C. Kendziorski, M. R. Kosorok, T. A. Louis, J. S. Marron, M. Newton, M. Ochs, J. Quackenbush, G. L. Rosner, I. Ruczinski, S. Skates, T. P. Speed, J. D. Storey, Z. Szallasi, R. Tibshirani, and S. Zeger. 2010. *Letter to Harold Varmus: Concerns about Prediction Models Used in Duke Clinical Trials*. Bethesda, MD, July 19, 2010. http://www.cancerletter.com/categories/documents (accessed January 18, 2012).

Bekelman, J. E., Y. Li, and C. P. Gross. 2003. Scope and impact of financial conflicts of interest in biomedical research. *JAMA: The Journal of the American Medical Association* 289(4):454-465.

Berry, D. 2012. Statisticians and clinicians: Collaborations based on mutual respect. *Amstat News*. http://magazine.amstat.org/blog/2012/02/01/collaborationpolic/ (accessed February 9, 2012).

Bird, S. J. 2001. Mentors, advisors and supervisors: Their role in teaching responsible research conduct. *Science and Engineering Ethics* 7(4):455-467.

Blumenthal, D., N. Causino, E. Campbell, and K. S. Louis. 1996. Relationships between academic institutions and industry in the life sciences—an industry survey. *New England Journal of Medicine* 334(6):368-374.

Bonnefoi, H., A. Potti, M. Delorenzi, L. Mauriac, M. Campone, M. Tubiana-Hulin, T. Petit, P. Rouanet, J. Jassem, E. Blot, V. Becette, P. Farmer, S. Andre, C. R. Acharya, S. Mukherjee, D. Cameron, J. Bergh, J. R. Nevins, and R. D. Iggo. 2007. Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: A substudy of the EORTC 10994/BIG 00-01 clinical trial. *Lancet Oncology* 8(12):1071-1078.

Bonnefoi, H., A. Potti, M. Delorenzi, L. Mauriac, M. Campone, M. Tubiana-Hulin, T. Petit, P. Rouanet, J. Jassem, E. Blot, V. Becette, P. Farmer, S. Andre, C. Acharya, S. Mukherjee, D. Cameron, J. Bergh, J. R. Nevins, and R. D. Iggo. 2011. Retraction−validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: A substudy of the EORTC10994/BIG 00-01 clinical trial. *Lancet Oncology* 12(2):116.

Brazma, A. 2009. Minimum Information About a Microarray Experiment (MIAME)−successes, failures, challenges. *The Scientific World Journal* 9:420-423.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum Information About a Microarray Experiment (MIAME) - toward standards for microarray data. *Nat Genet* 29(4):365-371.

Brown, C. 2003. The changing face of scientific discourse: Analysis of genomic and proteomic database usage and acceptance. *Journal of the American Society for Information Science and Technology* 54(10):926-938.

Brundage, M. D., D. Davies, and W. J. Mackillop. 2002. Prognostic factors in non-small cell lung cancer: A decade of progress. *Chest* 122(3):1037-1057.

Burke, H. B., and D. E. Henson. 1993. Criteria for prognostic factors and for an enhanced prognostic system. *Cancer* 72:3131-3135.

Burton, A., and D. G. Altman. 2004. Missing covariate data within cancer prognostic studies: A review of current reporting and proposed guidelines. *British Journal of Cancer* 91(1):4-8.

Buyse, M., S. Loi, L. J. van't Veer, G. Viale, M. Delorenzi, A. M. Glas, M. S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. Harris, J. Bogaerts, P. Therasse, A. Floore, M. Amakrane, F. Piette, E. T. Rutgers, C. Sortiriou, F. Cardoso, and M. J. Piccart. 2006. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *Journal of the National Cancer Institute* 98(17):1183-1192.

Califf, R. M. 2011a. *Discussion at the Workshop of the IOM Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials*, Washington, DC, March 30-31, Washington, DC,

Califf, R. M. 2011b. Discussion at the Discovery of Process Working Group Meeting with Representatives of Duke Faculty and Administration, Washington, DC, August 22.

Chahal, A. P. S. 2011. Informatics in clinical research in oncology: Current state, challenges, and a future perspective. *The Cancer Journal* 17(4):239-245 210.1097/PPO.1090b1013e31822c31827b31825.

Chan, A. W., and D. G. Altman. 2005. Identifying outcome reporting bias in randomised trials on PubMed: Review of publications and survey of authors. *British Medical Journal* 330(7494):753.

Chan, A. W., K. Krleza-Jeric, I. Schmid, and D. G. Altman. 2004a. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 171(7):735-740.

Chan, A. W., A. Hrobjartsson, M. T. Haahr, P. C. Gotzsche, and D. G. Altman. 2004b. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *Journal of the American Medical Association* 291(20):2457-2465.

Chan, M. M. 2009. *Letter to Division of Medical Oncology, Duke University Medical Center*. http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289102.pdf (accessed February 9, 2012).

Choi, B., S. Drozdetski, M. Hackett, C. Lu, C. Rottenberg, L. Yu, D. Hunscher, and D. Clauw. 2005. Usability comparison of three clinical trial management systems. *AMIA Annu Symp Proc* 2005:921.

Cokol, M., I. Iossifov, R. Rodriguez-Esteban, and A. Rzhetsky. 2007. How many scientific papers should be retracted? *EMBO Rep* 8(5):422-423.

Collins, F. 2010. Has the revolution arrived? *Nature* 464 (7289):674-675.

Concato, J., A. R. Feinstein, and T. R. Holford. 1993. The risk of determining risk with multivariable models. *Annals of Internal Medicine* 118(3):201-210.

Coombes, K. R., J. Wang, and K. A. Baggerly. 2007. Microarrays: Retracing steps. *Nature Medicine* 13(11):1276-1277.

CSE (Council of Science Editors). 2009. *CSE's White Paper on Promoting Integrity in Scientific Journal Publications, 2009 Update.* http://www.councilscienceeditors.org/i4a/pages/index.cfm?pageid=3331 (accessed August 4, 2011 ).

Curfman, G. D., S. Morrissey, and J. M. Drazen. 2006. Response to expression of concern regarding VIGOR study. *New England Journal of Medicine* 354(11):1196-1199.

DeAngelis, C. D., J. M. Drazen, F. A. Frizelle, C. Haug, J. Hoey, R. Horton, S. Kotzin, C. Laine, A. Marusic, A. J. P. M. Overbeke, T. V. Schroeder, H. C. Sox, and M. B. Van Der Weyden. 2004. Clinical trial registration. *JAMA: The Journal of the American Medical Association* 292(11):1363-1364.

DeAngelis, C. D., J. M. Drazen, F. A. Frizelle, C. Haug, J. Hoey, R. Horton, S. Kotzin, C. Laine, A. Marusic, A. J. P. M. Overbeke, T. V. Schroeder, H. C. Sox, and M. B. Van Der Weyden. 2005. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *JAMA: The Journal of the American Medical Association* 293(23):2927-2929.

DeMets, D. L. 2009. *"Minding the Gap": Driving Clinical and Translation Research by Eliminating the Shortage of Biostatisticians.* Bethesda, MD: Clinical Translational Science Award (CTSA) Consortium.

DeMets, D. L., R. Woolson, C. Brooks, and R. Qu. 1998. Where the jobs are: A study of *Amstat News* job advertisements. *The American Statistician* 52(4):303-307.

Deng, M. C., H. J. Eisen, M. R. Mehra, M. Billingham, C. C. Marboe, G. Berry, J. Kobashigawa, F. L. Johnson, R. C. Starling, S. Murali, D. F. Pauly, H. Baron, J. G. Wohlgemuth, R. N. Woodward, T. M. Klingler, D. Walther, P. G. Lal, S. Rosenberg, S. Hunt, and for the CARGO Investigators. 2006. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *American Journal of Transplantation* 6(1):150-160.

Dewald, W. G., J. G. Thursby, and R. G. Anderson. 1986. Replication in empirical economics: The journal of money, credit and banking project. *The American Economic Review* 76(4):587-603.

Dickersin, K., and I. Chalmers. 2010. *Recognising, Investigating and Dealing with Incomplete and Biased Reporting of Clinical Research: From Francis Bacon to the World Health Organization*. http://www.jameslindlibrary.org (accessed June 11, 2010).

Drazen, J., M. B. Van Der Weyden, P. Sahni, J. Rosenberg, A. Marusic, C. Laine, S. Kotzin, R. Horton, P. C. Hebert, C. Haug, F. Godlee, F. A. Frozelle, P. W. Leeuw, and C. D. DeAngelis. 2009. Uniform format for disclosure of competing interests in ICMJE journals. *New England Journal of Medicine* 361(19):1896-1897.

Drazen, J. M., P. W. de Leeuw, C. Laine, C. D. Mulrow, C. D. DeAngelis, F. A. Frizelle, F. Godlee, C. Haug, P. C. Hébert, S. Kotzin, A. Marusic, H. Reyes, and J. Rosenberg. 2010. Toward more uniform conflict disclosures: The updated ICMJE conflict of interest reporting form. *Annals of Internal Medicine* 153(4): 268-269.

Dressman, H. K., A. Potti, J. R. Nevins, and J. M. Lancaster. 2008. In reply. *Journal of Clinical Oncology* 26(7):1187-1188.

Dwan, K., D. G. Altman, J. A. Arnaiz, J. Bloom, A. Chan, E. Cronin, E. Decullier, P. J. Easterbrook, E. Von Elm, C. Gamble, D. Ghersi, J. P. A. Ioannidis, J. Simes, and P. R. Williamson. 2008. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 3(8):e3081.

Emanuel, E. J., D. Wendler, and C. Grady. 2000. What makes clinical research ethical? *JAMA: The Journal of the American Medical Association* 283(20):2701-2711.

Enserink, M. 2011. Authors pull the plug on second paper supporting viral link to chronic fatigue syndrome. *Science* December 28.

FDA (Food and Drug Administration). 2007. *Draft guidance for industry, clinical laboratories, and FDA staff - in vitro diagnostic multivariate index assays*. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079148.htm (accessed February 1, 2012).

FDA. 2011a. *Draft guidance for industry and Food and Drug Administration staff - in vitro companion diagnostic devices*. http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm262292.htm (accessed December 15, 2011).

FDA. 2011b. *FDA establishment inspection report, Duke University Medical Center*. http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289106.pdf (accessed February 9, 2012).

Fielding, L. P., C. M. Fenoglio-Preiser, and S. Freedman. 1992. The future of prognostic factors in outcome prediction for patients with cancer. *Cancer* 70:2367-2377.

Fleming, T. R. 2010. Clinical trials: Discerning hype from substance. *Annals of Internal Medicine* 153:400-406.

Fontanarosa, P. B., A. Flanagin, and C. D. DeAngelis. 2011. Reporting conflicts of interest, financial aspects of research, and role of sponsors in funded studies. *JAMA* 294(1):110-111.

Frankel, M. S. 1995. Commission on research integrity: Origins and charge. In *Professional Ethics Report*. http://www.aaas.org/spp/sfrl/per/per3.htm (accessed August 3, 2011).

Fung, E. T. 2010. A recipe for proteomics diagnostic test development: The ova1 test, from biomarker discovery to FDA clearance. *Clin Chem* 56(2):327-329.

Gasparini, G., F. Pozza, and A. L. Harris. 1993. Evaluating the potential usefulness of new prognostic and predictive indicators on node-negative breast cancer patients. *Journal of the National Cancer Institute* 85(15):1206-1219.

Gawande, A. 2009. *The checklist manifesto: How to get things right*. New York, NY: Metropolitan books.

Geller, G., A. Boyce, D. E. Ford, and J. Sugarman. 2010. Beyond "compliance": The role of institutional culture in promoting research integrity. *Academic Medicine* 85(8):1296-1302.

Hall, P. A., and J. J. Going. 1999. Predicting the future: A critical appraisal of cancer prognosis studies. *Histopathology* 35:489-494.

Helmreich, R. L. 2000. On error management: Lessons from aviation. *BMJ* 320(7237):781-785.

Hsu, D. S., B. S. Balakumaran, C. R. Acharya, V. Vlahovic, K. S. Walters, K. Garman, C. Anders, R. F. Riedel, J. Lancaster, D. Harpole, H. K. Dressman, J. R. Nevins, P. G. Febbo, and A. Potti. 2007. Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *Journal of Clinical Oncology* 25(28):4350-4357.

Hsu, D. S., B. S. Balakumaran, C. R. Acharya, V. Vlahovic, K. S. Walters, K. Garman, C. Anders, R. F. Riedel, J. Lancaster, D. Harpole, H. K. Dressman, J. R. Nevins, P. G. Febbo, and A. Potti. 2010. Retraction to *Journal of Clinical Oncology* 25(28):4350-4357.

Hudson, P. 2003. Applying the lessons of high risk industries to health care. *Qual Saf Health Care* 12(Suppl 1):i7-i12.

ICMJE. (International Committee of Medical Journal Editors). 2009a. *Ethical Considerations in the Conduct and Reporting of Research: Authorship and Contributorship*. http://www.icmje.org/ethical_1author.html (accessed February 2, 2012).

ICMJE. 2009b. *Manuscript preparation and Submission: Preparing a Manuscript for Submission to a BiomedicalJournal.* http://www.icmje.org/manuscript_1prepare.html (accessed February 2, 2012).

ICMJE. 2009c. *Publishing and Editorial Issues Related to Publication in Biomedical Journals: Corrections, Retractions and "Expressions of Concern"* . http://www.icmje.org/publishing_2corrections.html (accessed August 11, 2011).

Ioannidis, J. P. A., and M. J. Khoury. 2011. Improving validation practices in "omics" research. *Science* 334(6060):1230-1232.

IOM. 2002. *Integrity in Scientific Research: Creating an Environment that Promotes Responsible Conduct*. Washington, DC: National Academies Press.

IOM (Institute of Medicine). 2009a. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health through Research*. Washington, DC: The National Academies Press.

IOM. 2009b. *Conflict of Interest in Medical Research, Education and Practice*. Edited by B. Lo and M. Field. Washington, DC: The National Academies Press.

IOM. 2011. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: The National Academies Press.

IOTF (Interagency Oncology Task Force). 2011. *Joint fellowship training program.* http://iotftraining.nci.nih.gov/index.html (accessed October 28, 2011).

Jasny, B. R., G. Chin, L. Chong, and S. Vignieri. 2011. Again, and again, and again. *Science* 334(6060):1225.

Kaiser, J. 2011. *Public Health Service Issues Final Conflicts of Interest Rule*. http://news.sciencemag.org/scienceinsider/2011/08/new-us-conflict-of-interest-rule.html (accessed September 8, 2011).

Klimeck, G. 2011. *Platform for collaborative research with quantifiable impact on research and education*. Paper presented at Cyberinfrastructure Days Conference, Ann Arbor, MI.

Kolata, G. 2001. Johns Hopkins death brings halt to U.S.-financed human studies. *New York Times*, July 20.

Korn, D., and S. Ehringhaus. 2006. Principles for strengthening the integrity of clinical research. *PLoS Clinical Trials* 1(1):e1.

Kornbluth, S. 2011. Discussion at the Discovery of Process Working Group Meeting with Representatives of Duke Faculty and Administration, Washington, DC, August 22.

Kornbluth, S. A., and V. Dzau. 2011. *Predictors of chemotherapy response: Background information: Draft.* Duke University.

Laine, C., R. Horton, C. D. DeAngelis, J. M. Drazen, F. A. Frizelle, F. Godlee, C. Haug, P. C. Hébert, S. Kotzin, A. Marusic, P. Sahni, T. V. Schroeder, H. C. Sox, M. B. Van Der Weyden, and F. W. A. Verheugt. 2007. Clinical trial registration—looking back and moving ahead. *New England Journal of Medicine* 356(26):2734-2736.

Longo, D. R., J. E. Hewett, B. Ge, and S. Schubert. 2005. The long road to patient safety. *JAMA* 294(22):2858-2865.

Lowrance, W. W. 2006. *Access to Collections of Data and Materials for Health Research: A Report to the Medical Research Council and the Wellcome Trust.* http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_grants/documents/web_document/wtx030842.pdf (accessed August 10, 2011).

Marshall, E. 2011. Unseen world of clinical trials emerges from U.S. database. *Science* 333(6039):145.

Martinson, B. C., M. S. Anderson, and R. de Vries. 2005. Scientists behaving badly. *Nature* 435(7043):737-738.

McCullough, B. D. 2007. Got replicability? The journal of money, credit, and banking archive. *Econ Journal Watch* 4(3):326-337.

McGuire, W. L. 1991. Breast cancer prognostic factors: Evaluation guidelines. *J Natl Cancer Inst* 83:154-155.

McKinney, R. 2011. Discussion at Discovery of Process Working Group Meeting with representatives of Duke faculty and administration, Washington, DC, August 22.

McShane, L. M. 2010. *NCI address to Institute of Medicine committee convened to review Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials*. Presented at Meeting 1, Washington, DC, December 20.

McShane, L. 2010a. *Reanalysis report for cisplatin chemosensitivity predictor*. Bethesda, MD: NCI.

McShane, L. M. 2010b. December 20. *Nci address to institute of medicine committee convened to review omics-based tests for predicting patient outcomes in clinical trials*. Meeting 1: Review of Omics Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC.

McShane, L. M., D. G. Altman, W. Sauerbrei, S. E. Taube, M. Gion, and G. M. Clark. 2005. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Journal of the National Cancer Institute* 97(16):1180-1184.

Meldrum, D. R., and A. H. DeCherney. 2011. The who, why, what, when, where, and how of clinical trial registries. *Fertility and Sterility* 96(1):2-5.

Mischak, H., R. Apweiler, R. Banks, M. Conaway, J. Coon, A. Dominiczak, J. Ehrich, D. Fliser, M. Girolami, H. Hermjakob, D. Hochstrasser, J. Jankowski, B. Julian, W. Kolch, Z. Massy, C. Neususess, J. Novak, K. Peter, K. Rossing, J. Schanstra, J. Semmes, D. Theodorescu, V. Thongboonkerd, E. Weissinger, J. Van Eyk, and T. Yamamoto. 2007. Clinical proteomics: A need to define the field and to begin to set adequate standards. *PROTEOMICS− Clinical Applications* 1(2):148-156.

Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gotzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman. 2010. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology* 63(8):e1-e37.

*MoS* (*Manual of Style*). 2007. *AMA Manual of Style: A guide for Authors and Editors*, 10th ed. New York: Oxford University Press, Inc.

Naik, G. 2011a. Mistakes in scientific studies surge. *The Wall Street Journal*, August 10.

Naik, G. 2011b. Scientist' elusive goal: Reproducing study results. *The Wall Street Journal*, December 2.

NAS (National Academy of Sciences). 1992. *Responsible Science, Volume. I: Ensuring the Integrity of the Research Process*. Washington, DC: National Academy Press.

NAS. 2009. *On being a Scientist: A Guide to Responsible Conduct in Research*, 3rd ed. Washington, DC: The National Academies Press.

*Nature.* 2011. *Availability of Data and Material.* http://www.nature.com/authors/policies/availability.html (accessed August 15, 2011).

Nelson, D., and R. Weiss. 1999. *Hasty Decisions in the Race to a Cure? Gene Therapy Study Proceeded Despite Safety, Ethics Concerns.* http://www.washingtonpost.com/wp-srv/WPcap/1999-11/21/101r-112199-idx.html (accessed October 27, 2011).

Nevins, J. 2011. *Genomic Strategies to Address the Challenge of Personalizing Cancer Therapy.* Presented at the Workshop of the IOM Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30-31.

NIH (National Institutes of Health). 1998. *NIH Policy for Data and Safety Monitoring.* http://grants.nih.gov/grants/guide/notice-files/not98-084.html (accessed July 22, 2011).

NIH. 2010. *NIH Grants Policy Statement.* http://grants.nih.gov/grants/policy/nihgps_2010/index.htm (accessed July 22, 2010).

NIH. 2011. Mission. http://www.nih.gov/about/mission.htm (accessed October 19, 2011).

NRC (National Research Council). 1985. *Sharing Research Data*. Washington, DC: National Academy Press.

NRC. 2002. *Integrity in Scientific Research: Creating an Environment that Promotes Responsible Conduct*. Washington, DC: The National Academies Press

NRC. 2003. *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. Washington, DC: The National Academies Press.

NRC. 2005. *Catalyzing Inquiry at the Interface of Computing and Biology*. Washington, DC: The National Academies Press.

NSF (National Science Foundation). 2001. *Grant General Conditions (gc-1)*. http://www.nsf.gov/pubs/2001/gc101/gc101rev1.pdf (accessed August 11, 2011).

ORI (Office of Research Integrity). 2011. *About ORI.* http://ori.hhs.gov/about/index.shtml (accessed September 21, 2011).

OHSR. 2006. *Sheet 6: Guidelines for Writing Informed Consent Documents*. http://ohsr.od.nih.gov/info/sheet6.html (accessed October 27, 2011).

Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, and N. Wolmark. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine* 351(27):2817-2826.

Pathwork Diagnostics. 2010. *Pathwork Tissue of Origin Test for FFPE Cleared by U.S. Food and Drug Administration*. http://www.pathworkdx.com/News/M129_FDA_Clearance_Final.pdf (accessed November 17, 2011).

PCF (Prostation Cancer Foundation). 2011. *Prostate Cancer Research.* http://www.pcf.org/site/c.leJRIROrEpH/b.5780289/k.D2E4/Research.htm (accessed August 10, 2011).

PCSBI (Presidential Commission for the Study of Bioethical Issues). 2011. *Moral science: Protecting participants in human subjects research*. http://bioethics.gov/cms/sites/default/files/Moral%20Science%20-%20Final.pdf (accessed December 21, 2011).

Peng, R. D. 2009. Reproducible research and Biostatistics. *Biostatistics* 10(3):405-408.

Peng, R. D. 2011. Reproducible research in computational science. *Science* 334(6060):1226-1227.

Peng, R. D., F. Dominici, and S. L. Zeger. 2006. Reproducible epidemiologic research. *American Journal of Epidemiology* 163(9):783-789.

Philip, R. O., M. A. Payne, W. Andrew, B. S. Greaves, and T. J. Kipps. 2003. Crc clinical trials management system (CTMS): An integrated information management solution for collaborative clinical research. *AMIA Annu Symp Proc* 2003:967.

PhRMA Foundation. 2011. *2012 Awards in Pharmacology*. http://phrmafoundation.org/download/PhRMA%20Bro_pharmacology.pdf (accessed August 10, 2011).

Piwowar, H. A. 2011. Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE* 6(7):218657.

Piwowar, H. A., and W. W. Chapman. 2008. A review of journal policies for sharing research data. *AMIA Annual Symposium Proceedings* 2008:596–600.

Platt, J. R. 1964. Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* 146(3642):347-353.

Potti, A. 2009. *Letter to FDA's CDER from Division of Medical Oncology, Duke University Medical Center.* http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289103.pdf (accessed February 9, 2012).

Potti, A., and J. R. Nevins. 2007. Potti et al. Reply. *Nature Medicine* 13(11):1277-1278.

Potti, A., H. K. Dressman, A. Bild, R. F. Riedel, G. Chan, R. Sayer, J. Cragun, H. Cottrill, M. J. Kelley, R. Petersen, D. Harpole, J. Marks, A. Berchuck, G. S. Ginsburg, P. Febbo, J. Lancaster, and J. R. Nevins. 2006. Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine* 12(11):1294-1300.

Potti, A., H. K. Dressman, A. Bild, G. Chan, R. Sayer, J. Cragun, H. Cottrill, M. J. Kelley, R. Petersen, D. Harpole, J. Marks, A. Berchuck, G. S. Ginsburg, P. Febbo, J. Lancaster, and J. R. Nevins. 2011. Retraction: Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine* 17(1):135.

Pronovost, P. J., B. Weast, C. G. Holzmueller, B. J. Rosenstein, R. P. Kidwell, K. B. Haller, E. R. Reroli, J. B. Sexton, and H. R. Rubin. 2003. Evaluation of the culture of safety: Survey of clinicians and managers in an academic medical center. *Qual Saf Health Care* 12:405-410.

Quackenbush, J. 2009. Data reporting standards: Making the things we use better. *Genome Medicine* 1(11):111.

Quest Diagnostics. 2011. *Licenses and accreditation*. http://www.questdiagnostics.com/brand/company/b_comp_licenses.html (accessed November 21, 2011).

Ransohoff, D. F. 2002. Challenges and opportunities in evaluating diagnostic tests. *J Clin Epidemiol* 55(12):1178-1182.

Ransohoff, D. F., and A. R. Feinstein. 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 299(17):926-930.

Ranstam, J., M. Buyse, S. L. George, S. Evans, N. L. Geller, B. Scherrer, E. Lesaffre, G. Murray, L. Edler, J. L. Hutton, T. Colton, and P. Lachenbruch. 2000. Fraud in medical research: An international survey of biostatisticians. *Controlled Clinical Trials* 21(5):415-427.

Rennie, D. 1997. Thyroid storm. *JAMA: The Journal of the American Medical Association* 277(15):1238-1243.

Rhodes, R., and J. J. Strain. 2004. Whistleblowing in academic medicine. *J Med Ethics* 30:35-39.

Riley, R. D., K. R. Abrams, A. J. Sutton, P. C. Lambert, D. R. Jones, D. Heney, and S. A. Burchill. 2003. Reporting of prognostic markers: Current problems and development of guidelines for evidence-based practice in the future. *British Journal of Cancer* 88(8):1191-1198.

Rosenberg, S., M. R. Elashoff, P. Beineke, S. E. Daniels, J. A. Wingrove, W. G. Tingley, P. T. Sager, A. J. Sehnert, M. Yau, W. E. Kraus, K. Newby, R. S. Schwartz, S. Voros, S. G. Ellis, N. Tahirkhelli, R. Waksman, J. McPherson, A. Lansky, M. E. Winn, N. J. Schork, E. J. Topol, and for the PREDICT (Personalized Risk Evaluation and Diagnosis In the Conorary Tree) Investigators. 2010. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Annals of Internal Medicine* 153(7):425-434.

SACGHS (Secretary's Advisory Committee on Genetics, Health, and Society). 2010. *Gene patents and licensing practices and their impact on patient access to genetic tests*. http://oba.od.nih.gov/oba/sacghs/reports/SACGHS_patents_report_2010.pdf (accessed January 5, 2012).

Schein, E. 2004. *Organizational culture and leadership*, 3rd ed. The Jossey-Bass Business & Management Series. San Francisco, CA: John Wiley & Sons, Inc.

*Science*. 2011. *General Information for Authors.* http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail (accessed August 15, 2011).

*Science* Staff. 2011. Challenges and opportunities. *Science* 331(6018):692-693.

Segal, M. R., H. Xiong, H. Bengtsson, R. Bourgon, and R. Gentleman. 2012. Querying genomic databases: Refining the connectivity map. *Statistical Applications in Genetics and Molecular Biology* 11(2):1-34.

Sherpa. 2011. *Research Funders' Open Access Policies.* http://www.sherpa.ac.uk/juliet/ (accessed September 9, 2011).

Simon, R. 2008. The use of genomics in clinical trial design. *Clinical Cancer Research* 14(19):5984-5993.

Simon, R . 2010. Clinical trials for predictive medicine: New challenges and paradigms. *Clinical Trials* 7(5):Epub 2010 Mar.

Simon, R., and D. G. Altman. 1994. Statistical aspects of prognostic factor studies in oncology. *British Journal of Cancer* 69(6):979-985.

Sloane, A. 2003. Grading duke: "A" for acknowledgment. *Journal of Health Law* 36(4):627-645.

Song, F., S. Parekh-Bhurke, L. Hooper, Y. Loke, J. Ryder, A. Sutton, C. Hing, and I. Harvey. 2009. Extent of publication bias in different categories of research cohorts: A meta-analysis of empirical studies. *BMC Medical Research Methodology* 9(1):79.

Sprague, R. L., J. Daw, and G. C. Roberts. 2001. Influences on the ethical beliefs of graduate students concerning research. *Science and Engineering Ethics* 7(4):507-516.

Stelfox, H. T., G. Chua, K. O'Rourke, and A. S. Detsky. 1998. Conflict of interest in the debate over calcium-channel antagonists. *New England Journal of Medicine* 338(2):101-106.

Steneck, N. H. 2006. *ORI Introduction to the Responsible Conduct of Research.* http://ori.hhs.gov/education/products/RCRintro/index.html (accessed August 3, 2011).

Stodden, V., and Yale Roundtable Participants. 2010. Reproducible research: Addressing the need for data and code sharing in computational science. *Computing in Science and Engineering* 12(5):8-13.

Taylor, C. F., D. Field, S.-A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, C. A. Ball, P.-A. Binz, M. Bogue, T. Booth, A. Brazma, R. R. Brinkman, A. Michael Clark, E. W. Deutsch, O. Fiehn, J. Fostel, P. Ghazal, F. Gibson, T. Gray, G. Grimes, J. M. Hancock, N. W. Hardy, H. Hermjakob, R. K. Julian, M. Kane, C. Kettner, C. Kinsinger, E. Kolker, M. Kuiper, N. L. Novere, J. Leebens-Mack, S. E. Lewis, P. Lord, A.-M. Mallon, N. Marthandan, H. Masuya, R. McNally, A. Mehrle, N. Morrison, S. Orchard, J. Quackenbush, J. M. Reecy, D. G. Robertson, P. Rocca-Serra, H. Rodriguez, H. Rosenfelder, J. Santoyo-Lopez, R. H. Scheuermann, D. Schober, B. Smith, J. Snape, C. J. Stoeckert, K. Tipton, P. Sterk, A. Untergasser, J. Vandesompele, and S. Wiemann. 2008. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nat Biotech* 26(8):889-896.

Titus, S. L., J. A. Wells, and L. J. Rhoades. 2008. Repairing research integrity. *Nature* 453(7198):980-982.

TMQF Committee (Translational Medicine Quality Framework Committee). 2011. *A Framework for the Quality of Translational Medicine with a Focus on Human Genomics Studies: Principles from the Duke Translational Medicine Quality Framework Committee*. Durham, NC: Duke University.

Turner, E. H., A. M. Matthews, E. Linardatos, R. A. Tell, and R. Rosenthal. 2008. Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine* 358(3):252-260.

van't Veer, L. J., H. Dai, M. J. van de Vijver, Y. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkoven, C. Roberts, P. S. Linsley, R. Bernards, and S. F. Friend. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(31):530-536.

Vedula, S. S., L. Bero, R. W. Scherer, and K. Dickersin. 2009. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *New England Journal of Medicine* 361(20):1963-1971.

Vickers, A. 2008. Cancer data? Sorry, can't have it. *New York Times*, January 22.

Wellcome Trust. 2011. *Sharing Research Data to Improve Public Health: Full Joint Statement by Funders of Health Research.* http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm (accessed August 11, 2011).

Yarborough, M., and R. R. Sharp. 2009. Public trust and research a decade later: What have we learned since Jesse Gelsinger's death? *Molecular Genetics and Metabolism* 97(1):4-5.

Yarborough, M., K. Fryer-Edwards, G. Geller, and R. S. Sharp. 2009. Transforming the culture of biomedical research from compliance to trustworthiness: Insights from nonmedical sectors. *Academic Medicine* 84(4):472-476.

Zarin, D. A. 2005. Clinical trial registration. *New England Journal of Medicine* 352(15):1611.

Zarin, D. A., and T. Tse. 2008. Moving toward transparency of clinical trials. *Science* 319(5868):1340-1342.

Zarin, D. A., T. Tse, R. J. Williams, R. M. Califf, and N. C. Ide. 2011. The ClinicalTrials.gov results database—update and key issues. *New England Journal of Medicine* 364(9):852-860.

# 6
# Lessons from the Case Studies

Omics research is a continually evolving field, with rapidly advancing scientific techniques that are highly dependent on bioinformatics and rigorous statistical methods to effectively interpret high-dimensional data. Despite the challenges inherent in a new field and some setbacks (Dupuy and Simon, 2007; Ransohoff, 2005; Ransohoff and Gourlay, 2010; Simon et al., 2003), a number of omics-based tests derived from this research have reached the market and are being used to manage patient care across a broad range of medical specialties and subspecialties.

The purpose of the case studies was to examine the test development processes used in omics research and omics-based test development to help determine the criteria needed to effectively guide development of omics-based tests, and to consider the roles of the various responsible parties in test development (discussed in Chapter 5). These examples have influenced the emerging field of omics-based tests and will continue to guide the field as more information about these tests accrues. The case studies focus on factors such as:

- The discovery and confirmation of omics-based tests;
- Analytical validation;
- Statistical and bioinformatics validation;
- Clinical/biological validation; and
- Clinical utility and clinical use.

The most extensive case study centers on several omics-based tests developed by a Duke University laboratory to predict sensitivity to chemotherapeutic agents (see Appendix B). These tests were used to select patient therapy in the three clinical trials identified in the committee's statement of task. In addition, the committee examined six tests that are currently commercially available (Table 6-1): Onco*type* DX (Genomic Health), MammaPrint (Agendia), Tissue of Origin (Pathwork Diagnostics), OVA1 (Vermillion), AlloMap (XDx), and Corus CAD (CardioDx), and one test that did not advance to clinical use (Ovacheck). These cases reflect a number of diseases and types of omics-based research. The committee also reviewed the development and use of human epidermal growth factor receptor 2 (HER2) testing, one of the earliest single-biomarker tests to guide the choice of cancer therapy. This last case study illustrates the challenges involved in the development of a single analyte test and suggests how the complexity could be magnified for multianalyte, omics-based tests. Despite many years of research and development, challenges remain in defining the optimal test method and interpretation for HER2 testing.

## METHODS

Examination of what transpired in the development of the omics-based tests at Duke University was based on presentations to the committee, a panel discussion with Duke University

researchers and administrators, documents provided by the National Cancer Institute and Duke University, and the peer-reviewed literature. The Duke case study is illustrative of the more systemic challenges involved in the development of omics-based tests and the need for rigorous criteria for test development, and is discussed in more detail in Appendix B. For the other case studies, the committee reviewed publicly available materials, including peer-reviewed publications and Food and Drug Administration (FDA) 510(k) clearance decision summaries. The committee also invited presentations on the development of Onco*type* DX and MammaPrint (Shak, 2011; van 't Veer, 2011). Each of the six summaries for the commercially available tests was sent to the respective company for review of factual accuracy and completeness.

Detailed case studies were prepared for Onco*type* DX, MammaPrint, Tissue of Origin, and AlloMap. The organization of the detailed case studies reflects the test development process that the committee recommends in Chapters 2-4, including candidate omics-based test discovery and confirmation, development into a defined and validated omics-based test, and evaluation for clinical use. Shorter case studies were assembled for HER2, OVA1, Ovacheck, and Corus CAD, and include details on discovery, development, and clinical use. Case studies appear in Appendix A.

**TABLE 6-1** Overview of Commercially Available Omics-Based Tests

| Test Name | Company and Test Website | Test Indication (diagnostic, prognostic, effect modifier)[a] | FDA[b] Clearance? |
|---|---|---|---|
| Onco*type* DX | Genomic Health<br>www.oncotypedx.com/en-US/Breast.aspx | Prognostic<br>"a 21-gene assay that provides an individualized prediction of chemotherapy benefit and 10-year distant recurrence to inform adjuvant treatment decisions in certain women with early-stage breast cancer" (Genomic Health, 2011).<br>ASCO[c] guidelines state:<br>"Onco*type* DX may be used to identify patients who are predicted to obtain the most therapeutic benefit from adjuvant tamoxifen and may not require adjuvant chemotherapy. In addition, patients with high recurrence scores appear to achieve relatively more benefit from adjuvant chemotherapy than from tamoxifen" (Harris et al., 2007). | No |
| MammaPrint | Agendia<br>www.agendia.com/pages/mammaprint/21.php | Prognostic<br>"… a qualitative in vitro diagnostic test service, performed in a central laboratory, using the gene expression profile of fresh breast cancer tissue samples to assess a patients risk for distant metastasis (up to 10 years for patients less than 61 years old, up to 5 years for patients ≥ 61 years) for breast cancer patients, with Stage I or Stage II disease, with tumor size ≤ 5.0 cm and lymph node negative. [Test] result is indicated for use as a prognostic marker only" (FDA, 2011b). | Yes |
| Tissue of Origin | Pathwork Diagnostics<br>www.pathworkdx.com/TissueOfOriginTest/ | Diagnostic<br>"The Pathwork® Tissue of Origin Test is an in vitro diagnostic intended to measure the degree of similarity between the RNA expression patterns in a patient's formalin-fixed, paraffin-embedded (FFPE) tumor and the RNA expression patterns in a database of fifteen tumor types (poorly differentiated, undifferentiated and metastatic cases) that were diagnosed according to the current clinical and pathological practice" (FDA, 2010). | Yes |
| OVA1 | Vermillion<br>http://ova-1.com/ | Guides referral to a gynecologic oncologist<br>"…a qualitative serum test that combines the results of five immunoassays into | Yes |

| | | | |
|---|---|---|---|
| | | a single numerical score for women with an ovarian adnexal mass present for which surgery is planned as an aid to further assess the likelihood that malignancy is present when the physician's independent clinical and radiological evaluation does not indicate malignancy. The test is not intended as a screening or stand-alone diagnostic assay" (FDA, 2011c). | |
| AlloMap | XDx<br>www.allomap.com/ | Prognostic<br>"an In Vitro Diagnostic Multivariate Index assay (IVDMIA) test service, performed in a single laboratory, assessing the gene expression profile of RNA isolated from peripheral blood mononuclear cells (PBMC[s]). [It] is intended to aid in the identification of heart transplant recipients with stable allograft function [at least 2 months (≥ 55 days) posttransplant] with a low probability of moderate/severe acute cellular rejection (ACR) at the time of testing…" (FDA, 2008). | Yes |
| Corus CAD | Cardio Dx<br>www.cardiodx.com/corus-cad/product-overview/ | Diagnostic<br>"blood test that can quickly and safely assess whether or not [a] patient's symptoms are due to obstructive coronary artery disease (CAD) a decision-making tool that can help identify patients unlikely to have obstructive CAD…" (Cardio Dx, 2011). | No |

[a] The designation here reflects the FDA intended use, if there is one, or if not, the company-proposed use and/or guideline recommendation.
[b] Food and Drug Administration.
[c] American Society of Clinical Oncology.

## LESSONS LEARNED FROM THE CASE STUDIES

The review of the case studies highlighted several key concepts for omics-based test discovery, development, and validation. These include

- Importance of a well-designed development plan;
- Data and code availability;
- Avoidance of overlap between discovery and validation specimens;
- Locking down all aspects of the test prior to evaluation for clinical utility and use;
- Interaction with the FDA;
- Clinical/biological validation characteristics;
- Assessment of clinical utility; and
- Role of the investigators and institution in scientific oversight.

An important point to keep in mind is that these case studies are representative of an early period in the new field of omics-based test development, prior to broad agreement on standard processes and criteria for the various stages of test development. As the scientific community gains experience with omics-based research, more information about ideal test development processes will continue to emerge. Likewise, evidence on clinical utility for these tests will also become richer as more information accrues over time. Recommendations from this committee on the processes for omics-based test development aim to provide needed clarity to the field.

### Importance of a Well-Designed Development Plan

As described in Chapter 2, an important aspect of test development is a well-designed test development plan. Components of a development plan include starting with a clinically meaningful question, developing a candidate test on a training set of specimens, locking down the candidate test, and employing rigorous test validation procedures. The availability of appropriate archival tissue for clinical/biological validation can greatly facilitate rigorous development procedures. If appropriate archival samples are not available, then the development should include a prospective study design.

Tests developed in universities (or where the developmental process is started in universities) are likely to grow out of basic research on the biological meaning of the omics elements that comprise the test and gradually evolve into more formal test development. This is illustrated by the studies of HER2 and MammaPrint, although the development of MammaPrint crossed into a more formalized approach after Agendia was formed. In contrast, companies are likely to start with a focus on a specific test that will eventually have commercial value and thus have a clear development plan early on, as seems to be the case with Genomic Health. According to a presentation to the committee by Steven Shak, chief medical officer of Genomic Health, Onco*type* DX had a clearly articulated development plan. Investigators defined the purpose of the test and subsequently created and implemented a multistep, multistudy approach to develop their omics-based test "to provide the evidence regarding analytic performance, clinical[/biological] validity, and clinical utility to meet the needs of patients, physicians, payors, and regulators" (Shak, 2011).

## Data and Code Availability

As described in Chapter 2, the committee recommends that data and metadata used for identification of an omics-based test should be made available in an independently managed database in standard format, and that code and fully specified computational procedures for omics-based tests should be made sustainably available. Chapters 2 and 5 discuss the importance of sharing data within the scientific community, especially in the setting of omics-based research, where data and analyses are particularly complex. Sharing data can enable external verification of the results and allow other investigators to generate additional insights from the data. In Chapter 5, the committee recommends that journals and funders require the public availability of data, metadata, prespecified analysis plans, code, and fully specified computational models.

The importance of data availability was highlighted in the development of Ovacheck, in which publicly available datasets enabled external researchers to uncover serious problems in experimental design (Baggerly et al., 2004) that ultimately demonstrated that the classifier was based on artifacts and not onto a relevant biological signal. As a result, the test was not implemented in the clinic.

In the Duke case, external investigators did not have full access to the data and code, and this limited the ability to independently evaluate the tests to determine whether they were valid (Baggerly and Coombes, 2009; Baron et al., 2010). Conflicting and unclear information in the papers and cited references regarding the data and statistical methods contributed to the inability of colleagues in the scientific community to understand and replicate the generation of the computational models (Baggerly, 2011; McShane, 2010; *Review of genomic predictors for clinical trials from Nevins, Potti, and Barry,* 2009). When Baggerly and Coombes attempted to assess the validity of the tests, they determined that there was insufficient information to reproduce the published results using the available data and the methods published in the *Nature Medicine* paper (Baggerly, 2011; Potti et al., 2006).

A review of the six commercially available tests illustrates that public availability of all omics-based test data and code has not been the standard of practice (Table 6-2). The field of omics is early in its development and the standards for data sharing have been unclear and only slowly evolving toward more transparency. Commercial interests and protection of proprietary information may also limit the public availability of some data and information.

The cases highlight several examples in which test developers explicitly note the availability of data. For example, Paik et al. (2004), Deng et al. (2006), and Rosenberg et al. (2010) report the computational model for Onco*type* DX, AlloMap, and Corus CAD, respectively. Both tests developed as LDTs had published computational models (Oncotype DX and Corus CAD); only one FDA-cleared test has a published computational model (AlloMap). Discovery microarray data are available for MammaPrint, AlloMap, and Corus CAD (Deng et al., 2006; van 't Veer et al., 2002).[1] Buyse et al. (2006) report that raw microarray data and clinical data for the MammaPrint clinical validation study were deposited with the European Bioinformatics Institute ArrayExpress database. Although there are examples of developers reporting the availability of a test's computational model or data used in discovery or validation, as Table 6-2 shows, there often is not enough information publicly available for external investigators to fully reproduce a test.

[1] Microarray data from Corus CAD are available, but PCR data used in test development are unavailable. Personal communication, Steve Rosenberg, October 21, 2011.

**TABLE 6-2** Data Availability

| Test | Computational Model Published? | Raw Data from Discovery Publicly Available? | Went Through Food and Drug Administration Clearance Processes? | Is the Methodology to Derive the Computational Model Available, and in Sufficient Detail to Fully Reproduce? | Independent Clinical Research Entity Involved? |
|---|---|---|---|---|---|
| Onco*type* DX | Yes (Paik et al., 2004) | Discovery RT-PCR[a] data not available | No | Methodology used to derive and train the computational model is not available in sufficient detail to fully reproduce | Yes (NSABP)[b] |
| MammaPrint | No | Discovery microarray data available | Yes | Some of the details needed for independent replication are unclear from the supplementary materials, but inquiries about the computational model have always been answered[c] | Yes (TRANSBIG)[d] |
| Tissue of Origin | No | Raw data from discovery includes both publicly available (GEO[e] GSE number 2109) and private sources | Yes | Methodology used to derive and train the computational model is not available | No |
| OVA1 | No | Initial raw data used to find a preliminary panel of biomarkers is available; raw data used to finalize the panel of biomarkers in the OVA1 test is unavailable; data on samples used to train the computational model are available in the 510(k) decision summary, but are not available in peer-reviewed literature | Yes | Methodology used to derive the computational model is not available | Yes (PrecisionMed International housed specimens, Quest Diagnostics performed biomarker measurements, Applied Clinical Intelligence performed data analysis) |
| AlloMap | Yes (Deng et al., 2006) | The microarray data used to initially identify biomarkers during the discovery phase of product development are available in GEO | Yes | Unknown. A description of the methods for biomarker discovery, test development, and computational model are detailed in the Deng et al. (2006) supplement | Yes (CARGO[f] investigators) |

| | | | | | |
|---|---|---|---|---|---|
| | | under accession number GSE2445. Raw RT-PCR training data were provided to FDA but not reported in Deng et al. (2006) | | | |
| Corus CAD | Yes (Rosenberg et al., 2010) | Discovery microarray data available (GEO accession number GSE20686) | No | Some of the details needed for independent replication, such as penalization tuning parameters for Ridge regression, are not provided. RT-PCR data used to derive the computational model was not published but are available upon request to qualified investigators[g] | Yes "The analysis was independently performed under the supervision of Dr. Schork at Scripps Translational Science Institute" |

[a] Reverse-transcriptase polymerase chain reaction.

[b] National Surgical Adjuvant Breast and Bowel Project.

[c] Personal communication, Laura van 't Veer, Agendia, November 1, 2011.

[d] A consortium of the Breast International Group (BIG).

[e] Gene Expression Omnibus.

[f] Cardiac Allograft Rejection Gene Expression Observational.

[g] Personal communication, Steve Rosenberg, CardioDX, December 12, 2011.

The committee recognized that it might not always be possible to make this information publicly available due to the protection of intellectual property. For publicly funded research, the committee recommends that code and fully specified computational procedures should be made available at the time of publication or at the end of funding. For commercially developed tests, code and fully specified computational procedures would be submitted for FDA review if seeking approval or clearance, or would be described in a publication in the case of a laboratory-developed test (LDT). Companies that seek FDA clearance or approval for their tests would have had to submit data to the FDA as part of the 510(k) clearance processes or premarket approval (PMA) processes, respectively, but only the information reported in the FDA decision summary is made publicly available.

### Avoidance of Overlap of Discovery and Validation Specimens

One of the challenges of omics-based research is the lack of available specimens on which to conduct correlative science and exploration of candidate omics-based tests (reviewed in IOM, 2010). Partly due to this challenge, a number of tests have been developed with overlapping training and validation datasets. In the cases that the committee reviewed, two commercial omics-based tests used overlapping training and development datasets at some point in their development processes: MammaPrint and AlloMap (Table 6-3). Of the samples used in the development of the MammaPrint computational model, 78 percent were reused in the van de Vivjer et al. (2002) study, leading to criticism about the overlap between training and validation datasets (Kim and Paik, 2010; Ransohoff, 2003, 2004). A subsequent clinical validation study (Buyse et al., 2006) was performed, and suggested that overfitting due to the use of training samples was likely a problem in the 2002 study. In the AlloMap case, the first clinical validation used 63 specimens that had not been used in previous development of the test. This second validation included the 63 primary validation samples plus samples from patients who had contributed samples to the gene discovery and/or diagnostic development phases of the study. Overlap of training and validation datasets can lead to a number of problems in test discovery and development, including overstatement of the accuracy of an omics-based test and incorrect error estimation (Leek et al., 2010). The failure of Ovacheck illustrates the importance of independent validation datasets. Most of the tissue specimens used to validate Ovacheck were obtained from the same institution that provided the specimens used to train the algorithm (Petricoin et al., 2002). The committee recommends in Chapter 2 that test developers should clearly separate training and validation datasets in order to avert these problems, which are among the most common causes of a test ultimately failing clinical validation.

**TABLE 6-3** Statistical and Bioinformatics Validation Considerations

| Test | Lockdown Reported? | Overlap of Discovery and Validation Datasets at Some Point in Test Development? |
|---|---|---|
| Onco*type* DX | Yes (Paik et al., 2004) | No |
| MammaPrint | Yes (Personal communication [a]) | Yes |
| Tissue of Origin | Yes (Monzon et al., 2009; Pillai et al., 2011; personal communication [b]) | No |
| OVA1 | Yes (personal communication [c]) | No |
| AlloMap | Yes (Personal communication [d]) | Yes |
| Corus CAD | Yes (Rosenberg et al., 2010) | No |

[a] Personal communication, Laura van 't Veer, Agendia, November 28, 2011.
[b] Personal communication, Ed Stevens, Pathwork Diagnostics, October 18, 2011.
[c] Personal communication, Scott Henderson, Vermillion, December 12, 2011.
[d] Personal communication, Mitch Nelles, XDx, October 12, 2011.

### Locking Down All Aspects of the Test Prior to Evaluation for Clinical Use

In Chapter 4, the committee recommends that a validated omics-based test should not be changed during a clinical trial. The omics-based test should be locked down during this stage of development (more information on the importance on locking down an omics-based test can be found in Chapters 2, 3, and 4). As noted by Baggerly and colleagues and by McShane, the computational models developed by a Duke University lab were not locked down prior to use in the clinical trials (Baggerly, 2011; McShane, 2010). This was a serious shortcoming in the development of the Duke omics-based tests. For three other cases, the developers explicitly stated in the study publication that their test was locked down before clinical validation: Onco*type* DX (Paik et al., 2004), Tissue of Origin (Monzon et al., 2009; Pillai et al., 2011), Corus CAD (Rosenberg et al., 2010). Communication with test developers generated additional confirmation of lockdown for AlloMap, OVA1, and MammaPrint (Table 6-3).

### Interaction with the FDA

In Chapter 4, the committee recommends that the FDA clarify the regulation of omics-based tests by developing and finalizing a guidance or regulation defining which omics-based tests require FDA review, the type of review required, and when this review should occur. A similar guidance should be developed and finalized for oversight of LDTs that are currently not reviewed by the FDA.

Review of the committee's case studies demonstrates that companies have pursued both LDT and FDA pathways for developing an omics-based test. The use of multiple pathways indicates a lack of clarity and consistency on the regulatory requirements for omics-based tests. Five of the commercially available tests that the committee examined are performed exclusively by each company's proprietary *Clinical Laboratory Improvement Amendments of 1988*- (CLIA-) certified laboratory.[2] Two companies did not seek FDA clearance and market their tests as LDTs: Genomic Health (Onco*type* DX) and CardioDx (Corus CAD). Four companies received FDA 510(k) clearance of their tests: Agendia (MammaPrint), Pathwork Diagnostics (Tissue of Origin), Vermillion (OVA1), and XDx (AlloMap). The publicly available 510(k) clearance decision documents summarize the analytical and clinical validation results that the company/sponsor submitted to the FDA. However, FDA clearance does not mean that a test has clinical utility, and lack of FDA clearance does not mean that a test does not have clinical utility. This is an important distinction between the FDA approval process for drugs and devices (which includes these tests).

In Chapter 5, the committee also recommends that the FDA communicate the Investigational Device Exemption (IDE) requirements for omics-based tests used in clinical trials. Commercial developers, such as those examined in the case studies, may be more familiar with IDE requirements than academic institutions. In several of the case studies, companies and

[2] OVA1 is performed exclusively by Quest Diagnostics, which is subject to CLIA certification (Quest Diagnostics, 2011). Currently Pathwork Diagnostics offers Tissue of Origin exclusively through its CLIA- certified laboratory, but is developing an in vitro diagnostics test kit for other laboratories (Pathwork Diagnostics, 2010).

the FDA held a pre-IDE meeting to determine whether an IDE would be required for the company's test development process. The FDA determined that an IDE was not needed for both the AlloMap and Tissue of Origin tests because the test was not directing patient therapy in the studies proposed to assess the test.[3] Physicians can now use the test for that purpose, however. Agendia reported that they received an IDE for MammaPrint that helped clarify the process and requirements for the de novo 510(k),[4] and Vermillion reported that they received an IDE for OVA1.[5] Two ongoing prospective studies (the TAILORx and RxPONDER trials) direct patient management on the basis of Onco*type* DX Recurrence Score. For both trials, information required for approval of investigational use of Onco*type* DX in the trial was submitted as part of an investigational new drug application to FDA.[6] In the Duke case study, the investigators did consult the FDA regarding the need for an IDE (FDA, 2011a). In 2009, the FDA sent a letter to Duke stating that the omics-based tests being studied in the three clinical trials named in the statement of task needed to go through the IDE process (Chan, 2009). In response, the investigators made some changes to the protocol of the studies and Duke contacted the FDA for further clarification about whether an IDE was still required (FDA, 2011a; Potti, 2009). When the FDA failed to respond to these letters, the Duke Institutional Review Board (IRB) determined that an IDE was not needed (FDA, 2011a). However, in retrospect, the Duke IRB recognized that an IDE should have been obtained for the omics-based tests because the tests were used to direct patient management in the clinical trials (FDA 2011a).Regardless of which pathway is taken to market, consultation with the FDA can be beneficial and is recommended. For example, the developers of the OVA1 test sought FDA input, and this early dialogue with the FDA prompted Vermillion to include two different cut-off values for the test, depending on a patient's menopausal status (Fung, 2010). Although Genomic Health did not meet with the FDA, the company indicated that it benefited from past experience working with the FDA and from the extensive background material FDA provides on its website about assay validation.[7]

### Clinical/Biological Validation Characteristics

The choice of study design for the clinical/biological validation studies of the tests varied widely among the case studies (Table 6-4). Designs included retrospective studies, a prospective–retrospective study (using archived specimens from previously conducted, formal clinical trials that evaluated treatment options that might be affected by the test's use), prospective trials in which the test was not directing therapy, and in the case of the Duke omics-based tests, prospective trials with the tests directing therapy.

As described in Chapter 4, prospective trial designs where biomarkers direct patient management (as in Figures 4-4 and 4-5) are useful approaches to assess the clinical utility of the biomarker. However, they are also the study designs that pose the highest risk to trial participants, because treatment decisions are determined by the test. The justification for using these designs depends substantially on the amount of information known about the test. For example there were no reported validation attempts using clinical tumor samples from patients with lung cancer for the cisplatin test developed by the Duke University lab, even though the first trial in which the cisplatin test was used to guide therapy was the NCT00509366 trial for

[3] Personal communication, Mitch Nelles, XDx, October 12, 2011; Personal communication, Ed Stevens, Pathwork Diagnostics, October 18, 2011.
[4] Personal communication, Laura van' t Veer, Agendia, November 28, 2011.
[5] Personal communication, Scott Henderson, Vermillion, November 1, 2011.
[6] Personal communication, Lisa McShane, National Cancer Institute, February 9, 2012.
[7] Personal communication, Steven Shak, Genomic Health, December 13, 2011.

advanced lung cancer. This type of trial design may have been premature, given the lack of tumor-specific validation of the cisplatin test.

**TABLE 6-4** Choice of Trial Designs for Clinical/Biological Validation

| Test | Clinical Validation Designs |
|---|---|
| Onco*type* DX | Prospective–retrospective (Paik et al., 2004) |
| | Retrospective (Habel et al., 2006) |
| MammaPrint | Retrospective |
| Tissue of Origin | Retrospective |
| OVA1 | Prospective (not directing therapy—trial was for development and validation) |
| AlloMap | Prospective (not directing therapy—trial was for development and validation) |
| Corus CAD | Prospective (not directing therapy—trial was for development and validation) |

As described in Chapter 3, the committee recommended that the identity of the specimens used for clinical/biological validation of the omics-based test should be blinded to the individuals performing and interpreting the test results during clinical/biological validation of the test when feasible. In some instances, this may entail working with another independent group to conduct validation studies. In the case of Onco*type* DX, Genomic Health completed RT-PCR analysis of the specimens used for clinical validation and supplied the results to the National Surgical Adjuvant Breast and Bowel Project, who conducted the analyses evaluating the association between recurrence score and clinical outcome. In the clinical validation for MammaPrint by Buyse et al. (2006), the data were housed at TRANSBIG,[8] and the statistical analyses were conducted by the International Drug Development Institute. For Corus CAD, the publication describing the clinical/biological validation noted that an investigator at the Scripps Translational Science Institute was responsible for the data analyses, while the laboratory work was performed at CardioDX.

The development and validation pathways for Onco*type* DX and MammaPrint offer some interesting comparisons. Kim and Paik (2010) note that both MammaPrint and Onco*type* DX went through these steps either "by design or by demand from the community." Investigators largely chose different strategies for test discovery and development, including the type of tissue used in discovery, the type of risk score (dichotomous variable versus low-, intermediate-, and high-risk categories); different regulatory approaches; and a different intended use population. Onco*type* DX was designed for patients with estrogen-receptor positive, lymph-node negative, early-stage breast cancer. MammaPrint has a potentially larger indication, including patients with estrogen-receptor negative tumors. Despite these differences, MammaPrint and Onco*type* DX have shown around 80 percent agreement in outcome classification (Fan et al., 2006), although they only have one gene overlapping.

[8] A consortium launched by the Breast International Group (BIG) to promote international collaboration in translational research. (See www.breastinternationalgroup.org/Research/TRANSBIG.aspx.)

## Assessment of Clinical Utility

Clinical utility is defined as "evidence of improved measurable clinical outcomes, and [a test's] usefulness and added value to patient management decision-making compared with current management without [the] test" (Teutsch et al., 2009, p. 11). Assessment of clinical utility is not part of the FDA's evaluation of a test, and generally, clinical utility is determined after a test or device is on the market, sometimes decades later. Clinical utility is different from the intrinsic attributes of a test's performance characteristics, such as sensitivity and specificity. Clinical utility is not concerned with how a test performs, but rather how its use influences health outcomes. The ideal way to assess clinical utility is through prospective randomized controlled trials addressing, for example, whether the use of the new test results in an increase in the length or quality of life, a significant increase in progression-free survival, or avoidance of unnecessary treatment (especially toxic treatment) by patients who are unlikely to benefit. Clinical trials assessing clinical utility are important because they can inform how a test is used in practice.

Evidence on clinical utility evolves as new information about a test emerges, and the absence of data on clinical utility should not be interpreted as a lack of utility. A defining feature of both the MammaPrint and Onco*type* DX development stories is the need for a large, prospective trial to provide more information on each test's utility in clinical practice. In both cases, the prospective trial was not initiated until after the test was on the market and widely available for clinical use as a prognostic factor. The results of the prospective studies TAILORx and MINDACT will provide prospective information on the clinical utility of Onco*type* DX and MammaPrint, respectively, for the first time.

## The Role of Investigators and Institutions in Scientific Oversight

Chapter 5 comprehensively discusses the roles of responsible parties in the conduct of omics-based research and omics-based test development. Here, the roles of the investigators and institutions are highlighted for their central role in ensuring rigorous omics-based research and test development. As illustrated in the Duke case study (Appendix B), transparency and open communication are integral to the conduct of science, whether it be reporting of data and code, disclosure of conflicts of interest, or reporting of potential breaches in scientific procedures.

Investigators are responsible for the accuracy of their data, the fairness of their conclusions, and responding appropriately to criticism. Investigators ensure that clinical research is conducted with the engagement of appropriate scientific expertise, including the involvement of individuals with proper biostatistics and bioinformatics expertise, and that the research has the approval of relevant review bodies. It also is important for all members of a research team to understand the aims and intricacies of collaborative studies and for co-authors of a publication to keep each other informed about constructive criticism of the work and ways to improve ongoing research.

Institutions play an important role in establishing a culture of scientific integrity and transparency, including setting expectations of behavior, achievement, and integrity, and providing safe environments for reporting irregularities to prevent lapses in scientific integrity. Institutions are directly charged with being the "oversight" bodies when specific scientific questions or challenges arise, including investigating questions of misconduct, or simply in investigating "soundness of science." Oversight processes that will maintain integrity even in the presence of institutional conflicts of interest, both financial and non-financial (such as factors that impact an institution's reputation) may be especially important in addressing this charge.

Closer attention to such conflicts may have been helpful in avoiding the events that occurred in the Duke case (see Appendix B).

## CONCLUDING REMARKS

The case studies illustrate the multitude of considerations that must be taken into account to move omics-based research into test development for clinical use. The involvement of investigators, institutions, funders, and journals are essential for ensuring good research practices and oversight of omics-based test discovery and development. A well-designed test development plan addresses a clinically meaningful question and employs rigorous test discovery, development, and validation procedures. This includes locking down all aspects of an omics-based test prior to evaluation for clinical utility and use and avoiding overlap between discovery and validation specimens. Choosing an appropriate clinical/biological validation strategy and interacting with FDA prior to initiation of validation studies also reflect a well-designed test development plan. Making data and code available are critical aspects of test development because it enables external verification of the results and generation of additional insights that can advance science and patient care.

## REFERENCES

Baggerly, K. A. 2011. *Forensics Bioinformatics*. Presentation at Workshop of the IOM Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, March 30-31, 2011, Washington, DC.

Baggerly, K. A., and K. R. Coombes. 2009. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Annals of Applied Statistics* 3(4):1309-1334.

Baggerly, K. A., J. S. Morris, and K. R. Coombes. 2004. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 20(5):777-785.

Baron, A. E., K. Bandeen-Roche, D. A. Berry, J. Bryan, V. J. Carey, K. Chaloner, M. Delorenzi, B. Efron, R. C. Elston, D. Ghosh, J. D. Goldberg, S. Goodman, F. E. Harrell, S. Galloway Hilsenbeck, W. Huber, R. A. Irizarry, C. Kendziorski, M. R. Kosorok, T. A. Louis, J. S. Marron, M. Newton, M. Ochs, J. Quackenbush, G. L. Rosner, I. Ruczinski, S. Skates, T. P. Speed, J. D. Storey, Z. Szallasi, R. Tibshirani, and S. Zeger. 2010. *Letter to Harold Varmus: Concerns about prediction models used in Duke clinical trials*. Bethesda, MD, July 19, 2010. http://www.cancerletter.com/categories/documents (accessed January 18, 2012).

Buyse, M., S. Loi, L. J. van 't Veer, G. Viale, M. Delorenzi, A. M. Glas, M. S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. Harris, J. Bogaerts, P. Therasse, A. Floore, M. Amakrane, F. Piette, E. T. Rutgers, C. Sortiriou, F. Cardoso, and M. J. Piccart. 2006. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *Journal of the National Cancer Institute* 98(17):1183-1192.

Cardio Dx. 2011. *What is Corus CAD?* http://www.cardiodx.com/corus-cad/product-overview/ (accessed November 21, 2011).

Chan, M. M. 2009. *Letter to Division of Medical Oncology, Duke University Medical Center*. http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289102.pdf (accessed February 9, 2012).

Deng, M. C., H. J. Eisen, M. R. Mehra, M. Billingham, C. C. Marboe, G. Berry, J. Kobashigawa, F. L. Johnson, R. C. Starling, S. Murali, D. F. Pauly, H. Baron, J. G. Wohlgemuth, R. N. Woodward, T. M. Klingler, D. Walther, P. G. Lal, S. Rosenberg, S. Hunt, and for the CARGO Investigators. 2006. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *American Journal of Transplantation* 6(1):150-160.

Dupuy, A., and R. M. Simon. 2007. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *Journal of the National Cancer Institute* 99(2):147-157.

Fan, C., D. S. Oh, L. Wessels, B. Weigelt, D. S. A. Nuyten, A. B. Nobel, L. J. van 't Veer, and C. M. Perou. 2006. Concordance among gene-expression-based predictors for breast cancer. *New England Journal of Medicine* 355(6):560-569.

FDA (Food and Drug Administration). 2008. *510(k) Substantial equivalence determination decision summary assay and instrument combination template*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K073482.pdf (accessed November 21, 2011).

FDA. 2010. *510(k) Substantial equivalence determination decision summary (k092967)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K092967.pdf (accessed November 16, 2011).

FDA. 2011a. *FDA establishment inspection report, Duke University Medical Center*. http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289106.pdf (accessed February 9, 2012).

FDA. 2011b. *510(k) Substantial equivalence determination decision summary (k101454)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K101454.pdf (accessed September 19, 2011).

FDA. 2011c. *Substantial equivalence determination decision summary (k081754)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K081754.pdf (accessed October 11, 2011).

Fung, E. T. 2010. A recipe for proteomics diagnostic test development: the OVA1 test, from biomarker discovery to FDA clearance. *Clin Chem* 56(2):327-329.

Genomic Health. 2011. *Overview: What is the Oncotype DX assay?* http://www.oncotypedx.com/en-US/Breast/HealthcareProfessional/Overview.aspx (accessed September 14, 2011).

Gutierrez, A. 20101. Discussion with the Committee on the Review of Omics-based Tests for Predicting Patient Outcomes in Clinical Trials, August 19, 2011, Washington, DC

Harris, L., H. Fritsche, R. Mennel, L. Norton, P. Ravdin, S. E. Taube, M. R. Somerfield, D. F. Hayes, and R. C. Bast Jr. 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of Clinical Oncology* 25(33):5287-5312.

IOM (Institute of Medicine). 2010. *A National Clinical Trials System for the 21st Century: Reinvigorating the Cooperative Group Program*. Washington, DC: The National Academies Press.

Kim, C., and S. Paik. 2010. Gene-expression-based prognostic assays for breast cancer. *Nature Reviews* 7(6):340-347.

Leek, J. T., R. B. Scharpf, H. C. Bravo, D. Simcha, B. Langmead, W. E. Johnson, D. Geman, K. Baggerly, and R. A. Irizarry. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 11(10):733-739.

McShane, L. M. 2010a December 20, 2010. *NCI address to Institute of Medicine Committee Convened to Review Omics-Based tests for Predicting Patient Outcomes in Clinical Trials*. Presented at Meeting 1: Review of Omics Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC.

Monzon, F. A., M. Lyons-Weiler, L. J. Buturovic, C. T. Rigl, W. D. Henner, C. Sciulli, C. I. Dumur, F. Medeiros, and G. G. Anderson. 2009. Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin. *Journal of Clinical Oncology* 27(15):2503-2508.

Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, and N. Wolmark. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine* 351(27):2817-2826.

Pathwork Diagnostics. 2010. *Pathwork Tissue of Origin test for FFPE cleared by U.S. Food and Drug Administration*. http://www.pathworkdx.com/News/M129_FDA_Clearance_Final.pdf (accessed November 17, 2011).

Petricoin, E. F., A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta. 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359(9306):572-577.

Pillai, R., R. Deeter, C. T. Rigl, J. S. Nystrom, M. H. Miller, L. Buturovic, and W. D. Henner. 2011. Validation and reproducibility of a microarray-based gene expression test for tumor identification in formalin-fixed, paraffin-embedded specimens. *Journal of Molecular Diagnostics* 13(1):48-56.

Potti, A. 2009. *Letter to FDA's CDER from Division of Medical Oncology, Duke University Medical Center.* http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289103.pdf (accessed February 9, 2012).

Potti, A., H. K. Dressman, A. Bild, R. F. Riedel, G. Chan, R. Sayer, J. Cragun, H. Cottrill, M. J. Kelley, R. Petersen, D. Harpole, J. Marks, A. Berchuck, G. S. Ginsburg, P. Febbo, J. Lancaster, and J. R. Nevins. 2006a. Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine* 12(11):1294-1300.

Quest Diagnostics. 2011. *Licenses and Accreditation*. http://www.questdiagnostics.com/brand/company/b_comp_licenses.html (accessed November 21, 2011).

Ransohoff, D. F. 2003. Gene-expression signatures in breast cancer. *New England Journal of Medicine* 348(17):1716.

Ransohoff, D. F. 2004. Rules of evidence for cancer molecular-marker discovery and validation. *Nature Reviews Cancer* 4(4):309-314.

Ransohoff, D. F. 2005. Bias as a threat to the validity of cancer molecular-marker research. *Nature Reviews Cancer* 5:142-148.

Ransohoff, D. F., and M. L. Gourlay. 2010. Sources of bias in specimens for research about molecular markers for cancer. *J Clin Oncol* 28(4):698-704.

*Review of Genomic Predictors for Clinical Trials from Nevins, Potti, and Barry.* 2009. Durham, NC: Duke University.

Rosenberg, S., M. R. Elashoff, P. Beineke, S. E. Daniels, J. A. Wingrove, W. G. Tingley, P. T. Sager, A. J. Sehnert, M. Yau, W. E. Kraus, K. Newby, R. S. Schwartz, S. Voros, S. G. Ellis, N. Tahirkhelli, R. Waksman, J. McPherson, A. Lansky, M. E. Winn, N. J. Schork, E. J. Topol, and for the PREDICT (Personalized Risk Evaluation and Diagnosis In the Conorary Tree) Investigators. 2010. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Annals of Internal Medicine* 153(7):425-434.

Shak, S. 2011. *Case Study: Oncotype DX Breast Cancer Assay*. Presented at Meeting 2 of the Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30.

Simon, R., M. D. Radmacher, K. Dobbin, and L. M. McShane. 2003. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute* 95(1):14-18.

Teutsch, S. M., L. A. Bradley, G. E. Palomaki, J. E. Haddow, M. Piper, N. Calonge, D. Dotson, M. P. Douglas, and A. O. Berg. 2009. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP working group. *Genetics in Medicine* 11(1):3-14.

van de Vijver, M. J., Y. D. He, L. J. van ‘t Veer, H. Dai, A. A. M. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteven, A. M. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E. T. Rutgers, S. F. Friend, and R. Bernards. 2002. A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* 347(25):1999-2009.

van ‘t Veer, L. J. 2011. *Case study−MammaPrint*. Presented at Meeting 2 of the Committee on Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30.

van ‘t Veer, L. J., H. Dai, M. J. van de Vijver, Y. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Wittereveen, G. J. Schreiber, R. M. Kerkoven, C. Roberts, P. S. Linsley, R. Bernards, and S. F. Friend. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(31):530-536.

# Appendix A
# Case Studies

This appendix reviews the case studies that the committee examined, including six commercially available omics-based tests, an early single-marker test, and an omics-based test that did not advance to clinical use. These case studies appear in the following order:

- Human epidermal growth factor receptor 2 (HER2)
- Onco*type* DX
- MammaPrint
- Tissue of Origin
- OVA1
- OvaCheck
- AlloMap
- Corus CAD

The case study focusing on several omics-based tests developed by a Duke University laboratory to predict sensitivity to chemotherapeutic agents appears in Appendix B.

## HER2

HER2 is one of the earliest biomarker tests for guiding therapeutic decisions, and is widely used in clinical practice. The development of HER2 as an effect modifier biomarker has transformed breast cancer treatment by identifying the 20-30 percent of patients with overexpression of the HER2 oncogene who are likely to benefit from therapy targeting HER2 (De et al., 2010; Phillips et al., 2009). At least seven tests to detect HER2 gene amplification and protein overexpression have Food and Drug Administration (FDA) approval for use as effect modifier markers for tumor response to trastuzumab (reviewed by Shah and Chen, 2010; FDA, 2009b,c). However, despite more than 20 years of research and development, difficulties remain in defining optimal implementation of this single-marker test (De et al., 2010), illustrating some of the profound challenges confronting developers of multianalyte, omics-based tests.

These difficulties include the number of modalities for evaluating HER2 (IHC [immunohistochemistry], FISH [fluorescent in situ hybridization], and others), the subjectivity of test results, lab-to-lab variability (central or reference lab versus smaller labs), laboratory errors leading to false positives and false negatives, differences in cut-off recommendations, and some uncertainty regarding clinical benefit of trastuzumab for patients with borderline HER2-positive results. Accurate selection of patients for therapy targeting HER2, or conversely, identification of those patients who are not likely to

benefit from HER2-targeted therapy, depends on reliable HER2 testing and appropriate cut-off criteria (Kroese et al., 2007).

## HER2 Testing in Clinical Practice

The 2007 American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) panel recommended HER2 status determination for all invasive breast cancers (Wolff et al., 2007a,b) and clarified some of the technical limitations of both IHC and FISH (Schmitt, 2009). In 2002, substantial discordance was reported for both IHC and FISH results performed in community labs versus a central reference lab in the course of two clinical trials (Paik et al., 2002; Roche et al., 2002). In response, the ASCO/CAP panel issued recommendations for the HER2 testing process (e.g., ways to reduce lab-based errors) and interpretation (Wolff et al., 2007a,b). These guidelines alleviated some lab effects within a single HER2 testing modality, though interlab reproducibility continues to be an area of substantial concern for HER2 testing.

The choice of HER2 testing modality is also debated (Sauter et al., 2009; Schmitt, 2009). Historically, IHC has been the primary method for HER2 testing, and FISH has been used to confirm these findings, when IHC testing is equivocal. However, some assert that FISH should be the primary HER2 testing platform (Sauter et al., 2009), while others have advised that IHC alone should never be relied on for selecting anti-HER2 treatment (De et al., 2010). The ASCO/CAP recommendations did not recommend one method for HER2 testing over another, and the National Cancer Institute (NCI) website currently states that "limitations in assay precision make it inadvisable to rely on a single method to rule out potential Herceptin benefit" (NCI, 2011). New methodologies are also in development for HER2 testing, including real-time reverse-transcriptase polymerase chain reaction (qRT-PCR)-based detection of HER2 gene overexpression, which has been presented as the most quantitative platform to date (Baehner et al., 2010).

Differences in cut-off recommendations and equivocal HER2 test results also present challenges. For example, a tumor in which 10 percent of tumor cells show +3 IHC immunoreactivity, and another in which 99 percent of tumor cells display intermediate +2 immunoreactivity might both respond to the same treatment (De et al., 2010). Recently published highly exploratory studies from two of the largest randomized trials of the anti-HER2 therapy trastuzumab have suggested that patients who have some HER2 expression (but below the established cut-off points and not amplified in a FISH test) might benefit from adjuvant trastuzumab (Paik et al., 2008; Perez et al., 2010). A prospective randomized trial (National Surgical Adjuvant Breast and Bowel Project [NSABP] B-47) aims to address this question by randomizing patients with HER2 IHC scores of 1+ or 2+ (but not amplified according to FISH) to chemotherapy plus or minus trastuzumab.

False-positive and false-negative results remain a significant concern for HER2 testing as well. False negatives result in a potentially life-saving anti-HER2 therapy being withheld from a patient. False positives result in treatment with anti-HER2 therapies in the adjuvant or neoadjuvant setting, despite a small chance of benefiting from such treatment. This is a concern given trastuzumab's association with cardiotoxicity, as well as the expense of treatment ($800-$1,000/week for 26-52 weeks) (Sauter et al., 2009).

## Case Highlights

The development of HER2 testing and HER2 targeted therapy represent a significant advance in the treatment of breast cancer and the field of molecularly targeted medicine, but the challenges in implementing HER2 testing in practice have been substantial. Different testing modalities, subjectivity of test results, lab-to-lab variability, false-positive and false-negative results, differences in cut-off recommendations, and some uncertainty regarding clinical benefit of trastuzumab for patients with borderline HER2-positive results make it difficult to determine how best to conduct HER2 testing. There is not yet complete consensus on the standardization of HER2 testing, and as new testing methodologies emerge, new questions about HER2 testing will arise. The challenges involved in developing a single-analyte test such as HER2 are informative as the community is moving toward the development of multianalyte, omics-based tests in which these challenges may be magnified.

## ONCOTYPE DX

Onco*type* DX (Genomic Health Inc.) is a multigene expression test developed to predict the risk of recurrence for node-negative, estrogen receptor-positive breast cancer. Onco*type* DX estimates the likelihood of distant recurrence at 10 years, and classifies individuals at low (scores less than 18), intermediate (18-30) and high (31-100) risk of breast cancer recurrence, assuming the use of adjuvant endocrine therapy, such as tamoxifen and/or the aromatase inhibitors without chemotherapy. Developed as a laboratory developed test (LDT), the test has not been submitted to the FDA for clearance or approval, however Genomic Health indicated that the company benefited from prior interaction with FDA and the extensive background material FDA provides on its website about assay validation.[1] Two ongoing prospective studies (the TAILORx and RxPONDER trials, see section below on Clinical Utility) direct patient management on the basis of Onco*type* DX Recurrence Score. For both trials, information required for approval of investigational use of Onco*type* DX in the trial was submitted as part of an investigational new drug application to FDA.[2]

Developers of Onco*type* DX sought to identify a subgroup of patients who were at such a low risk of recurrence that even if chemotherapy is active, the risks of chemotherapy would outweigh the benefit. Large randomized trials had previously demonstrated the benefit of adding chemotherapy to tamoxifen therapy in patients with estrogen receptor-positive tumors (Berry et al., 2005; EBCTCG, 2005; Fisher et al., 1989, 1997, 2004). Adjuvant chemotherapy studies have generally demonstrated that the relative risk reduction from chemotherapy is constant across risk groups, and studies by the developers of Onco*type* DX in patients with node-negative and node-positive estrogen receptor-positive early breast cancer randomized to chemotherapy suggested that the relative risk reduction of chemotherapy in women with low Recurrence Scores was lower (Albain et al., 2010; Paik et al., 2006). This suggested that the absolute benefit of chemotherapy is lowest for those with the smallest risk of recurrence, and many women treated with tamoxifen alone are likely to remain free of distant recurrence with

[1] Personal communication, Steven Shak, Genomic Health, December 13, 2011.
[2] Personal communication, Lisa McShane, National Cancer Institute, February 9, 2012.

minimal, if any, benefit from the addition of chemotherapy. In this regard, Onco*type* DX is used as a prognostic factor.

## Discovery Phase

The discovery of Onco*type* DX is described by Paik et al. (2004) and the presentation on the "Development and Clinical Validation of Onco*type* DX" located on the Onco*type* DX website (Genomic Health, 2011c). Investigators optimized the methods using a high-throughput, RT-PCR assay for quantifying RNA expression in formalin-fixed, paraffin-embedded (FFPE) tissue (Cronin et al., 2004). Two hundred-fifty candidate genes were selected for assay development based on microarray expression data and information from genomic databases, published literature, and experiments in molecular and cell biology. The relationship between gene expression and recurrence was analyzed in archival tissue from 447 breast cancer patients in 3 separate clinical studies (see Table A-1). Investigators generated a 21-gene panel (16 cancer-related genes and 5 reference genes) and computational model for determining the Recurrence Score. Five steps were used to develop the final gene list and Recurrence Score computational model. First, univarible analysis of each gene was performed separately for the three studies. Second, 16 cancer-related genes were selected based on their performance in predicting recurrence across all three studies. Third, based on coexpression by cluster and principal component analysis, 13 of the 16 genes were put into four gene groups (proliferation, estrogen receptor, HER2, and invasion). Fourth, martingale residual analysis was used to identify linear or non-linear functions forms for each of the gene groups. Fifth, regression analysis performed on each of the three studies was used to select the coefficients for each of the four gene groups and the remaining three individual genes.[3] Additional analyses indicated that inclusion of additional cancer-related genes beyond 16 did not increase the robustness of prediction across the three datasets and that inclusion of fewer than 16 reduced the robustness (Paik et al., 2004). Many, but not all, of the 16 cancer-related genes in Onco*type* DX were previously well established in the cancer literature for their association with prognosis (Kim and Paik, 2010).

**TABLE A-1** Archival Tissue Used in the Development of Onco*type* DX Computational Model and Gene List

| Study | Paik et al., 2003 | Cobleigh et al., 2005 | Esteban et al., 2003 |
|---|---|---|---|
| Tissue Source | Tamoxifen arm of NSABP[a] B-20 | Rush Presbyterian-St. Luke's Hospital | Providence St. Joseph's Hospital |
| Sample # | 233 | 78 | 136 |
| Lymph-node status | Negative | > 10 positive nodes | Positive or negative |
| Estrogen-receptor Status | Positive | Positive and negative | Positive and negative |
| Treatment | Tamoxifen (100%) | Tamoxifen (54%)<br>Chemotherapy (80%) | Tamoxifen (41%)<br>Chemotherapy (39%) |

[a]National Surgical Adjuvant Breast and Bowel Project.
SOURCE: Shak (2011).

[3] Personal communication, Steven Shak, Genomic Health, December 13, 2011.

## Test Validation Phase

More than 150 standard operating procedures (SOPs) were developed for the 5-step 21-Gene Recurrence Score, including SOPs for equipment, histopathology, information technology, pre- and postanalytical methods, production and quality control, and quality assurance (Shak, 2011).

### *Analytical validation*

The analytical validity of Onco*type* DX was assessed in the Agency for Healthcare Research and Quality (AHRQ) report, *Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes* (AHRQ, 2008). The report noted there is evidence about Onco*type* DX's assay performance and laboratory characteristics as well as some limited information on its reproducibility. Cronin et al. (2007) found that Onco*type* DX met acceptable operational performance ranges with minimal assay imprecision due to instrument, operator, reagent, and day-to-day baseline variation. Investigators also conducted technical feasibility studies during assay development, including analysis of preanalytical factors such as variability in preparation, tumor block age, and dissection (Shak, 2011). Paik and colleagues (2004) measured and reported the reproducibility within and between blocks in the clinical validation study.

### *Statistical and Bioinformatics Validation*

The computational model to determine the Recurrence Score is published, and there is public information that provides an overview of how the computational model was generated (see Discovery Phase) (Paik et al., 2004). The supplementary materials of Paik et al. (2004) note that the investigators weighted the NSABP B-20 results most heavily in selecting the final gene list and developing the computational model because investigators planned to clinically validate the test in similar archival tissue from NSABP B-14 patients. However, more detailed information on model development and the RT-PCR and clinical data used in the development of 21-Gene RS are not publicly available.[4]

The test was locked down prior to clinical validation. The investigators reported that "the prospectively defined assay methods and endpoints were finalized in a protocol signed on August 27, 2003, and RT-PCR data were transferred to the NSABP for analysis on September 29, 2003" (Paik et al., 2004, p. 2820). Genomic Health was blinded to the clinical outcome data until the RT-PCR data were locked and transferred to NSABP.[5]

### *Clinical/Biological Validation*

Archival tissue from breast cancer patients in three studies was used to clinically validate the prognostic value of Onco*type* DX (Table A-2). Paik et al. (2004) found that the Recurrence Score quantified the likelihood of distant recurrence in tamoxifen-treated patients with lymph node-negative, estrogen receptor-positive breast cancer. Investigators prospectively defined the endpoints for validation and prespecified the cut-off values for low, intermediate, and high risk of recurrence. They had a large number of patient

[4] Personal communication, Steven Shak, Genomic Health, December 13, 2011.
[5] Personal communication, Steven Shak, Genomic Health, December 13, 2011.

samples on which to clinically validate the prognostic value of the Recurrence Score, and did not use samples from the discovery phase in the validation studies. Although this study was not a true prospective clinical validation, many assert the prospective-retrospective study design has evidentiary value close to a prospective study (AHRQ, 2008; Harris et al., 2007; Simon et al., 2009).

The second study, Habel et al. (2006), assessed the prognostic value of the Recurrence Score using archival tissue from patients treated within the Northern California Kaiser Permanente health plan. Investigators found that the Recurrence Score was associated with the risk of breast cancer death among patients with estrogen receptor-positive breast cancer who were treated with tamoxifen or were not treated with systemic adjuvant therapy.

The third study, Esteva et al. (2005), used a smaller number of archival tissue samples and did not find an association between the Recurrence Score and risk of distant recurrence. Investigators hypothesized that this result could be due to potential selection bias or confounding factors. However, investigators did find a high degree of concordance between RT-PCR and immunohistochemical assays for estrogen receptor, progesterone receptor, and HER2.

**TABLE A-2** Clinical/Biological Validation Studies for Onco*type* DX

| Study | Paik et al. (2004) | Habel et al. (2006) | Esteva et al. (2005) |
|---|---|---|---|
| Tissue source | NSABP[a] B-14 | Kaiser Permanente | MD Anderson Cancer Center |
| Study question | Does 21-Gene RS[b] correlate with likelihood of distant recurrence? | Does 21-Gene RS predict risk of breast cancer-specific mortality in women treated and not treated with tamoxifen? | Does 21-Gene RS predict risk of recurrence in women not treated with systemic therapy? |
| Study design | Prospective–retrospective | Retrospective; matched case control<br><br>Cases = patients who died from breast cancer<br>Controls = breast cancer patients individually matched to cases alive at the date of death of their matched case | Retrospective with case inclusion criteria |
| Patient characteristics | Lymph-node negative; ER+[c] | Lymph-node negative; ER +/- | Lymph-node negative; ER +/- |
| Treatment | Tamoxifen; no chemotherapy | +/-Tamoxifen; no chemotherapy | No systemic therapy |
| Sample # | 668 | 220 cases, 570 controls | 149 |
| Blinding | Yes | Yes | Yes |

| | | | |
|---|---|---|---|
| Independence | Different specimens than used in discovery<br><br>NSABP control of clinical outcome data | Different specimens than used in discovery<br><br>Kaiser Permanente control of clinical outcome data | Different specimens than used in discovery<br><br>MD Anderson control of clinical outcome data |
| Results | Rate of recurrence significantly lower ($p < 0.001$) with low-risk RSs compared to high-risk RSs; RS provided significant predictive power independent of age and tumor size; RS was predictive of overall survival | RS associated with risk of breast cancer death in ER+, tamoxifen-treated and – untreated patients ($p = 0.003$ and $p = 0.03$, respectively) | No association between RS and distant recurrence-free survival in ER+/- patients with no adjuvant systemic therapy |

[a] National Surgical Adjuvant Breast and Bowel Project
[b] Recurrence Score
[c] Estrogen receptor

**Chemotherapy benefit** In an exploratory analysis designed to assess the test's ability to predict the benefit of chemotherapy treatment, investigators used archival tissue from the NSABP B-20 study, in which patients were randomized to tamoxifen or tamoxifen plus chemotherapy. There were two chemotherapy arms: the cyclophosphamide, methotrexate, and fluorouracil (CMF) arm and the methotrexate and fluorouracil (MF) arm (Paik et al., 2006). In this study, there appeared to be a relative treatment modifier effect independent of the prognostic role of Onco*type* DX ($p = 0.038$ for interaction between Recurrence Score and chemotherapy treatment). Prognosis was more favorable in patients with a low Recurrence Score who received only tamoxifen, and chemotherapy did not appear to be active in this subgroup. In contrast, patients with a high risk of recurrence based on their Recurrence Score had a worse prognosis if treated with tamoxifen only, and achieved a large benefit from chemotherapy. Tissue from the tamoxifen plus chemotherapy arms were not previously used in the development of Onco*type* DX, but tissue from the tamoxifen-only arm had been previously used in the test discovery phase. As noted in Chapter 2, the use of discovery phase tissue samples to assess test performance is not ideal because it can lead to overfitting. The TAILORx trial (see below) will provide higher quality evidence to assess the benefit from chemotherapy treatment in a subset of patients with Recurrence Scores of 11-25 because it will prospectively evaluate the impact of Onco*type* DX on treatment within a large, randomized clinical trial population and will not use tissue samples from the discovery phase.

**Lymph node-positive patients** Tumor samples from lymph node-positive patients were used to help develop Onco*type* DX (Cobleigh et al., 2005), and a recent prospective-retrospective analysis of a large trial found that the Recurrence Score is prognostic for

tamoxifen-treated patients with positive nodes and, as expected, their prognosis was worse than for patients with negative lymph nodes (Albain et al., 2010). The analysis also evaluated the effect modifier role of Onco*type* DX and indicated that node-positive patients with low Recurrence Scores did not benefit from chemotherapy treatment, but node-positive patients with high Recurrence Scores had an improvement in disease-free survival when treated with chemotherapy. The relative effects of chemotherapy rose with increasing Recurrence Scores. The RxPONDER trial (see below) will provide more information on the clinical utility of Onco*type* DX in lymph node-positive patients.

## Clinical Utility

### *TAILORx Trial*

The NCI initiated the Trial Assigning IndividuaLized Options for Treatment (TAILORx) in 2006 (now fully accrued) to assess the performance of Onco*type* DX in a large, prospective, randomized clinical trial. The primary objective of TAILORx is to assess the effect of chemotherapy, in addition to hormonal therapy, in women with Recurrence Scores between 11 and 25.[6] The benefit of chemotherapy for women in this mid-range risk group is currently unclear. The study involves over 10,000 women recently diagnosed with estrogen receptor-positive and/or progesterone-receptor positive, HER2-negative breast cancer without lymph node involvement at 900 sites in the United States, Canada, and several other countries outside of North America. Women with a low Recurrence Score received hormonal therapy alone while women with a high Recurrence Score received hormonal therapy and chemotherapy. Women with Recurrence Scores in the mid-range risk group were randomized to receive either chemotherapy plus hormonal therapy or hormonal therapy alone. Women will be studied for 10 years, with additional follow-up 20 years after initial therapy (NCI, 2006, 2010a,b).

### *RxPONDER Trial*

Onco*type* DX is also being evaluated in lymph node-positive patients in a prospective trial, RxPONDER (Rx for POsitive NoDe, Endocrine Responsive breast cancer), which will recruit 4,000 patients with Recurrence Scores of 25 or less who have estrogen-receptor positive tumors and 1-3 positive lymph nodes. Patients will be randomly assigned to treatment with chemotherapy plus hormonal therapy or hormonal therapy alone. The trial seeks to determine whether these women may safely forego chemotherapy treatment and whether there is an optimal Recurrence Score cutpoint for recommending chemotherapy or not (SWOG, 2011).

[6] While the Oncotype validation studies prespecified cutoff values of 0-18, 18-30, and 31 and above for low, intermediate, and high risk of recurrence, the TAILORx investigators defined a mid-range risk of recurrence as scores of 11-25 to roughly correlate with a 10 to 20 percent risk of distant recurrence at 10 years. In TAILORx, patients classified at mid-range risk will be randomized to receive either hormonal therapy or hormonal therapy and chemotherapy, while patients at very low risk (below 11) will be assigned to hormonal therapy only and those at high risk (above 25) will receive hormonal therapy and chemotherapy.

## Clinical Use

As an LDT, Onco*type* DX is performed in Genomic Health's *Clinical Laboratory Improvement Amendments of 1988* (CLIA) -certified laboratory. Onco*type* DX has been incorporated into guidelines from ASCO and the National Comprehensive Cancer Network (NCCN). The ASCO *2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer* stated that "Onco*type* DX may be used to identify patients who are predicted to obtain the most therapeutic benefit from adjuvant tamoxifen and may not require adjuvant chemotherapy. In addition, patients with high recurrence scores…appear to achieve relatively more benefit from adjuvant chemotherapy (specifically [C]MF) than from tamoxifen" (Harris et al., 2007, p. 5299). The guidelines specify that there are insufficient data to suggest whether these conclusions can be generalized to other hormonal therapies (e.g., aromatase inhibitors) or other chemotherapy regimens. However, a recent study using specimens from the Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial found that the Recurrence Score was an independent predictor of distant recurrence in women with node-negative and node-positive, hormone receptor-positive patients treated with anastrozole (Arimidex), an aromatase inhibitor (Dowsett et al., 2010). NCCN guidelines note that Onco*type* DX is an option when evaluating certain patients with breast cancer, and assert that "the recurrence score should be used for decision-making only in the context of other elements of risk stratification for an individual patient" (NCCN, 2011a, p. 85).

Over 7,500 physicians have ordered the Onco*type* DX test for more than 175,000 patients (Genomic Health, 2011a), with 55,000 Onco*type* DX tests ordered in the past year.[7] Onco*type* DX is covered by almost all private insurers and is a covered benefit for Medicare beneficiaries and some Medicaid beneficiaries (Genomic Health, 2011b). The Blue Cross Blue Shield Technology Evaluation Center (TEC) found that the use of Onco*type* DX meets the TEC criteria (BCBS, 2008). The current list price for the assay is $4,175.[8]

A meta-analysis of 912 patients found that physicians using Onco*type* DX in clinical practice altered their treatment decisions in more than one third of patients, leading to a 28 percent reduction in the use of chemotherapy (Hornberger and Chien, 2010).

## Case Highlights

According to a review by AHRQ, Onco*type* DX is one of the more well-established omics-based breast cancer tests due to its validation pathway (AHRQ, 2008). In a presentation to the committee, Steven Shak, chief medical officer of Genomic Health, stated that the company had a clearly articulated development plan that involved a multistep, multistudy approach.

The NSABP B-14 clinical validation was a large, blinded, prospective-retrospective study that provided evidence for the test's ability to discriminate among low, intermediate, and high risk of distant recurrence among a well-defined patient cohort with node-negative, estrogen receptor-positive cancer who were treated with tamoxifen, but not with chemotherapy. Although this was not a true prospective clinical validation study, many assert that this study design has an evidentiary value close to a prospective study

[7] Personal communication, Steven Shak, Genomic Health, December 13, 2011.
[8] Personal communication Steven Shak, Genomic Health, December 13, 2011.

(AHRQ, 2008; Harris et al., 2007; Simon et al., 2009). The Kaiser clinical validation study demonstrated that the Recurrence Score was associated with risk of breast cancer death among a population-based sample. Ongoing clinical trials will further inform clinical use of Onco*type* DX, including the benefit of chemotherapy among women with intermediate Recurrence Scores and women with 1-3 positive lymph nodes.

Onco*type* DX was developed as an LDT without FDA review. The computational model for Onco*type* DX was published, but several aspects of test development are not specified in detail in the published literature, including how the 250-gene list was selected during test discovery and how the archival tissue from three clinical studies was used in test training and development of the computational model. Gene expression and clinical data from the discovery phase are not publicly available.

## MAMMAPRINT

MammaPrint is a prognostic test designed to predict the risk of recurrence of distant breast cancer following surgery for patients with both estrogen receptor-positive and -negative tumors. MammaPrint uses a 70-gene RNA expression signature to classify individuals as having either high or low risk of recurrence. MammaPrint was developed by investigators at the Netherlands Cancer Institute, who founded a spin-off company, Agendia, to develop the commercial test.

Agendia first met with the FDA in 2005, in which a pre-IDE was submitted. In 2006, FDA approved Agendia's IDE. The IDE clarified the process and requirements for the de novo 510(k), and yielded useful customer information. Agendia subsequently submitted a draft 510(k) in June 2006 and de novo 510(k) in September 2006.[9] In 2007, MammaPrint became the first FDA-cleared molecular test profiling genetic activity (FDA, 2007b).

### Discovery Phase

The 70-gene signature was developed using archival samples of primary invasive breast tissue from 78 breast cancer patients (34 patients developed distant metastases within 5 years, 44 patients were disease-free after 5 years) (van't Veer et al., 2002). All patients were lymph-node negative, under age 55, and had tumors of less than 5 centimeters. Only 6 percent of patients received adjuvant systemic therapy.

RNA was isolated from snap-frozen tissue. Each RNA sample underwent 2 hybridizations on microarrays with 25,000-gene sequences. An intensity ratio was calculated using a reference RNA pool containing equal amounts of RNA from each tissue sample.

Investigators used an unsupervised hierarchical clustering algorithm that identified approximately 5,000 genes with expression significantly increased or decreased relative to random chance in more than 3 tumors out of the 78.

Using supervised classification, investigators found that 231 genes were significantly associated with disease outcome. Subsets of five genes were sequentially added to evaluate their power in correct classification using the leave-one-out method for

[9] Personal communication, Laura van 't Veer, Agendia, November 28, 2011.

cross-validation.[10] The optimal gene expression signature was composed of 70 genes and correctly predicted whether the patient was still recurrence free or not at 5 years for 65 of 78 patients (83 percent). To reduce the number of false negatives (patients who actually had recurred but who were identified by the classifier as having a good prognosis), investigators set the threshold so that no more than 10 percent of patients with a poor prognosis were misclassified. This optimized sensitivity (as opposed to optimized specificity) threshold resulted in 15 misclassifications (81 percent correct) (van't Veer et al., 2002).

## Test Validation Phase

### *Analytical validation*

The test platform was converted into a new microarray, MammaPrint, containing the 70 genes identified in the discovery phase (Glas et al., 2006). Investigators reanalyzed RNA from 162 patient samples from the discovery and clinical validation (described in the next section) of the 70-gene signature to confirm that the results from the commercial microarray were consistent with the results generated on the discovery phase microarray. Investigators reported that the original analyses and reanalysis on the commercial MammaPrint microarray showed high correlation of prognosis prediction ($p < 0.0001$), with seven discordant cases.

High intralaboratory and interlaboratory reproducibility was reported for MammaPrint in three different laboratories when RNA from four different patient samples was assessed (Ach et al., 2007). A report from AHRQ, *Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes*, noted that evidence to support analytic validity was obtained from a limited number of patients and a moderate number of replication experiments, and the impact of RNA labeling variation on risk classification was not thoroughly investigated (AHRQ, 2008).

The FDA accepted a modification to Agendia's 510(k) premarket notification, in which the specimen type was switched from fresh frozen tissue to fresh tissue stored in an RNA preservative solution (FDA, 2007a). Investigators showed that shipment in the RNA preservative did not affect the MammaPrint test results, with no statistically significant difference in MammaPrint risk group assignment or index between fresh frozen and preserved tissue (FDA, 2007a).

### *Statistical and Bioinformatics Validation*

Discovery microarray data and clinical information are available and reported in van 't Veer et al. (2002), which was uncommon at the time of its publication.[11] Although the method used to derive the MammaPrint computational model is described in the

[10] Leave-one-out cross-validation is a statistical method used to assess the generalizability of an analysis to an independent dataset. One observation is removed from the dataset to use as the test set and the remaining observations are used as the training set. This process is repeated for all observations and the results of all iterations are averaged.

[11] Microarray data were first hosted at http://www.rii.com/publications/default.htm, as mentioned in the van't Veer (2002) paper and are now available through the Netherlands Cancer Institute website (http://bioinformatics.nki.nl/data.php) and the Stanford open access microarray site. Personal communication, Laura van't Veer, Agendia, November 1, 2011.

supplementary materials, the final computational model is not reported, and some of the details needed for independent replication are unclear from the supplementary materials, including the methodology for gene selection and details about the statistical analysis.

Several statistical problems in the early development studies were reported in the published literature (Ransohoff, 2003; Simon, 2003). There was overlap of 61 discovery samples used in the second clinical validation study, which was acknowledged by the authors in the van de Vijver paper (van de Vijver et al., 2002). Model performance was assessed using the same data that were used in development of the 70-gene signature, and Simon (2003) asserted that an incomplete cross-validation process was performed in van't Veer et al. (2002) because the method did not include reselection of the differentially expressed genes. According to Simon and colleagues, this resulted in a biased underestimate of the error rate that, when combined with the small number of samples used in discovery and validation, likely led to overfitting and overstatement of the accuracy of the 70-gene signature (Simon et al., 2003). A subsequent study was performed in acknowledgment of some of these shortcomings (Buyse et al., 2006).

Microarray data and clinical information from the third clinical validation study are publicly available (Buyse et al., 2006).[12] A statistician independently applied the MammaPrint computational model to reproduce the MammaPrint risk classification, and independent statisticians assessed the concordance (100 percent) between risk classification produced by Agendia and the external statistician. Independent auditors visited clinical centers to carry out source data verification; statistical analyses were conducted by the International Drug Development Institute; and clinical, pathologic, and gene signature data were centralized at the TRANSBIG Secretariat.

The MammaPrint test was locked down twice. The 70-gene signature research assay was locked down after its development described in van 't Veer et al. (2002), and as the commercial microarray (Glas et al., 2006) as developed by Agendia.[13]

*Clinical/Biological Validation*

Table A-3 lists the clinical validation studies that have been performed to assess MammaPrint. The first validation, published as part of the discovery phase paper, assessed 19 patient samples. The authors reported that the 70-gene signature predicted disease outcome better than traditional clinical prognostic factors (van't Veer et al., 2002).

A second study evaluated the 70-gene signature in 295 patients, which, as noted previously, included 61 samples that were used in the discovery phase of test development (representing 78 percent of the tumor samples in the development of the signature) (van de Vijver et al., 2002). Investigators asserted that leaving out these patient samples would have resulted in selection bias because the signature was developed using a disproportionately large number of patients in whom distant metastases developed within 5 years (van de Vijver et al., 2002). However, critics do not consider this a true validation because of the overlap between training and validation datasets (Kim and Paik, 2010; Ransohoff, 2003, 2004).

[12] Accession number E-TABM-77 (European Bioinformatics Institute ArrayExpress database).
[13] Personal communication, Laura van 't Veer, University of California, San Francisco, November 28, 2011.

A third validation study (Buyse et al., 2006) was performed using archived samples not collected as part of a prospective clinical trial protocol from 307 patients in 5 European centers. The authors concluded that MammaPrint outperformed traditional prognostic factors in predicting distant metastases, overall survival, and disease-free survival. However, they also noted that the hazard ratios[14] reported in the earlier study (van de Vijver et al., 2002) were much higher than those reported in Buyse et al. (2006) (see Table A-3), echoing earlier concerns "that the inclusion in the [van de Vijver et al. 2002 study] of patients whose data were used in the development of the 70-gene signature may have inflated the discriminatory power of the signature in that study, even though analytic measures had been taken to limit this effect" (Buyse et al., 2006, p. 1190). The authors also suggested that the longer period of follow up in the third study may have also contributed to this difference in hazard ratios.

**TABLE A-3** Clinical/Biological Validation Studies for MammaPrint/70-gene Expression Signature

| Study | van't Veer et al. (2002) | van de Vijver et al. (2002) | Buyse et al. (2006) |
|---|---|---|---|
| Tissue Source | Netherlands Cancer Institute | Netherlands Cancer Institute | 5 European centers |
| Test platform | Experimental microarray | Experimental microarray | MammaPrint (Glas et al., 2006) |
| Study question(s) | Does 70-gene expression signature show comparable performance to development tissues? | Does 70-gene signature confirm results from previous validation study in lymph node-negative patients? What is the performance of signature in lymph node-positive patients? | Does MammaPrint have prognostic value in a group of independent patients, beyond clinical risk classifications? |
| Study design | Retrospective | Retrospective | Retrospective |
| Patient characteristics | Under 55 years old<br>Lymph-node negative<br>Tumor size < 5 cm | Under 53 years old<br>Lymph-node negative and positive<br>Tumor size < 5 cm | Under 61 years old<br>Lymph-node negative<br>Tumor size < 5 cm |
| Treatment | No adjuvant systemic therapy | 10 lymph node-negative patients and 120 lymph node-positive patients received adjuvant chemotherapy (n = | No adjuvant systemic therapy |

[14] A hazard ratio is an expression of the risk of an event in one arm of a study as compared to the risk of the event happening in the other arm over time. This differs from the relative risk ratio, which is a proportion of the number of events that occur in one arm of the study as compared to the other arm.

| | | | |
|---|---|---|---|
| | | 90), hormonal therapy (n = 20), or both (n = 20) | |
| Sample # | 19 | 295 | 307 |
| Blinding | Not stated | Rosetta Inpharmatics carried out microarray analysis; all raw data were available to all investigators | Data centralized at the TRANSBIG Secretariat; 100% concordance between risk classification by Agendia and Swiss Institute of Bioinformatics; statistical analysis carried out by IDDI[a] |
| Independence | Different specimens than used in discovery | Included 61 samples from discovery phase | Different specimens than used in discovery |
| | Not conducted by a separate group | Not conducted by a separate group | Involvement of the Swiss Institute of Bioinformatics, TRANSBIG[b] Secretariat, IDDI, and independent auditors |
| Results | Disease outcome was predicted by gene signature in 17 of 19 (Fisher's exact test for association, p = 0.0018) | Estimated HR for distant metastases in poor vs good signature groups: 5.1 (95% CI 2.9- 9.0; p < 0.001)<br><br>Multivariate model<br>• Poor prognosis signature of 70-gene test the strongest predictor of the likelihood of distant metastases (HR = 4.6; 95% CI 2.3- 9.2), with only tumor size and lack of adjuvant chemotherapy remaining in the model. | Distant metastasis:<br>• Unadjusted HR = 2.32 (95% CI 1.35- 4.00)<br>• Adjusted HR ranged from 2.13- 2.15 after adjustment for various estimates of clinical risk<br><br>Overall survival:<br>• Unadjusted HR = 2.79 (95% CI 1.60-4.87)<br>• Adjusted HR ranged from 2.63- 2.89 after adjustment for various estimates of clinical risk |

[a]International Drug Development Institute.
[b]A consortium of the Breast International Group (BIG).

**Expanding the eligibility age for MammaPrint** Investigators assessed MammaPrint performance in a retrospective study of 131 patients with node-negative breast cancer who were older than 55 years and not treated with adjuvant therapy (FDA, 2009a). The clinical sensitivity and specificity of the test in older women were comparable to previous

data submitted in support of using the test in younger women. The FDA expanded the intended use of the test to include breast cancer patients of all ages (FDA, 2009a). In an analysis of a predominantly postmenopausal cohort of 100 women, MammaPrint correctly identified 100 percent of women at low risk for distant metastases at 5 years (Wittner et al., 2008). However, the positive predictive value of MammaPrint (women who develop distant metastases who are classified as having a poor prognosis) was lower than previously observed (12 percent versus 52 percent) which, according to the investigators, was unexpected because older women are generally thought to have a lower risk of recurrence.

**Lymph node-positive patients** Mook et al. (2008) used a retrospective analysis with 241 patients to assess whether MammaPrint could identify patients with 1-3 positive lymph nodes who have a good prognosis. Investigators found that MammaPrint was significantly better than traditional prognostic factors in predicting breast cancer-specific survival with a multivariate HR = 7.17 (95% CI 1.81- 28.43; p = 0.005).

**Chemotherapy benefit** MammaPrint was assessed for its value in predicting chemotherapy benefit in a retrospective analysis evaluating 541 patient samples from a pool of 1,637 patients (Knauer et al., 2010).[15] Investigators concluded that patients with poor prognosis, defined by MammaPrint, derive a significant benefit from the addition of chemotherapy. The MINDACT trial (see below) will provide higher quality evidence to assess the benefit of chemotherapy because data will be collected prospectively within a large, randomized clinical trial population.

### Clinical Utility

The European Organisation for Research and Treatment of Cancer, with partial support from Agendia, is conducting a large multicenter, prospective randomized trial to compare MammaPrint with common clinicopathological criteria (using Adjuvant! Online) in selecting patients for adjuvant chemotherapy. The trial, "Microarray In Node negative and 1-3 positive lymph node Disease may Avoid ChemoTherapy" (MINDACT), will randomly assign chemotherapy to women who have discordant prognosis (either good prognosis with MammaPrint and high risk of recurrence based on clinicopathologic factors or poor prognosis with MammaPrint and low risk of recurrence based on clinicopathologic factors). All other women will be assigned to either (1) chemotherapy if they are at high risk for recurrence based on both their MammaPrint signature and clinicopathologic characteristics or (2) no chemotherapy if they are classified at low risk based on both. The trial has fully accrued more than 6,600 women who will be followed to assess the primary outcomes of distant metastasis-free survival and disease-free survival (Clinicaltrials.gov, 2011b). First results are expected in 2015.

[15] The pooled sample included patients from the original discovery phase study (van't Veer et al., 2002), clinical/biological validation studies (Buyse et al., 2006; van't Veer et al., 2002; van de Vijver et al., 2002), Mook et al. (2008), and several community-based registries (Knauer et al., 2010).

## Clinical Use

The MammaPrint test is conducted in Agendia's two CLIA-certified laboratories. Since its 2007 FDA clearance, MammaPrint has undergone several modifications to the test over time, which have been documented by the FDA (Table A-4) (Agendia, 2009). In the United States, MammaPrint is available for patients of all ages with invasive Stage I or II breast cancer, who are lymph-node negative, have tumors of less than 5 cm, and are either estrogen-receptor positive or negative. The intended use statement accompanying the FDA clearance specifies that "[t]he MammaPrint® result is indicated for use by physicians as a prognostic marker only, along with other clinicopathological factors" (FDA, 2011b, p.1). In addition, the special conditions for use statement notes that "MammaPrint® is not intended for diagnosis, or to predict or detect response to therapy, or to help select the optimal therapy for patients" (FDA, 2011b, p. 1). According to Agendia, insurance coverage for MammaPrint is available from the Centers for Medicare & Medicaid Services, private health insurers, and third party payers. More than 14,000 MammaPrint test results have been reported (Agendia, 2011b).

**TABLE A-4** FDA 510(k) Clearances for MammaPrint

| Decision Date | 510(k) Number | Summary |
|---|---|---|
| 02/06/2007 | K062694 | Original clearance for MammaPrint test<br>• Single laboratory<br>• Fresh frozen tissue<br>• Patients less than 61 years old<br>• Includes laboratory procedures and software computational models |
| 06/22/2007 | K070675 | Modifications<br>• Changed specimen type to fresh tissue stored in specific RNA preservative<br>• Change in software version |
| 07/21/2008 | K080252 | • Addition of a second scanner<br>• Replacement of low density microarray with high-density microarray |
| 12/11/2009 | K081092 | • Modified intended use by adding 5-year prognostic information for breast cancer patients 61 years and older |
| 01/28/2011 | K101454 | • Added 2 scanners, 2 bioanalyzers, and a new laboratory for MammaPrint testing |

SOURCE: FDA (2011a).

Guidelines from ASCO on the use of tumor markers specify that the clinical utility and appropriate application of MammaPrint is under investigation, and further state that: "MammaPrint profiling does appear to identify groups of patients with very good or very poor prognosis. However, due to the nature of the study design, it is difficult to tell if these data pertain to an inherently favorable outcome in untreated patients, to patients whose prognosis is favorable because of the therapy, or to those with poor outcomes in the absence of treatment or despite treatment" (Harris et al., 2007, p. 5301). NCCN is awaiting the results from the MINDACT trial before determining their recommendations for use (NCCN, 2011a). The Blue Cross Blue Shield Medical Advisory

Panel determined that MammaPrint did not meet the TEC criteria (BCBS, 2008). However, the 2009 update of the St. Gallen International Expert Consensus stated that "the Panel agreed that validated multigene tests, if readily available, could assist in deciding whether to add chemotherapy in cases where its use was uncertain after consideration of conventional markers" (Goldhirsch et al., 2009, p. 1324). According to Agendia, MammaPrint has also been included in the 2008 update to guidelines for the Dutch Institute for Healthcare Improvement (Agendia, 2011a).

### Case Highlights

Initial statistical approaches in MammaPrint test validation, including overlap between discovery and validation datasets and assessment of model performance using an incomplete cross-validation procedure that led to overfitting, were criticized in the literature (Ransohoff, 2003, 2004; Simon et al., 2003). A subsequent study (Buyse et al., 2006) provided a clinical validation with patients who had not participated in the discovery and validation studies, and confirmed MammaPrint as a prognostic test. However, the hazard ratios between the poor and good prognosis groups reported in Buyse et al. (2006) were lower than those reported in van de Vijver et al. (2002).

Discovery microarray and clinical data are available for MammaPrint, but the fully specified computational model is not published, and several details needed for independent replication are unclear from the supplemental materials. MammaPrint received FDA clearance and more than 14,000 MammaPrint results have been reported, but some technology assessment groups have asserted that more information on MammaPrint is needed to determine how it should be used in clinical practice (AHRQ, 2008; BCBS, 2008; Harris et al., 2007; NCCN, 2011a). It is hoped that the prospective validation trial, MINDACT, will provide this information, especially in determining if MammaPrint can accurately predict chemotherapy benefit.

## TISSUE OF ORIGIN

The Tissue of Origin test (Pathwork® Diagnostics) is a gene expression-based test designed to identify the primary tissue of origin for tumors that are difficult to classify, including metastatic, poorly differentiated, and undifferentiated tumors. Only about 20-25 percent of patients with tumors of unknown origin receive a primary tumor diagnosis, despite extensive clinical and pathological assessments and advanced imaging (Hillen, 2000; Pavlidis et al., 2003; Pavlidis and Merrouche, 2006). This is problematic because identifying the tissue of origin can have important ramifications for treatment decisions. The Tissue of Origin test aims to assist in the classification of such tumors by comparing the similarity of gene expression in a tumor with unknown origin to a panel of gene expression data from 15 common tumors[16] (Dumur et al., 2008).

The FDA has cleared two versions of the Tissue of Origin test—the test for frozen tissue in 2008 and the test for FFPE tissue in 2010. Pathwork Diagnostics consulted with the FDA on several occasions, including a pre-IDE meeting at which the FDA

[16] The tumor types include bladder, breast, colorectal, gastric, hepatocellular, kidney, melanoma, non-small cell lung, non-Hodgkin's lymphoma, ovarian, pancreatic, prostate, sarcoma, testicular germ cell, and thyroid. Pathwork Diagnostics is also in the process of developing gene expression panels for endometrial and head and neck cancers.

determined that an IDE was not required because the proposed study design involved an analysis of archived samples.[17]

## Discovery Phase

Published information on gene discovery and computational model development for the Tissue of Origin test is limited. The Tissue of Origin test for frozen and FFPE specimens include the same 15 tumor types, but they use different computational models and processing methods (Pillai et al., 2011). The test uses two computational models, one for standardization and one for classification. The standardization computational model was developed from analysis of more than 5,000 tissue samples. The classification computational model for FFPE specimens was developed from more than 2,000 frozen and 100 FFPE tissue specimens. Training specimens were assigned a tissue of origin diagnosis according to standard clinical and pathological practices, and the test set consisted of FFPE specimens. Investigators state that "[m]achine learning techniques guided selection of the 2,000-gene profile and the optimal model needed to classify the tumor" for FFPE tissue specimens (Pillai et al., 2011). In comparison, the optimal model for frozen tissue specimens included 1,550 genes in the computational model (FDA, 2008b). The Tissue of Origin test produces a set of 15 similarity scores that describe the probability that the gene expression of a tumor of unknown origin is comparable to gene expression of the 15 tumor types included in the test.

## Test Validation Phase

### *Analytical Validation*

Information on analytical validity is available for both versions of the Tissue of Origin test, for frozen tissue specimens (Dumur et al., 2008; FDA, 2008b) and FFPE specimens (FDA, 2010; Pillai et al., 2011). The standardization computational model used in both versions of the Tissue of Origin test corrects for variations in RNA quality, sample storage and preparation, operators, and microarray procedures (Pillai et al., 2011).

Dumur et al. (2008) evaluated the same 60 frozen tissue specimens of poorly differentiated and undifferentiated tumors from 15 tumor types with established origin in 4 different laboratories. Blinded microarray data were sent to Pathwork for generation of Tissue of Origin scores, which were then sent to pathologists blinded to the original tissue type diagnosis from the surgical reports for their use in generating an interpretation of tissue of origin based on predetermined cut-offs for the test. Investigators found that Tissue of Origin results were highly reproducible across all four laboratories (Dumur et al., 2008). One potential limitation of this analysis was that 57 of the 60 tissue specimens were obtained from the same biospecimen bank, and therefore it is possible that only some aspects of preanalytical variability were considered. The 2008 510(k) decision summary reports additional information on analytical performance for the Tissue of Origin test for frozen samples (FDA, 2008b).

The 2010 510(k) decision summary reports analytical performance information for the FFPE version of the Tissue of Origin Test, including assay precision and

[17] Personal communication, Ed Stevens, Pathwork Diagnostics, October 18, 2011.

reproducibility, quality controls, detection limits, analytical specificity, and assay cut-off values (FDA, 2010). Pillai et al. (2011) conducted a multisite reproducibility study that showed 89.3 percent concordance among three laboratories.

*Statistical and Bioinformatics Validation*

Information about the methods used to develop the computational model has not been made available in the peer-reviewed literature, and the proprietary computational model is not available.[18] Data used in computational model development includes publicly available information (GEO accession number 2109), commercial data sources, and private correspondence (Pillai et al., 2011).

The assay was locked down prior to the validation and reproducibility studies, and this information was documented in the design control procedures submitted to the FDA (Monzon et al., 2009; Pillai et al., 2011).[19]

*Clinical/Biological Validation*

Table A-5 lists the clinical validation studies for the Tissue of Origin test. The clinical validation for the Tissue of Origin test for frozen samples was a blinded, multicenter study that found overall sensitivity of 87.8 percent and specificity of 99.4 percent (Monzon et al., 2009). The site of origin (called the reference diagnosis) was known for all tissue specimens, but a limitation noted by the authors was the "inability to independently verify the reference diagnosis used to assess the accuracy of the test" (Monzon et al., 2009) because the reference diagnoses originated from the surgical pathology report that accompanied the banked specimen. Investigators also considered the possibility that an unknown primary tumor might originate from a tissue site that was not covered by the panel.

**TABLE A-5** Clinical/Biological Validation Studies for the Tissue of Origin Test

| Study | Monzon et al. (2009) | Pillai et al. (2011) |
|---|---|---|
| Tissue source(s) | 2 academic, 3 commercial biospecimen banks; electronic microarray files for 271 tumors obtained from the International Genomics Consortium | 7 tissue banks |
| Specimen preparation | Frozen | Formalin fixed paraffin embedded (FFPE) tissue |
| Study purpose | Determine the performance characteristics of the frozen Tissue of Origin test in specimens representative of those likely to be classified as uncertain primary cancers | Determine the performance characteristics of the FFPE tissue of Origin test in specimens representative of those likely to be classified as uncertain primary cancers, and to assess |

[18] Personal communication, Ed Stevens, Pathwork Diagnostics, October 18, 2011.
[19] Personal communication, Ed Stevens, Pathwork Diagnostics, October 18, 2011.

| | | |
|---|---|---|
| | | interlaboratory reproducibility |
| Study design | Retrospective | Retrospective |
| Tumor characteristics | No fewer than 25 specimens for each tumor type | 25- 57 specimens for each tumor type |
| | Approximately half of specimens were metastatic | |
| Sample # | 547 | 462 |
| Blinding | Yes | Yes |
| Independence | Different specimens than used in discovery | Different specimens than used in discovery |
| | Not conducted by a separate group | Not conducted by a separate group |
| Results | Overall sensitivity (positive percentage agreement with reference diagnosis) = 87.8% (95% CI 84.7%- 90.4%) | Overall sensitivity (positive percentage agreement with reference diagnosis) = 88.5% (95% CI 85.3%- 91.3%) |
| | Overall specificity (negative percentage agreement with reference diagnosis) = 99.4% (95% CI 98.3%- 99.9%) | Negative percentage agreement with reference diagnosis = 99.1% (no CI reported) |
| | Performance of metastatic tumors was slightly lower than poorly differentiated and undifferentiated tumors | |

The second clinical validation, for the FFPE version of the Tissue of Origin test, was a blinded study that used only specimens that were not used in discovery. Pillai et al. (2011) analyzed 462 specimens and found that overall agreement of the test result to reference diagnosis was 88.5 percent.

The Tissue of Origin test was also assessed in several small studies. In a study of 15 fresh frozen metastatic brain cancer specimens, a correct diagnosis was given in 12 out of 13 viable specimens (92.3 percent) (Wu et al., 2010). In another study of 21 fresh frozen tumor samples, Monzon et al. (2010) used the Tissue of Origin test to classify specimens from patients with a carcinoma of unknown primary (CUP). Investigators determined that the test yielded a clear identification of the primary site for 76 percent of specimens, but noted that the major limitation of the study was that CUP specimens, by definition, lack a reference diagnosis. Grenert et al. (2011) found that 35 of 37 viable FFPE specimens (95 percent) from an academic pathology department agreed with the reference diagnosis. The Tissue of Origin test was also assessed using FFPE cell blocks of cytologic body fluid specimens; investigators found that 16 of 17 viable samples, or

94.1 percent, were in agreement with reference diagnosis (Stancel et al., 2011). Dumur and colleagues (2011) assessed the Tissue of Origin test on 43 poorly differentiated and undifferentiated frozen tumor specimens, including 6 tumor samples that are not represented in the Tissue of Origin panel of 15 tumors (off-panel) and 7 CUP specimens. Investigators found 97 percent agreement between the Tissue of Origin test result and diagnosis, but noted that for CUP and off-panel specimens, the tissue type and cell type may be confounded by the Pathwork of Origin test.

**Clinical use**

Pathwork transitioned the Tissue of Origin Test for FFPE tissue, originally offered through its CLIA-certified laboratory as an LDT, to an FDA-cleared in vitro diagnostic (IVD) (Pathwork Diagnostics, 2010). Currently, all testing is performed in the Pathwork Diagnostics laboratory, but the company plans to make an IVD kit available for pathologists to run in their own clinical laboratories (Pathwork Diagnostics, 2011b).

The FDA intended use statement specifies that the Tissue of Origin test "measure[s] the degree of similarity between the RNA expression patterns in a patient's … tumor and the RNA expression patterns in a database of fifteen tumor types (poorly differentiated, undifferentiated and metastatic cases) that were diagnosed according to the current clinical and pathological practice. This test should be evaluated by a qualified physician in the context of the patient's clinical history and other diagnostic test results" (FDA, 2010, p. 1). The FDA limitation statement notes that Tissue of Origin is not intended to:

- Establish origin for tumors that cannot be diagnosed according to current practices;
- Subclassify or modify classification of tumors that can be diagnosed by current practice;
- Predict disease course, survival, or treatment efficacy, or distinguish between primary and metastatic tumors; and
- Distinguish tumors that are not in the test's database.

The Tissue of Origin test has not been incorporated yet into ASCO or NCCN guidelines; NCCN deemed the test an exciting new area of molecular profiling, but concluded that further clinical trials were still necessary before incorporation into its guidelines (Mulcahy, 2010). In a preliminary analysis of 59 patients, Hornberger et al. (2011) reported that oncologists changed treatment plans in 53 percent of patients with difficult to diagnose tumors after using the Tissue of Origin test. Another analysis of 284 consecutive cases found the Tissue of Origin test suggested a change in diagnosis in 81 percent (95% CI 76%- 85%) of cases and confirmed a suspected primary tumor site for 15 percent of cases (95% CI 12%- 20%) (Laouri et al., 2011). Studies have not yet evaluated whether clinical outcomes improve following an altered course of patient management based on the Tissue of Origin test.

In 2011, the company announced that the Tissue of Origin test received Medicare coverage and it is working to secure additional insurance coverage for the test (Pathwork Diagnostics, 2011a). As a private company, Pathwork Diagnostics does not release

information on the number of tests ordered per year, but stated that the company has processed results for thousands of Tissue of Origin Tests in the past year.[20]

### Case Highlights

There is limited publicly available information on Tissue of Origin test discovery, and the computational model is proprietary. Pathwork initially offered its Tissue of Origin test for FFPE as an LDT through its CLIA-certified laboratory, but in 2010 received FDA clearance for this version of the test. Clinical validation studies blinded Pathwork to reference diagnoses, specified lockdown, and involved tissue specimens from a number of biospecimen banks. The effect of the Tissue of Origin test on treatment decisions and patient outcomes is under evaluation.

## OVA1

OVA1® (Vermillion, Inc.) is a test that measures five proteins in serum (CA125-II, beta-2-microglobulin, transferrin, apolipoprotein A1, and transthyretin) to generate a score reflecting the likelihood of ovarian malignancy in patients with an adnexal mass for whom surgery is planned. The test is intended to help physicians determine which patients are more likely to have cancer and thus should be referred to a gynecologic oncologist for surgery.

OVA1 was FDA cleared in September 2009. Investigators met with the FDA several times, including before starting clinical trials, during clinical trials, and during the submission process.[21]

Initially, investigators intended to develop a screening test for ovarian cancer, but abandoned this goal because it would have required large studies, exceeding time and budgetary constraints, and necessitated a level of clinical specificity that would have been difficult to achieve (Fung, 2010). Indeed, AHRQ's evidence report on ovarian cancer detection noted that model simulations suggest that frequent screening for ovarian cancer, even with a highly specific test, would result in a very low positive predictive value because the test would identify a number of false positives for ovarian cancer (AHRQ, 2006). Investigators then sought to develop a diagnostic test to aid patient management decisions for women in whom an ovarian mass had already been identified.

### Discovery Phase

An early study in the discovery process for OVA1 aimed to identify serum biomarkers for the detection of early stage ovarian cancer (Zhang et al., 2004). Proteomic profiles were generated by mass spectrometry from 645 archived serum samples from healthy women and patients with ovarian cancer. The investigators identified three proteins (apolipoprotein A, transthyretin, and inter-α-trypsin inhibitor heavy chain 4) that, combined with CA125, appeared to provide a modest improvement over CA125 alone in identifying women with ovarian cancer. In subsequent studies, investigators identified seven candidate proteomic markers of ovarian cancer in addition to CA125: the three from the initial study plus four more, including beta-2-microglobulin and transferrin

[20] Personal communication, Ed Stevens, Pathwork Diagnostics, October 18, 2011.
[21] Personal Communication, Scott Henderson, Vermillion Inc., October 26, 2011.

(Fung, 2010; Zhang and Chan, 2010). According to Vermillion, this biomarker panel was further refined in a series of studies encompassing over 2,000 subjects, but detailed information describing the discovery and development process that led to the panel of 5 biomarkers comprising the OVA1 computational model have not been made available in the peer-reviewed literature.[22]

## Test Validation Phase

### *Analytical Validation*

The investigators determined that reproducibility on the mass spectrometry platform was not adequate for routine clinical use (Fung, 2010), so the test platform was changed to immunoassays, which were already available for several proteins in the panel. Data regarding the precision and reproducibility of the test were reported to the FDA (FDA, 2011c). Variability in results was measured across runs, over time, across serum lots, on different machines, and in different labs. Stability of specimens and reagents was also assessed under various conditions, such as storage temperature.

### *Statistics and Bioinformatics Validation*

The FDA 510(k) clearance states that the OVA1 computational model was derived using two independent training datasets. The first consisted of 284 preoperative serum samples from women with adnexal masses obtained from the University of Kentucky (175 benign disease and 109 malignancies). The second consisted of 125 evaluable specimens from a randomly selected subset of 146 preoperative serum samples that were set aside in the clinical validation trial (89 benign disease, 36 malignancies). This information has not been published in the peer-reviewed literature, but details of the training methodology were made available to the FDA.[23] The computational model is proprietary.

### *Clinical/Biological Validation*

The clinical validation study was a prospective, double-blind study involving 27 subject enrollment sites (Miller et al., 2011; Ueland et al., 2011). Enrollment was limited to women who had a documented pelvic mass following physical and clinical examination and planned surgical intervention; 743 patients were enrolled in the validation study, and 146 patient samples were randomly selected and set aside for the training set described above. Seventy-four specimens were eliminated due to missing information or an unevaluable sample, resulting in 524 evaluable patient samples of which 516 were evaluated by physician assessment. According to the FDA clearance decision summary, the OVA1 test was informative for both premenopausal and postmenopausal patients. Use of the test in conjunction with clinical presurgical assessment increased sensitivity for malignancy from 72 percent to 92 percent (Table A-6) (FDA, 2011c).

[22] Personal Communication, Scott Henderson, Vermillion Inc., October 26, 2011.
[23] Personal Communication, Scott Henderson, Vermillion Inc., October 26, 2011

**TABLE A-6** Performance Characteristics for OVA1 Applied to Pre- and Postmenopausal Subjects Evaluated by Non-Gynecologic Oncologist Physicians

| Performance Measure | Presurgical Clinical Assessment | OVA1 Test | Dual Assessment (Clinical Assessment and OVA1 Test) |
|---|---|---|---|
| Sensitivity | 72.2% (52/72) | 87.5% (63/72) | 91.7% (66/72) |
| Specificity | 82.7% (163/197) | 50.8% (100/197) | 41.6% (82/197) |
| Positive Predictive Value | 60.5% (52/86) | 39.4% (63/160) | 36.5% (66/181) |
| Negative Predictive Value | 89.1% (163/183) | 91.7% (100/109) | 93.2% (82/88) |
| Prevalence | | 26.8% (72/269) | |

SOURCE: FDA (2011c).

In a comparison of physician assessment with OVA1, the test correctly identified 70 percent of malignancies missed by physician assessment among gynecologists and gynecologic oncologists, and 95 percent of malignancies missed by physician assessment among gynecologic oncologists (Ueland et al., 2011).

### Clinical Use

The OVA1 test is available through Quest Diagnostics, which has exclusive rights to offer the test in the clinical reference laboratory market and is subject to CLIA certification (PR Newswire, 2009; Quest Diagnostics, 2011). As described in the FDA clearance decision summary (FDA, 2011c), the test consists of the OvaCalc Software as well as the instruments, assays, and reagents recommended by Vermillion, which are sold separately from the OvaCalc Software. For example, CA-125 levels are assessed with a Roche Elecsys 2010, while the remaining four proteins are detected by a Siemens BN II (FDA, 2011c). To determine the OVA1 score, a user manually enters the results for the five protein analytes into an Excel spreadsheet together with the headers from the OvaCalc Software to generate a numerical score from 0 to 10 from the 5 analyte results (with 10 indicating the highest probability of cancer). Dialogue with the FDA prompted the use of different OVA1 score cutoffs depending on menopausal status (5.0 for premenopausal women and 4.4 for post-menopausal women). The test is intended for use only as an adjunctive test to complement, not replace, other diagnostic and clinical procedures.

The American College of Obstetricians and Gynecologists and Society of Gynecologic Oncologists (SGO) issued a committee opinion in March 2011 that OVA1 "appears to improve the predictability of ovarian cancer in women with pelvic masses" but noted that the clinical utility of this test has not been established (ACOG and SGO, 2011). OVA1 has not yet been incorporated into guidelines from ASCO. NCCN guidelines note that the "SGO and [FDA] have stated that the OVA-1 test should not be used as a screening tool to detect ovarian cancer. The OVA-1 screening test uses 5 markers…to assess who should undergo surgery by an experienced gynecologic oncologist and who can have surgery in the community. Based on data documenting an

increased survival, the NCCN panel recommends that all patients should undergo surgery by an experienced gynecologic oncologist" (NCCN, 2011b).

According to Vermillion, the company is working to secure coverage and reimbursement for OVA1 (Vermillion, 2011). Currently, the test is covered under Medicare, 22 Blue Cross Blue Shield plans, and a number of other private U.S. health plans.

### Case Highlights

Few details of the discovery and development of the OVA1 test, including information on the computational model, are available in the peer-reviewed literature. The clinical/biological validation study is published (Miller et al., 2011; Ueland et al., 2011), and OVA1 is FDA-cleared. Input from FDA prompted the use of two different OVA1 cutoff values based on menopausal status.

## OVACHECK

Ovarian cancer is the leading cause of gynecologic cancer deaths in the United States, and its high mortality rate is frequently due to failure to detect the cancer at an early stage (ACS, 2011). In the late 1990s, Drs. Emanuel Petricoin and Lance Liotta, investigators from the FDA and the NCI, collaborated with researchers at the bioinformatics company Correlogic to develop a proteomics-based approach for ovarian cancer screening using serum samples. Investigators developed the OvaCheck Test using mass spectrometry analysis of circulating serum proteins to develop a statistical algorithm to discriminate between healthy patients and those with ovarian cancer (Diamandis, 2004).

The development of the OvaCheck test garnered public attention as well as early controversy (Pollack, 2004). The two laboratories licensed to perform the test, Quest Diagnostics and LabCorp, had planned to begin marketing the test in 2004 (Pollack, 2004). Before a validation study for OvaCheck had been published and prior to commercial availability, marketing materials for the test were distributed at an SGO conference (Wagner, 2004). Around the same time, the FDA sent Correlogic a letter specifying that the agency "has determined that the OvaCheck test is subject to regulation under the device provisions of the Federal Food, Drug, and Cosmetic Act."[24] In Correlogic's reply to the FDA, the company said that it "respectfully do[es] not agree … with FDA's position that the software used to provide the OvaCheck testing service is a medical device … subject to FDA premarket review" (Correlogic, 2004). Analyses of the data by independent statisticians, made possible by the willingness of the investigators to make data publicly available, uncovered problems with the data and methods. As a result, this unvalidated test was not made available to the public.

### Discovery and Test Development Process

In 2002, Petricoin, Liotta, and investigators at Correlogic published their findings in *The Lancet* (Petricoin et al., 2002). The investigators analyzed serum samples from

[24] FDA 2004. Letter Re: OvaCheck to Peter J. Levine from Steven I Gutman, July 12, 2004.

100 women with ovarian cancer, 16 with benign gynecologic conditions, and 100 healthy women using a hydrophobic interaction protein chip, analyzed by surface-enhanced laser desorption/ionization time-of-flight or SELDI-TOF mass spectrometry. The cases and controls were obtained as frozen samples from the National Ovarian Cancer Early Detection Program (NOCEDP); 17 additional controls were obtained at the Simone Protective Cancer Institute (SPCI) (Petricoin et al., 2002).

SELDI-TOF mass spectrometry was used to produce an m/z (or mass to charge ratio) for each ion, and the ion spectra generated were then analyzed computationally. Spectra generated by SELDI-TOF were analyzed by combining "genetic algorithms" (a set of features "survives" if it can discriminate affected cases from controls in a training set, and feature sets that cannot survive this test are discarded) with "cluster analysis" (cases and controls are used to form clusters so that an unknown sample can be classified by its similarity to a cluster set) (Petricoin et al., 2002). The cancer and normal spectra were split into sets of 50, with half used to train the computational model and half, along with the 16 benign spectra, used to validate the resulting computational model. None of the samples used in training were used to validate the test. The proprietary computational models were not described in the publication and the final form (i.e., the equations) of the computational model developed on the training data was not provided. In addition, little information was provided about data preprocessing, and it is unclear whether the model was locked down in the development process.

*Validation*

After initial components of the discovery and development process were called into question, two additional datasets were made publicly available by NCI and FDA through the Clinical Proteomics Program databank (NCI, 2002). The original dataset described in the initial paper (Petricoin et al., 2002) was derived with a Ciphergen H4 protein chip array, and was baseline corrected to allow for the removal of background "noise" or unnecessary peaks by running a blank set of samples that are subtracted from the data. The second set used the same samples as in dataset 1 and were also baseline corrected but were run on the Ciphergen WCX2 protein chip array. The third dataset contained new samples—91 controls and 162 with cancer—that were prepared robotically, rather than by hand, and were not baseline corrected. They were derived from the Ciphergen WCX2 protein chip array, as in dataset 2 (NCI, 2002).

Independent investigators began looking into mass spectrometry profiling of serum samples and the OvaCheck test because of its clinical importance, interest of the scientific community, and potential progression into clinical use.[25] Using these three publicly available datasets (NCI, 2002) they concluded that there were inadvertent changes in protocol mid-experiment (Baggerly et al., 2004). The m/z values reported suggested that no external calibration was supplied. Furthermore, when differing instrument calibration is employed, results may vary from lab to lab, rendering an assay that is inconsistent across settings (Baggerly et al., 2004). Standardized methods of calibration, or lack thereof, were not reported by Petricoin and Liotta. Baggerly concluded that the resulting proteomic patterns were attributable to "artifacts of sample

[25] Personal communication, Keith Baggerly, December 8, 2011.

processing, not the underlying biology of cancer" (Baggerly et al., 2004). This analysis also indicated that there were statistically significant differences in the noise regions of the cancer and normal spectra. These differences provided evidence that normal and cancer controls were not processed in an identical manner (Baggerly et al., 2004). Inconsistencies in the three datasets also implied that the clustering approach outlined by Petricoin and colleagues would not work (Baggerly et al., 2004). It is also possible that patient populations and sample handling differed between these 17 controls and the original case/control cohort. Ideally, the entire validation set would have been obtained independently at a separate institution (Diamandis, 2004).

In addition, the investigators found feature sets in the noise region of the spectra that were able to classify samples as cancer or controls, indicative of artifacts. They found that spectra that were not biologically relevant (i.e. noise spectra) differed between cases and controls, reflecting some non-biological differences between cases and controls that may have stemmed from differences in sample preparation, experimental protocol, etc. These batch effects should not have occurred if, as was claimed, the cases and controls were run together.

In another independent analysis using routine statistical methods, investigators again identified peaks in the noise spectra that distinguished between cancer and controls due to a significant non-biological source of bias in the data (Sorace and Zhan, 2003).

## Case Highlights

The OvaCheck development process provides lessons about the potential sources of bias that may result from improper experimental design. Variability in specimen quality or handling can impact the ability to train and validate a computational model; when specimens are "inherently different or are handled in a way that systematically introduces a signal into the data for one of the compared groups," bias can ensue (Ransohoff, 2005). Ensuring that samples are randomized, operators are blinded to the nature of samples, instruments are properly calibrated, and results are revalidated after every shift in protocol can help to ensure that the data are unbiased and that the resulting statistical predictor will generalize to independent data sets (Baggerly et al., 2004). Any protocol changes that occur during the experiment should be carefully documented to address potential sources of bias. These include a shift between chip types and routine maintenance of instruments mid-experiment that require recalibration of formulas, etc. The OvaCheck story also highlights the dangers of batch effects. Both the training and test sets suffered from batch effects that caused the noise region of the spectra to differ between the cases and controls. If the test samples had been obtained independently, preferably drawn from another institution at a different point in time, and prepared and run through the mass spectrometer separately, then the risk of such batch effects would have been greatly reduced. The study by Sorace and Zhan (2003) also demonstrated that use of a simpler statistical analysis may have illuminated mistakes early in the development process, possibly preventing the publication of the original *Lancet* paper.

Despite its failure, the OvaCheck development process serves as an example of the benefits of making data publicly available, which allowed for independent assessment of the data and computational model and thus prevented the clinical implementation of the test.

Problems in the development of ovarian proteomic-based tests are not unique to OvaCheck. In 2008, Yale University investigators (Visintin et al., 2008) reported the results for a six-biomarker combination (including CA125) for ovarian cancer detection. The test, OvaSure, became available on the market in June 2008 as an LDT offered by LabCorp (Pollack, 2008a). In August, FDA sent a notice to the company that the agency believed OvaSure had not received adequate clinical validation (FDA, 2008c), and then sent the company a warning letter (FDA, 2008d) that specified that the test required FDA oversight. LabCorp stopped sale of the OvaSure test, but disagreed with the FDA's position (Pollack, 2008b).

## ALLOMAP

AlloMap® Molecular Expression Testing (XDx Expression Diagnostics) was developed to aid identification of heart transplant recipients who have a low risk of moderate or severe acute cellular rejection[26] (ACR) at the time of testing. AlloMap is a blood test that measures RNA expression of 11 genes to obtain a single score on a scale of 0 to 40, with a lower score reflecting lower probability of ACR at the time of testing. Prior to AlloMap, the standard of care was a more invasive method for monitoring heart transplant patients for ACR. Endomyocardial biopsy (EMB) is a procedure that can cause rare, but potentially serious, complications (Baraldi-Junkins et al., 1993) and is also subject to inter-observer variability in histologic evaluation (Marboe et al., 2005; Nielsen et al., 1993).

XDx met with the FDA for a pre-IDE meeting to discuss the Cardiac Allograft Rejection Gene Expression Observational (CARGO) study, AlloMap development, and analytical and clinical validation procedures. At this time, it was determined that an IDE was not needed because the AlloMap test would not be directing patient management decisions in the CARGO study.[27] FDA informed XDx that AlloMap would be classified as a Class II device using the de novo 510(k) process.

### Discovery Phase

The data used to develop the gene expression profile for AlloMap were generated from patients enrolled in the CARGO study (Deng et al., 2006). After heart transplantation, all patients were followed prospectively with EMB at each subsequent clinical visit using standard techniques and grading by local pathologists according to International Society for Heart and Lung Transplantation (ISHLT) guidelines (Deng et al., 2006). At the same time, blood was drawn to isolate RNA from peripheral blood mononuclear cells. A subset of biopsies were graded by three independent pathologists blinded to the clinical information before selecting samples for test discovery, computational model development, and test validation.

A custom microarray representing 7,370 genes was used for the discovery phase of test development. Statistical analyses were used to select 97 candidate genes from the microarray expression data. A literature review identified an additional 155 genes based

[26] Acute cellular rejection (ACR) is when a transplanted organ is not accepted by the body of the organ recipient. In the AlloMap case study, ACR refers to a transplanted heart not being accepted by the body of the transplant recipient.

[27] Personal communication, Mitch Nelles, XDx, October 12, 2011.

on molecular pathways implicated in transplant rejection. The expression levels of the 252 candidate genes were further assessed by qRT-PCR to identify 68 genes whose expression correlated with moderate or severe rejection (as determined by EMB). Six genes were eliminated due to variation in expression based on blood sample processing time. From these remaining 62 genes, statistical modeling of gene expression correlations yielded a 20-gene classifier (11 informative genes, 9 control/normalization genes). The 11 informative genes, 6 derived from the literature and 5 from the microarray analysis, were used to calculate the AlloMap test score. Known functions of these genes include roles in hematopoiesis, platelet activation, T lymphocyte activation and migration, and response to steroids.

## Test Validation Phase

### *Analytical Validation*

Analytical validation for AlloMap was documented in the 510(k) decision summary (FDA, 2008a). XDx reported the results for the following variabilities: run-to-run, operator-to-operator (interoperator), within operator (intraoperator), lot-to-lot, plate-to-plate within a lot, and section-to-section within a plate. The range of RNA purity that is acceptable for testing was determined and reported. XDx also reported that test performance was not compromised by presence of the following in blood samples: immunosuppressants, cytomegalovirus, heparin, hemoglobin, acetylsalicylic acid, acetaminophen, triglyceride, bilirubin, and genomic DNA.

### *Statistics and Bioinformatics Validation*

Discovery-phase microarray data are available in a publicly accessible database (GEO, accession number GSE2445). Raw data from qRT-PCR training were provided to the FDA as part of premarket notification processes, but were not reported in Deng et al. (2006).

All samples for training and validation originated from the CARGO study. Samples used in the primary clinical validation study did not overlap with samples used in the discovery phase. The secondary clinical validation reused all 63 patient samples from the primary validation as well as some samples that had been used in the discovery phase (Figure A-1). Deng et al. (2006) noted that the secondary validation "may provide improved power but may be biased to the extent that a longitudinal set of samples from an individual patient are not completely independent with respect to gene expression."

Details of the computational model development were provided in the Deng et al. (2006) supplemental material and provided to the FDA as part of the 510(k) submission. The test was locked down after discovery, prior to final validation.[28]

[28] Personal communication, Mitch Nelles, XDx, October 12, 2011.

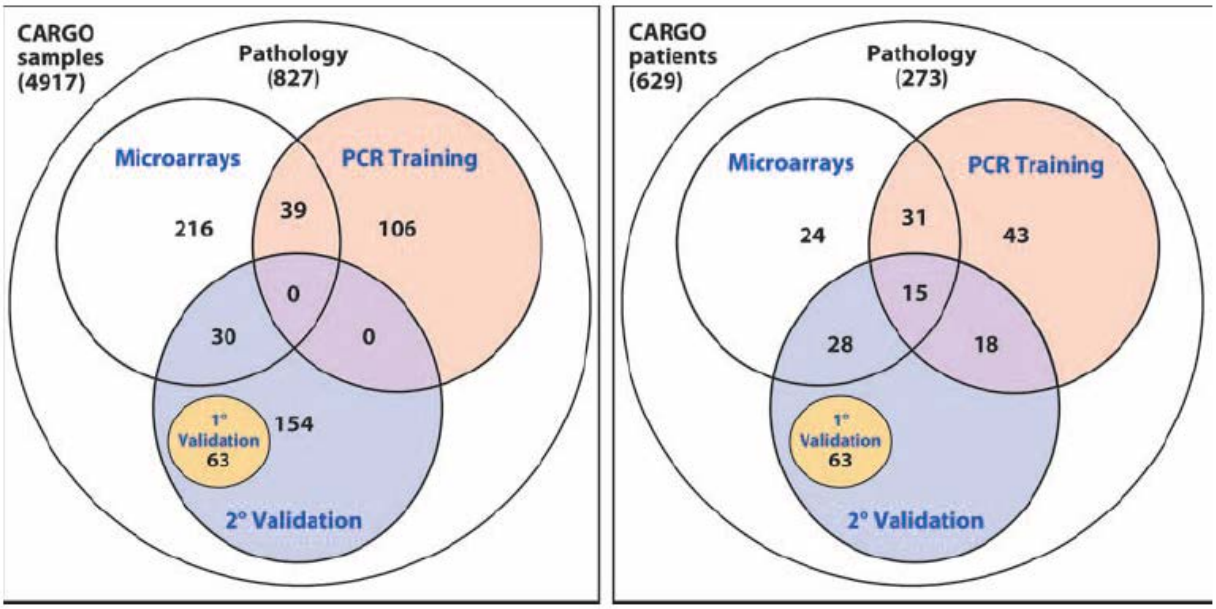


**FIGURE A-1** Venn diagrams illustrating overlap in patient blood samples used for AlloMap development. A total of 4,917 samples were drawn from 629 patients: 827 biopsy samples were analyzed by centralized pathology, 285 samples from 98 patients were analyzed by microarray and the data were used in the discovery phase, 145 samples from 107 patients were analyzed for PCR training, 63 samples from 63 patients were analyzed in the primary clinical validation (1° Validation), 63 patients from the primary clinical validation plus 61 patients analyzed in the microarray and PCR training steps of the discovery phase were analyzed in a secondary clinical validation (2° Validation).

NOTE: PCR = polymerase chain reaction.

SOURCE: Deng et al., (2006).

*Clinical/Biological Validation*

AlloMap was evaluated in three clinical validation studies and a prevalent population study (Table A-7) (Deng et al., 2006; FDA, 2008a). The histological characteristics of cardiac transplant rejection based on an expert panel reading of the individual EMB were used as the clinical endpoint. In the primary clinical validation study, samples from 63 patients were blinded and prospectively evaluated. Samples from these patients had not previously been introduced into any phase of test development. In this primary clinical validation, AlloMap distinguished between patients with moderate or severe rejection and those with no rejection ($p = 0.0018$). With the prospectively defined threshold score of 20 or greater as indicative of rejection, the investigators reported that the test correctly classified 84 percent (95% CI 66%- 94%) of patients with rejection and 38% (95% CI 22%- 56%) of patients with no rejection.

The secondary clinical validation study was performed to confirm the results in a larger sample set of CARGO patients. This clinical validation study evaluated 184 samples from 124 patients, which included the 63 samples used in the primary clinical validation study as well as samples that were used in the discovery phase of development

(see Figure A-1). In this secondary clinical validation, AlloMap correctly classified 76 percent (95% CI 63%- 85%) of patients with rejection and 41 percent (95% CI 32% - 50%) of patients with no rejection (p = 0.0001).

In the prevalent population study, investigators evaluated AlloMap's performance among patients likely to be seen in clinical practice. This analysis included 281 CARGO samples from 166 patients who were at least a year posttransplant. None of these samples had been used in AlloMap test discovery or training. At a threshold score of 30, which was selected to maximize the negative predictive value (NPV), the positive predictive value (PPV) for rejection was 6.8 percent, the NPV was 99.6 percent, and 68 percent of tests were below this value (Deng et al., 2006).

**TABLE A-7** Clinical/Biological Validation Studies for AlloMap

| Study | CARGO primary clinical validation study (Deng et al., 2006) | CARGO secondary clinical validation study (Deng et al., 2006) | CARGO prevalent population study (Deng et al., 2006) | CARGO dataset used for FDA clearance (k073482) |
|---|---|---|---|---|
| Tissue source | CARGO trial | CARGO trial | CARGO trial | CARGO trial |
| Test platform | qRT-PCR and EMB | qRT-PCR and EMB | qRT-PCR and EMB | qRT-PCR and EMB |
| Study question | Does AlloMap distinguish rejection from no rejection? | Does AlloMap distinguish rejection from no rejection? | Does AlloMap distinguish rejection from no rejection in samples representing expected clinical population? | Does AlloMap distinguish rejection from no rejection in samples representing expected population? |
| Study design | Prospective<br>Not marker directed | Prospective<br>Not marker directed | Prospective<br>Not marker directed | Prospective<br>Not marker directed |
| Patient characteristics | Time posttransplant not stated | Time posttransplant not stated | ≥1 year posttransplant | > 55 days posttransplant, >30 days post rejection treatment |
| Sample number | 63 samples; 63 patients | 184 samples; 124 patients | 281 samples; 166 patients | 300 samples; 154 patients |
| Blinding | Yes | Yes | Yes | Yes |
| Independence | Different specimens than used in discovery<br>Development involved XDx and transplant | All specimens used in primary clinical validation plus some specimens used in discovery | Different specimens than used in discovery<br>Development involved XDx and transplant cardiologists | Different specimens than used in discovery<br>Development involved XDx and transplant cardiologists |

| | | | | |
|---|---|---|---|---|
| | cardiologists directing CARGO[a] | Development involved XDx and transplant cardiologists directing CARGO | directing CARGO | directing CARGO |
| Results | At threshold score of 20, test correctly classified 84% (95% CI 66%-94%) of patients with rejection and 38% (95% CI 22%-56%) of patients with no rejection (p = 0.0018) | At threshold score of 20, test correctly classified 76% (95% CI 63%-85%) of patients with rejection and 41% (95% CI 32%-50%) of patients with no rejection (p = 0.0001) | At threshold score of 30, PPV for rejection was 6.8%, NPV was 99.6% (Deng et al., 2006) | AUC of 0.67 (95% CI 0.56-0.78). At threshold score[b] of 34 at 2-6 months post transplant PPV for rejection = 5.0%, NPV = 98.2%. At threshold score of 34 at >6 months post transplant PPV for rejection = 4.1%, NPV = 98.9% |

[a] Personal communication, Mitch Nelles, XDx, December 9, 2011.

[b] Starling et al., 2006.

For FDA clearance (FDA, 2008a), an analysis of the prevalent population study was evaluated. The analysis included 300 CARGO samples from 154 patients who were at least 55 days post transplant and more than 30 days beyond a rejection episode. None of these samples had been used in the discovery or training phases of AlloMap product development. The area under the curve (AUC) was 0.67 (95 percent CI 0.56-0.78) for a receiver operator characteristic develop from the full dataset.

## Clinical Utility

*Monitoring risk of future ACR*

In a prospective evaluation of 104 CARGO patients who were at least 30 days past a heart transplantation and whose blood samples had not been used to develop the AlloMap computational model, results suggest that the gene expression score used to determine rejection at the time of testing may determine the likelihood of ACR in the subsequent 12 weeks (Mehra et al., 2007, 2008). The score (mean ± standard deviation) for patients with rejection within the following 12 weeks was 27.4 ± 6.3 (n = 39) and for patients with no rejection in the following 12 weeks was 23.9 ± 7.1 (p = 0.01). During this time, no samples from patients with rejection within the following 12 weeks had scores of less than 20 (Mehra et al., 2007).

*International population*

The CARGO II trial was designed to evaluate the correlation between AlloMap and presence or absence of ACR as determined by EMB in a mostly European cohort of

heart transplant patients. The study was completed in February 2009, but the results are not yet available (ClinicalTrials.gov, 2011a). Of the 17 transplant centers participating in the study, 4 were in North America and 13 were in Europe.

*Noninferiority of AlloMap to EMB for Clinical Management of Heart Transplant Patients*

The IMAGE (Invasive Monitoring Attenuation through Gene Expression) trial demonstrated that AlloMap was non-inferior to EMB for monitoring posttransplant patients for ACR and reduced the number of biopsies that needed to be performed on heart transplant patients (Pham et al., 2010). Patients were randomly assigned to monitoring for rejection by either AlloMap testing or EMB. The two groups were compared with respect to a composite primary outcome of rejection with hemodynamic compromise, graft dysfunction due to other causes, death, or retransplantation. The 2-year cumulative rates of the primary outcome were 14.5 percent for patients monitored with AlloMap and 15.3 percent for patients monitored with EMB ($p = 0.86$). The rates of death from any cause were 6.3 percent and 5.5 percent respectively ($p = 0.82$). The frequency of biopsy per patient year of follow-up was 0.5 and 3.0, respectively ($p < 0.001$).

## Clinical Use

The FDA 510(k) decision summary (2008a) stated that the intended clinical use of AlloMap testing is "to aid in the identification of heart transplant recipients with stable allograft function who have a low probability of moderate/severe acute cellular rejection (ACR) at the time of testing in conjunction with standard clinical assessment." AlloMap testing is performed at the XDx CLIA-certified laboratory. Since AlloMap was initially marketed in 2005, more than 32,000 commercial tests have been performed in U.S. heart transplant patients. In 2010, 7,147 tests were performed.[29]

ISHLT guidelines recommend that AlloMap can be used in low-risk patients who are 6 months to 5 years post heart transplantation for ruling out presence of ACR (ISHLT, 2010). The California Technology Assessment Forum determined that AlloMap "meets Technology Assessment Criteri[a] 1 through 5 for safety, effectiveness and improvement in health outcomes when used to manage heart transplant patients at least one year post-transplant" (CTAF, 2010, p. 15). The Blue Cross Blue Shield Association Technology Evaluation Center decided in September 2011 that AlloMap did not meet its utility criteria as a method to monitor cardiac allograft rejection because the clinical validation studies were small and the cut points defining a positive test had not been independently validated (BCBSA, in press).

Currently, those who provide coverage and/or regular payment for the AlloMap test include Medicare, MediCal, New York Medicaid, United Healthcare, Anthem Wellpoint, Aetna, Kaiser, and several other private insurers.[30]

[29] Personal communication, Mitch Nelles, XDx, October 12, 2011.
[30] Personal communication, Mitch Nelles, XDx, December 9, 2011.

### Case Highlights

In the CARGO trial, there was no overlap between patient samples used for discovery and those used for the primary clinical validation study, but there was overlap between patient samples used for discovery and those used in a secondary clinical validation study. XDx indicated that overlap could not be avoided due to the difficulty in obtaining the requisite number of blood samples from heart transplant recipients associated with a consensus biopsy reading of ACR.[31]

XDx met with the FDA early in the process to determine the appropriate pathway for developing AlloMap, and the test was cleared in 2008. The test was locked down prior to final validation. Some additional performance characterization has been conducted with the goal of making better use of the output from the test, but no changes have been made to the computational model that generates the AlloMap scores.[32]

## CORUS CAD

Corus® CAD (CardioDx, Inc.) was developed as a less invasive method than angiography to identify obstructive coronary artery disease (CAD).[33] Corus CAD is a blood test that measures the expression level of 23 genes to get a score on a scale of 1 to 40. The score is used by primary care clinicians and cardiologists for determining whether a non-diabetic patient's symptoms of cardiovascular disease are due to CAD.

CardioDx met with the FDA for a pre-IDE meeting prior to derivation of the final computational model for Corus CAD,[34] but this test has not been submitted to the FDA for clearance or approval. CardioDx sought the LDT pathway to market and noted that at that time, the "FDA had draft guidance for IVDMIAs and did not require 510(k) clearance or PMA."[35]

### Discovery Phase

Initial proof-of-concept work in two retrospective cohorts undergoing coronary angiography demonstrated a set of genes that differentiated between patients with obstructive CAD and those without (Wingrove et al., 2008). The discovery process to develop Corus CAD entailed two microarray gene expression analyses (Rosenberg et al., 2010; Elashoff et al., 2011). Gene expression was measured in whole-blood cells, which were collected from patients prior to coronary angiography. The first microarray gene expression analysis was a retrospective study of samples from the repository of the Duke University CATHGEN registry,[36] which contained blood samples from patients with and without diabetes (Elashoff et al., 2011). Microarray analysis of 195 patient samples suggested 2,438 CAD-associated genes. Eighty-eight genes that had the greatest statistical significance and biological relevance were selected for confirmation by RT-PCR in these same 195 samples. Diabetes was the clinical factor that had the most

[31] Personal communication, Mitch Nelles, XDx, October 21, 2011.
[32] Personal communication, Mitch Nelles, XDx, October 12, 2011.
[33] CAD is the damage to the heart caused by atherosclerotic constriction of arteries supplying blood to the heart.
[34] Personal communication, Steve Rosenberg, Cardio Dx, October 21, 2011.
[35] Personal communication, Steve Rosenberg, Cardio Dx, October 21, 2011.
[36] See http://cathgen.duhs.duke.edu/modules/cath_about/index.php?id=1 (accessed January 18, 2012).

significant effect on gene expression ($p = 0.0006$). Analysis in non-diabetic and diabetic subsets ($n = 124$ and 71, respectively) showed expression of 42 and 12 significant CAD genes, respectively ($p < 0.05$), with no intersection (Elashoff et al., 2011). Therefore, the authors limited further work to patients without diabetes. A second microarray analysis to further define genes that could be a hallmark of CAD, and all subsequent work to develop Corus CAD, was performed in the prospective clinical trial PREDICT (Personalized Risk Evaluation and Diagnosis in the Coronary Tree). The development of the classifier entailed a three-step approach. Patients from 39 U.S. centers were assigned to a group based on date of study enrollment. Blood samples from 198 PREDICT patients were used for the second microarray gene expression analysis. The analysis suggested 5,935 CAD-associated genes. There were 655 genes that overlapped with the CATHGEN microarray results.

A total of 113 genes was selected based on biological relevance, statistical significance, and CAD-associated gene expression measured with RT-PCR in 640 PREDICT patient samples. Gene expression correlation clustering and cell-type analyses of these genes were used to determine the final 23-gene computational model (20 CAD-related genes and 3 normalization genes). These genes have functions in neutrophil activation and apoptosis, natural killer cell activation, innate immunity, cell necrosis, and adaptive immune response (Rosenberg et al., 2010).

## Test Validation Phase

### *Analytical Validation*

Detailed information regarding analytical validation for Corus CAD is not publicly available. The following preanalytical clinical and demographic variables were measured in patients used in the microarray studies: sex, age, race, body mass index, current smoker, systolic blood pressure, diastolic blood pressure, hypertension, dyslipidemia, neutrophil count, and lymphocyte count (Elashoff et al., 2011).

### *Statistics and Bioinformatics Validation*

Discovery-phase microarray data are available in a publicly accessible database (GEO, accession number GSE20686). PCR data in the development and validation sets are not publicly available, although Cardio Dx has indicated they would be available upon request by qualified investigators.[37] Analysis of the RT-PCR results from the validation study was performed at Scripps Translational Science Institute.[38]

Blood samples from 640 patients were used to develop the computational model (Elashoff et al., 2011); blood samples from another 526 patients were used for validation. There was no overlap in patient samples used in each phase of the study.

Details regarding development of the computational model were published (Rosenberg et al., 2010). The model was locked down prior to the start of the validation study. CardioDx funded the study and was involved in the design and conduct of the study.

[37] Personal communication, Steve Rosenberg, Cardio DX, October 21, 2011, December 12, 2011.
[38] Personal communication, Steve Rosenberg, Cardio DX, October 21, 2011.

*Clinical/Biological Validation*

Rosenberg et al. (2010) reported statistical analyses only for the validation group. The investigators predefined the primary endpoint as the receiver-operating characteristic (ROC) curve area for prediction of disease status by the test score. In a set of 526 PREDICT patients not used for gene discovery or computational model development, the area under the curve (AUC) for ROC was $0.70 \pm 0.02$ ($p < 0.001$). At a threshold score of 14.75, which corresponded to a 20 percent likelihood of obstructive CAD, the sensitivity and specificity were 85 percent and 43 percent respectively. This yielded an NPV of 83 percent and a PPV of 46 percent, with 33 percent of patient scores below this threshold (Rosenberg et al., 2010).

## Clinical Utility

The COMPASS (Coronary Obstruction Detection by Molecular Personalized Gene Expression) trial prospectively evaluated use of Corus CAD in patients referred for myocardial perfusion imaging (MPI) due to suspected CAD. 431 patients were analyzed. In the primary analysis of 63 cases, Corus CAD AUC was 0.79 ($p < 0.001$), while MPI AUC was 0.59 ($p < 0.001$). At a threshold score of 15, Corus CAD sensitivity and NPV were 89 percent and 96 percent, respectively; MPI sensitivity and NPV were 27 percent and 88 percent, respectively. In the secondary case analysis, Corus CAD AUC was also higher than MPI (0.77 vs 0.64, $p < 0.01$) (Thomas et al., 2011). Estimated study completion date is March 2012 (ClinicalTrials.gov, 2011a).

## Clinical Use

As an LDT, Corus CAD is performed in the company's CLIA-certified laboratory. Approximately 13,000 tests were ordered between October 2010 and September 2011.[39]

Corus CAD is a very newly developed test, which may account for why it has not been incorporated into any guidelines. Multiple insurance plans currently pay for the test on a patient-by-patient basis.[40] As noted on the CardioDx website, "CardioDx is actively pursuing third-party payer reimbursement for Corus CAD."

## Case Highlights

This study highlights the importance of publishing detailed information regarding derivation of the computational model used for the diagnostic test. The work by Rosenberg and colleagues has been described as having an "elegant design" and being "at the vanguard of clinical genetics in cardiovascular care" but "the report offers too little information about the derivation of the algorithm for readers to determine whether the screening tool provides internally valid results" (Arnett, 2010, p. 473). Six months after the publication of the clinical work by Rosenberg et al. (2010), Elashoff et al. (2011) published more detailed information regarding the derivation of the computational model.

[39] Personal communication, Steve Rosenberg, Cardio DX, October 24, 2011.
[40] Personal communication, Steve Rosenberg, Cardio DX, December 12, 2011.

## REFERENCES

Ach, R. A., A. Floore, B. Curry, V. Lazar, A. M. Glas, R. Pover, A. Tsalenko, H. Ripoche, F. Cardoso, M. S. d'Assignies, L. Bruhn, and L. J. van't Veer. 2007. Robust interlaboratory reproducibility of a gene expression signature measurement consistent with the needs of a new generation of diagnostic tools. *BMC Genomics* 8(148):10.1186/1471-2164-8-148.

ACOG and SGO (American College of Obstetricians and Gynecologists and Society of Gynecologic Oncologists). 2011. Committee opinion no. 477: The role of the obstetrician-gynecologist in the early detection of epithelial ovarian cancer. *Obstetrics and Gynecology* 117(3):742-746.

ACS (American Cancer Society). 2011. *What are the key statistics about ovarian cancer?* http://www.cancer.org/Cancer/OvarianCancer/DetailedGuide/ovarian-cancer-key-statistics. (accessed September 8, 2011).

Agendia. 2009. *FDA broadens clearance for Agendia's MammaPrint*. http://www.agendia.com/pages/press_release/70.php?aid=90 (accessed March 16, 2011).

Agendia. 2011a. *International recognition for pioneering work in translation research and personalized medicine for breast cancer*. http://www.agendia.com/pages/awards_and_recognition/97.php (accessed September 21, 2011).

Agendia. 2011b. *MammaPrint has extensive international clinical validation*. http://www.agendia.com/pages/validation/32.php (accessed March 27, 2011).

AHRQ (Agency for Healthcare Research and Quality). 2006. *Genomic tests for ovarian cancer detection and management*. Rockville, MD: AHRQ.

AHRQ. 2008. *Impact of gene expression profiling tests on breast cancer outcomes*. Rockville, MD: AHRQ.

Albain, K. S., W. E. Barlow, S. Shak, G. N. Hortobagyi, R. B. Livingston, I.-T. Yeh, P. Ravdin, R. Bugarini, F. L. Baehner, N. E. Davidson, G. W. Sledge, E. P. Winer, C. Hudis, J. N. Ingle, E. A. Perez, K. I. Pritchard, L. sheperd, J. R. Gralow, C. Yoshizawa, D. C. Allred, C. K. Osborne, and D. F. Hayes. 2010. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: A retrospective analysis of a randomized trial. *The Lancet Oncology* 11(1):55-65.

Arnett, D. K. 2010. Gene expression algorithm for prevalent coronary artery disease: A first step in a long journey. *Annals of Internal Medicine* 153(7):473-474.

Baehner, F. L., N. Achacoso, T. Maddala, S. Shak, C. P. Quesenberry, Jr., L. C. Goldstein, A. M. Gown, and L. A. Habel. 2010. Human epidermal growth factor receptor 2 assessment in a case-control study: Comparison of fluorescence in situ hybridization and quantitative reverse transcription polymerase chain reaction performed by central laboratories. *Journal of Clinical Oncology* 28(28):4300-4306.

Baggerly, K. A., J. S. Morris, and K. R. Coombes. 2004. Reproducibility of SELDI-TOF protein patterns in serum: Comparing datasets from different experiments. *Bioinformatics* 20(5):777-785.

Baraldi-Junkins, C., H. R. Levin, E. K. Kasper, B. K. Rayburn, A. Herskowitz, and K. L. Baughman. 1993. Complications of endomyocardial biopsy in heart transplant patients. *Journal of Heart and Lung Transplantation* 12(1 Pt 1):63-67.

BCBS (Blue Cross and Blue Shield Association). 2008. Gene expression profiling of breast cancer to select women for adjuvant chemotherapy. *Technology Evaluation Center* 22(13):1-51.

BCBS. In press. Gene expression profiling as a noninvasive method to monitor for cardiac allograft rejection. *Technology Evaluation Center.*

Berry, D. A., C. Cirrincione, I. C. Henderson, M. L. Citron, D. R. Budman, L. J. Goldstein, S. Martino, E. A. Perez, H. B. Muss, L. Norton, C. Hudis, and E. P. Winer. 2005. Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. *Journal of the American Medical Association* 295(14):1658-1667.

Buyse, M., S. Loi, L. J. van't Veer, G. Viale, M. Delorenzi, A. M. Glas, M. S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. Harris, J. Bogaerts, P. Therasse, A. Floore, M. Amakrane, F. Piette, E. T. Rutgers, C. Sortiriou, F. Cardoso, and M. J. Piccart. 2006. Validation and clinical utility of a 70-

gene prognostic signature for women with node-negative breast cancer. *Journal of the National Cancer Institute* 98(17):1183-1192.

ClinicalTrials.gov. 2011a. *Cardiac Allograft Rejection Gene Expression Observational (CARGO) II STUDY(CARGOII*). http://www.clinicaltrials.gov/ct2/show/NCT00761787?term=CARGO&rank=1 (accessed November 15, 2011).

Clinicaltrials.gov. 2011b. *Genetic testing or clinical assessment in determining the need for chemotherapy in women with breast cancer that involves no more than 3 lymph nodes*. http://clinicaltrials.gov/ct2/show/NCT00433589?term=mindact&rank=1 (accessed March 27, 2011).

Cobleigh, M. A., B. Tabesh, P. Bitterman, J. Baker, M. Conin, M. L. Liu, R. Borchik, J. M. Mosquera, M. G. Walker, and S. Shak. 2005. Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clinical Cancer Research* 11(24 Pt 1):8623-8631.

Correlogic. 2004. Re: Correlogic Systems Inc. Reference Laboratory – OvaCheck Testing Service. http://www.correlogic.com/pdfs/July14SteveGutmanLetter.pdf (accessed December 2, 2011).

Cronin, M., M. Pho, D. Dutta, J. C. Stephans, S. Shak, M. C. Kiefer, J. M. Esteban, and J. Baker. 2004. Measurement of gene expression in archival paraffin-embedded tissues: Development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *American Journal of Pathology* 164(1):35-42.

Cronin, M., C. Sangli, M.-L. Liu, M. Pho, D. Dutta, A. Nguyen, J. Jeong, J. Wu, K. C. Langone, and D. Watson. 2007. Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. *Clinical Chemistry* 53(6):1084-1091.

CTAF (California Technology Assessment Forum). 2010. *Gene Expression Profiling for the Diagnosis of Heart Transplant Rejection.* http://ctaf.org/content/assessment/detail/1208 (accessed January23, 2012).

De, P., B. R. Smith, and B. Leyland-Jones. 2010. Human epidermal growth factor receptor 2 testing: Where are we? *Journal of Clinical Oncology* 28(28):4289-4292.

Deng, M. C., H. J. Eisen, M. R. Mehra, M. Billingham, C. C. Marboe, G. Berry, J. Kobashigawa, F. L. Johnson, R. C. Starling, S. Murali, D. F. Pauly, H. Baron, J. G. Wohlgemuth, R. N. Woodward, T. M. Klingler, D. Walther, P. G. Lal, S. Rosenberg, and S. Hunt. 2006. Noninvasive Discrimination of Rejection in Cardiac Allograft Recipients Using Gene Expression Profiling. *American Journal of Transplantation* 6(1):150-160.

Diamandis, E. 2004. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: Opportunities and potential limitations. *Molecular and Cellular Proteomics* 3(4):367-378.

Dowsett, M., J. Cuzick, C. Wale, J. Forbes, E. A. Mallon, J. Salter, E. Quinn, A. Dunbier, M. Baum, A. Buzdar, A. Howell, R. Bugarini, F. L. Baehner, and S. Shak. 2010. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: A TransATAC study. *Journal of Clinical Oncology* 28(11):1829-1834.

Dumur, C. I., M. Lyons-Weiler, C. Sciulli, C. T. Garrett, I. Schrijver, T. K. Holley, J. Rodriguez-Paris, J. R. Pollack, J. L. Zehnder, M. Price, J. M. Hagenkord, C. T. Rigl, L. J. Buturovic, G. G. Anderson, and F. A. Monzon. 2008. Interlaboratory performance of a microarray-based gene expression test to determine tissue of origin in poorly differentiated and undifferentiated cancers. *Journal of Molecular Diagnostics* 10(1):67-77.

Dumur, C. I., C. E. Fuller, T. L. Blevins, J. C. Schaum, D. S. Wilkinson, C. T. Garrett, C. N. Powers. 2011. Clinical verification of the performance of the Pathwork Tissue of Origin test. *American Journal of Clinical Pathology* 136(6):924-933.

EBCTCG (Early Breast Cancer Trialists Collaborative Group). 2005. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. *The Lancet* 365(9472):1687-16717.

Elashoff, M. R., J. A. Wingrove, P. Beineke, S. E. Daniels, W. G. Tingley, S. Rosenberg, S. Voros, W. E. Kraus, G. S. Ginsburg, R. S. Schwartz, S. G. Ellis, N. Tahirkheli, R. Waksman, J. McPherson, A. J. Lansky, and E. J. Topol. 2011. Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in non-diabetic patients. *BMC Medical Genomics* 4(1):26.

Esteban, J., J. Baker, M. Cronin, M. L. Liu, M. G. Llamas, M. G. Walker, R. Mena, and S. Shak. 2003. Tumor gene expression and prognosis in breast cancer: Multi-gene RT-PCR assay of paraffin-embedded tissue. *Proceedings of the American Society of Clinical Oncology* 22:Abstract 3416.

Esteva, F. J., A. A. Sahin, M. Cristofanilli, K. Coombes, S.-J. Lee, J. Baker, M. Cronin, M. Walker, D. Watson, S. Shak, and G. N. Hortobagyi. 2005. Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. *Clinical Cancer Research* 11(9):3315-3319.

FDA (Food and Drug Administration). 2007a. *510(k) Substantial Equivalence Determination Decision Summary (k070675)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K070675.pdf (accessed September 20, 2011).

FDA. 2007b. *FDA clears breast cancer specific molecular prognostic test*. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2007/ucm108836.htm (accessed March 14, 2011).

FDA. 2008a. *510(k) substantial equivalence determination decision summary assay and instrument combination template* (k073482). http://www.accessdata.fda.gov/cdrh_docs/reviews/K073482.pdf (accessed November 23, 2011).

FDA. 2008b. *510(k) substantial equivalence determination decision summary (K080896)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K080896.pdf (accessed November 15, 2011).

FDA. 2008c. OvaSure manufacturer letter. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/IVDRegulatoryAssistance/ucm125130.htm (accessed November 23, 2011).

FDA. 2008d. Laboratory Corporation of America 29-Sep-08. http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2008/ucm1048114.htm (accessed November 23, 2011).

FDA. 2009a. *510(k) Substantial Equivalence Determination Decision Summary (k081092)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K081092.pdf (accessed September 20, 2011).

FDA. 2009b. *InSite™ Her-2/neu kit—P040030.* http://www.fda.gov/medicaldevices/productsandmedicalprocedures/deviceapprovalsandclearances/recently-approveddevices/ucm079253.htm (accessed November 28, 2011).

FDA. 2009c. *Invitrogen SPOT-Light® HER2 CISH Kit.* http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm074029.htm (accessed November 28, 2011).

FDA. 2010. *510(k) Substantial Equivalence Determination Decision Summary (k092967)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K092967.pdf (accessed November 16, 2011).

FDA. 2011a. *510(k) Premarket Notification CDRH SuperSearch*. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm (accessed October 18, 2011).

FDA. 2011b. *510(k) Substantial Equivalence Determination Decision Summary (k101454)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K101454.pdf (accessed September 19, 2011).

FDA. 2011c. *Substantial equivalence determination decision summary (k081754)*. http://www.accessdata.fda.gov/cdrh_docs/reviews/K081754.pdf (accessed October 11, 2011).

Fisher, B., J. Costantino, C. Redmond, R. Poisson, D. Bowman, J. Couture, N. V. Dimitrov, N. Wolmark, D. L. Wickerham, and E. R. Fisher. 1989. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *New England Journal of Medicine* 23(320):479-484.

Fisher, B., J. Dignam, N. Wolmark, A. DeCillis, B. Emir, D. L. Wickerham, J. Bryant, N. V. Dimitrov, N. Abramson, J. N. Atkins, H. Shibata, L. Deschenes, and R. G. Margolese. 1997. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *Journal of the National Cancer Institute* 89(22):1673-1682.

Fisher, B., J. Jeong, J. Bryant, S. Anderson, J. Dignam, E. R. Fisher, and N. Wolmark. 2004. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *The Lancet* 364(9437):858-868.

Fung, E. T. 2010. A recipe for proteomics diagnostic test development: The OVA1 test, from biomarker discovery to FDA clearance. *Clinical Chemistry* 56(2):327-329.

Genomic Health. 2011a. *Oncotype DX Breast Cancer Assay*. http://www.oncotypedx.com/en-US/Breast/HealthcareProfessional/Overview.aspx (accessed January 20, 2011).

Genomic Health. 2011b. *Oncotype DX Breast Cancer Assay: Insurance Information*. http://www.oncotypedx.com/en-US/Breast/HealthcareProfessional/InsuranceInformation.aspx (accessed January 31, 2011).

Genomic Health. 2011c. *Presentation of the Development and Clinical Validation of Oncotype DX*. http://www.oncotypedx.com/en-US/Breast/HealthcareProfessional/Overview.aspx (accessed July 21, 2011).

Glas, A. M., A. Floore, L. Delahaye, A. T. Wittereveen, R. C. F. Pover, N. Bakx, J. S. T. Lahti-Domenici, T. J. Bruinsma, M. O. Warmoes, R. Bernards, L. F. A. Wessels, and L. J. van't Veer. 2006. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 7(278).

Goldhirsch, A., J. N. Ingle, R. D. Gelber, A. S. Coates, B. Thurlimann, and H.-J. Senn. 2009. Thresholds for therapies: Highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Annals of Oncology* 20(8):1319-1329.

Grenert, J. P., A. Smith, W. Ruan, R. Pillai, and A. H. Wu. 2011. Gene expression profiling from formalin-fixed, paraffin-embedded tissue for tumor diagnosis. *Clinica Chimica Acta* 412(15-16):1462-1464.

Habel, L. A., S. Shak, M. Jacobs, A. Capra, C. Alexander, M. Pho, J. Baker, M. Walker, D. Watson, J. Hackett, N. T. Blick, D. Greenberg, L. Fehrenbacher, B. Langholz, and C. P. Quesenberry. 2006. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Research* 8(3):R25.

Harris, L., H. Fritsche, R. Mennel, L. Norton, P. Ravdin, S. E. Taube, M. R. Somerfield, D. F. Hayes, and R. C. Bast, Jr. 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of Clinical Oncology* 25(33):5287-5312.

Hillen, H. F. 2000. Unknown primary tumours. *Postgraduate Medical Journal* 76(901):690-693.

Hornberger, J., and R. Chien. 2010. *P2-09-06: Meta-analysis of the decision impact of the 21-gene breast cancer recurrence score in clinical practice*. Poster presented at the 33rd Annual San Antonio Breast Cancer Symposium, San Antonio, Texas, December 8-12.

Hornberger, J. C., M. Amin, G. R. Varadhachary, W. D. Henner, and J. S. Nystrom. 2011. Effect of a gene expression-based tissue of origin test's impact on patient management for difficult-to-diagnose primary cancers. *Journal of Clinical Oncology* 29(Suppl 4; abstr 459).

ISHLT (International Society of Heart and Lung Transplantation). 2010. *The International Society of Heart and Lung Transplantation Guidelines for the Care of Heart Transplant Recipients. Task Force 2: Immunosuppression and Rejection*. http://www.ishlt.org/ContentDocuments/ISHLT_GL_TaskForce2_110810.pdf (accessed January 23, 2012).

Kim, C., and S. Paik. 2010. Gene-expression-based prognostic assays for breast cancer. *Nature Reviews* 7(6):340-347.

Knauer, M., S. Mook, E. J. Rutgers, R. A. Bender, M. Hauptmann, M. J. van de Vijver, R. H. T. Koornstra, J. Bueno-de-Mesquita, S. C. Linn, and L. J. van't Veer. 2010. The predictive value of the 70-gene

signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Research and Treament* 120(3):655-661.

Kroese, M., R. L. Zimmern, and S. E. Pinder. 2007. HER2 status in breast cancer--an example of pharmacogenetic testing. *J R Soc Med* 100(7):326-329.

Laouri, M., M. Halks-Miller, W. D. Henner, and S. Nystrom. 2011. Potential clinical utility of gene expression profiling in identifying tumors of uncertain origin. *Personalized Medicine* 8(6):615-622.

Marboe, C. C., M. Billingham, H. Eisen, M. C. Deng, H. Baron, M. Mehra, S. Hunt, J. Wohlgemuth, J. Prentice, and G. Berry. 2005. Nodular endocardial infiltrates (quality lesions) cause significant variability in diagnosis of ISHLT grade 2 and 3A rejection in cardiac allograft recipients. *Journal of Heart and Lung Transplant* 24(7 Suppl.): S219-226.

Mehra, M., J. Kobashigawa, M. Deng, K. Fang, T. Klingler, P. Lal, S. Rosenberg, P. Uber, R. Starling, and S. Murali. 2007. Transcriptional Signals of T-cell and Corticosteroid-sensitive Genes Are Associated With Future Acute Cellular Rejection in Cardiac Allografts. *Journal of Heart and Lung Transplantation* 26(12):1255-1263.

Mehra, M. R., J. A. Kobashigawa, M. C. Deng, K. C. Fang, T. M. Klingler, P. G. Lal, S. Rosenberg, P. A. Uber, R. C. Starling, S. Murali, D. F. Pauly, R. Dedrick, M. G. Walker, A. Zeevi, and H. J. Eisen. 2008. Clinical Implications and Longitudinal Alteration of Peripheral Blood Transcriptional Signals Indicative of Future Cardiac Allograft Rejection. *Journal of Heart and Lung Transplantation* 27(3):297-301.

Miller, R. W., A. Smith, C. P. DeSimone, L. Seamon, S. Goodrich, I. Podzielinski, L. Sokoll, J. R. van Nagell, Jr., Z. Zhang, and F. R. Ueland. 2011. Performance of the American College of Obstetricians and Gynecologists' ovarian tumor referral guidelines with a multivariate index assay. *Obstet Gynecol* 117(6):1298-1306.

Monzon, F. A., M. Lyons-Weiler, L. J. Buturovic, C. T. Rigl, W. D. Henner, C. Sciulli, C. I. Dumur, F. Medeiros, and G. G. Anderson. 2009. Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin. *Journal of Clinical Oncology* 27(15):2503-2508.

Monzon, F. A., F. Medeiros, M. Lyons-Weiler, and W. D. Henner. 2010. Identification of tissue of origin in carcinoma of unknown primary with a microarray-based gene expression test. *Diagnostic Pathology* 5(3):10.1186/1746-1596-5-3.

Mook, S., M. K. Schmidt, G. Viale, G. Pruneri, I. Eekhout, A. Floore, A. M. Glas, J. Bogaerts, F. Cardoso, M. J. Piccart-Gebhart, E. T. Rutgers, and L. J. van't Veer. 2008. The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Research and Treament* 116(2):295-302.

Mulcahy, N. 2010. *NCCN guideline on occult cancer show immunohistochemistry is "rapidly changing*." http://www.medscape.com/viewarticle/718870 (accessed September 26, 2011).

NCCN (National Comprehensive Cancer Network). 2011a. *NCCN Guidelines Version 2.2011 Breast Cancer*. http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf (accessed August 30, 2011).

NCCN. 2011b. *NCCN Guidelines Version 2.2012 Ovarian Cancer.* http://www.nccn.org/professionals/physician_gls/pdf/ovarian.pdf (accessed December 17, 2011).

NCI (National Cancer Institute). 2002. *Clinical Proteomics Program*. http://home.crr.cancer.gov/ncifdaproteomics/ppatterns.asp (accessed October 19, 2011).

NCI. 2006. *Personalized treatment trial for breast cancer launched*. http://www.cancer.gov/newscenter/pressreleases/TAILORxRelease (accessed January 27, 2011).

NCI. 2010a. *Phase III randomized study of adjuvant combination chemotherapy and hormonal therapy versus adjuvant hormonal therapy alone in women with previously resected axillary node-negative breast cancer with various levels of recurrence (TAILORx trial)*. http://www.cancer.gov/clinicaltrials/ECOG-PACCT-1 (accessed January 27, 2011).

NCI. 2010b. *TAILORx: Testing personalized treatment for breast cancer*. http://www.cancer.gov/clinicaltrials/noteworthy-trials/tailorx (accessed January 27, 2011).

NCI. 2011. *FDA approval for trastuzumab*. http://www.cancer.gov/cancertopics/druginfo/fda-trastuzumab (accessed September 26, 2011).

Nielsen, H., F. B. Sorensen, B. Nielsen, J. P. Bagger, P. Thayssen, and U. Baandrup. 1993. Reproducibility of the acute rejection diagnosis in human cardiac allografts. The Stanford Classification and the International Grading System. *Journal of Heart and Lung Transplantation* 12(2):239-243.

Paik, S., J. Bryant, E. Tan-Chiu, E. Romond, W. Hiller, K. Park, A. Brown, G. Yothers, S. Anderson, R. Smith, D. L. Wickerham, and N. Wolmark. 2002. Real-world performance of HER2 testing--National Surgical Adjuvant Breast and Bowel Project experience. *Journal of the National Cancer Institute* 94(11):852-854.

Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, R. Baehner, M. Walker, D. Watson, and T. Park. 2003. *Multi-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients--NSABP studies B-20 and B-14: Abstract #16*. Paper presented at San Antonio Breast Cancer Symposium, San Antonio, TX.

Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, and N. Wolmark. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine* 351(27):2817-2826.

Paik, S., G. Tang, S. Shak, C. Kim, J. Baker, W. Kim, M. Cronin, F. L. Baehner, D. Watson, J. Bryant, J. Costantino, C. E. Geyer, Jr., D. L. Wickerham, and N. Wolmark. 2006. Gene expression and benefit of chemotherapy in node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology* 24(23):3726-3734.

Paik, S., C. Kim, and N. Wolmark. 2008. HER2 status and benefit from adjuvant trastuzumab in breast cancer. *New England Journal of Medicine* 358(13):1409-1411.

Pathwork Diagnostics. 2010. *Pathwork Tissue of Origin test for FFPE cleared by U.S. Food and Drug Administration*. http://www.pathworkdx.com/News/M129_FDA_Clearance_Final.pdf (accessed November 17, 2011).

Pathwork Diagnostics. 2011a. *Pathwork Reimbursement Assistance Program (RAP)*. http://www.pathworkdx.com/patient_information/reimbursement1/ (accessed November 22, 2011).

Pathwork Diagnostics. 2011b. *The Pathwork Tissue of Origin Test*. http://www.pathworkdx.com/TissueOfOriginTest/IVDKit/ (accessed November 16, 2011).

Pavlidis, N., and Y. Merrouche. 2006. The importance of identifying CUP subsets. In *Carcinoma of an Unknown Primary Site,* edited by K. Fizazi,. New York: Taylor & Francis Group.

Pavlidis, N., E. Briasoulis, J. Hainsworth, and F. A. Greco. 2003. Diagnostic and therapeutic management of cancer of unknown primary. *European Journal of Cancer* 39(14):1990-2005.

Perez, E. A., M. M. Reinholz, D. W. Hillman, K. S. Tenner, M. J. Schroeder, N. E. Davidson, S. Martino, G. W. Sledge, L. N. Harris, J. R. Gralow, A. C. Dueck, R. P. Ketterling, J. N. Ingle, W. L. Lingle, P. A. Kaufman, D. W. Visscher, and R. B. Jenkins. 2010. HER2 and chromosome 17 effect on patient outcome in the N9831 adjuvant trastuzumab trial. *J Clin Oncol* 28(28):4307-4315.

Petricoin, E. F., A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta. 2002. Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* 359(9306): 572-577.

Pham, M. X., J. J. Teuteberg, A. G. Kfoury, R. C. Starling, M. C. Deng, T. P. Cappola, A. Kao, A. S. Anderson, W. G. Cotts, G. A. Ewald, D. A. Baran, R. C. Bogaev, B. Elashoff, H. Baron, J. Yee, and H. A. Valantine. 2010. Gene-expression profiling for rejection surveillance after cardiac transplantation. *New England Journal of Medicine* 362(20):1890-1900.

Phillips, K. A., D. A. Marshall, J. S. Haas, E. B. Elkin, S. Y. Liang, M. J. Hassett, I. Ferrusi, J. E. Brock, and S. L. Van Bebber. 2009. Clinical practice patterns and cost effectiveness of human epidermal growth receptor 2 testing strategies in breast cancer patients. *Cancer* 115(22):5166-5174.

Pillai, R., R. Deeter, C. T. Rigl, J. S. Nystrom, M. H. Miller, L. Buturovic, and W. D. Henner. 2011. Validation and reproducibility of a microarray-based gene expression test for tumor identification in formalin-fixed, paraffin-embedded specimens. *Journal of Molecular Diagnostics* 13(1):48-56.

Pollack, A. 2004. New cancer test stirs hope and concern. *The New York Times* http://www.nytimes.com/2004/02/03/science/new-cancer-test-stirs-hope-and-concern.html?src=pm (accessed November 23, 2011).

Pollack, A. 2008a. Cancer test for women raises hope, and concern. *The New York Times*. http://www.nytimes.com/2008/08/26/health/26ovar.html?pagewanted=all (accessed November 23, 2011).

Pollack, A. 2008b. Sales of test for ovarian cancer halted. *The New York Times* http://www.nytimes.com/2008/10/25/business/25cancer.html (accessed November 23, 2011).

PR Newswire. 2009. U.S. Food and Drug Administration clears Vermillion's OVA1(TM) test to determine likelihood of ovarian cancer in women with pelvic mass. http://www.prnewswire.com/news-releases/us-food-and-drug-administration-clears-vermillions-ova1tm-test-to-determine-likelihood-of-ovarian-cancer-in-women-with-pelvic-mass-62150212.html (accessed December 17, 2011).

Quest Diagnostics. 2011. *Licenses and accreditation*. http://www.questdiagnostics.com/brand/company/b_comp_licenses.html (accessed November 21, 2011).

Ransohoff, D. F. 2003. Gene-expression signatures in breast cancer. *New England Journal of Medicine* 348(17):1716.

Ransohoff, D. F. 2004. Rules of evidence for cancer molecular-marker discovery and validation. *Nature Reviews Cancer* 4(4):309-314.

Ransohoff, D. F. 2005. Lessons from controversy: ovarian cancer screening and serum proteomics. *Journal of the National Cancer Institute* 97(4):315-319.

Roche, P. C., V. J. Suman, R. B. Jenkins, N. E. Davidson, S. Martino, P. A. Kaufman, F. K. Addo, B. Murphy, J. N. Ingle, and E. A. Perez. 2002. Concordance between local and central laboratory HER2 testing in the breast intergroup trial N9831. *J Natl Cancer Inst* 94(11):855-857.

Rosenberg, S., M. R. Elashoff, P. Beineke, S. E. Daniels, J. A. Wingrove, W. G. Tingley, P. T. Sager, A. J. Sehnert, M. Yau, W. E. Kraus, L. K. Newby, R. S. Schwartz, S. Voros, S. G. Ellis, N. Tahirkheli, R. Waksman, J. McPherson, A. Lansky, M. E. Winn, N. J. Schork, and E. J. Topol. 2010. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Ann Intern Med* 153(7):425-434.

Sauter, G., J. Lee, J. M. Bartlett, D. J. Slamon, and M. F. Press. 2009. Guidelines for human epidermal growth factor receptor 2 testing: Biologic and methodologic considerations. *Journal of Clinical Oncology* 27(8):1323-1333.

Schmitt, F. 2009. HER2+ breast cancer: How to evaluate? *Advanced Therapeutics* 26(Suppl 1):S1-S8.

Shah, S., and B. Chen. 2010. Testing for HER2 in breast cancer: A continuing evolution. *Pathology Research International* 2011:903202.

Shak, S. 2011. *Case study: Oncotype DX breast cancer assay*. Presentation at Meeting 2 of the Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30.

Simon, R., M. D. Radmacher, K. Dobbin, and L. M. McShane. 2003. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute* 95(1):14-18.

Simon, R. M., S. Paik, and D. F. Hayes. 2009. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *Journal of the National Cancer Institute* 101(21):1446-1452.

Sorace, J. M. and M. Zhan. 2003. A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* 4(24): 10.1186/1471-2105-4-24.

Stancel, G. A., D. Coffey, K. Alvarez, M. Halks-Miller, A. Lal, D. Mody, T. Koen, T. Fairley, and F. A. Monzon. 2011. Identification of tissue of origin in body fluid specimens using a gene expression microarray assay. *Cancer Cytopathology* doi: 10.1002/cncy.20167.

Starling, R. C., M. Pham, H. Valantine, L. Miller, H. Eisen, E. R. Rodriguez, D. O. Taylor, M. H. Yamani, J. Kobashigawa, K. McCurry, C. Marboe, M. R. Mehra, A. Zuckerman, M. C. Deng, and Working Group on Molecular Testing in Cardiac Transplantation. 2006. Molecular testing in the management of cardiac transplant recipients: Initial clinical experience. *Journal of Heart and Lung Transplation* 25(12):1389-1395.

SWOG (Southwest Oncology Group). 2011. *Spotlight: RxPONDER trial will evaluate whether gene expression test can drive chemotherapy choice*. http://swog.org/visitors/newsletters/2011/04/index.asp?a=spotlight (accessed May 5, 2011).

Thomas, G. S., S. Voros, J. A. McPherson, A. J. Lansky, F. L. Weiland, S. C. Cheng, S. A. Bloom, H. Salha, M. R. Elashoff, B. O. Brown, H. D. Lieu, A. Johnson, S. E. Daniels, S. Rosenberg. 2011. The Compass trial (NCT01117506): A prospective multi-center, doubleblind study assessing a whole blood gene expression test for the detection of obstructive coronary artery disease In symptomatic patients referred for myocardial perfusion imaging. Abstract presented at American Heart Association Meeting, November 15, 2011.

Ueland, F. R., C. P. Desimone, L. G. Seamon, R. A. Miller, S. Goodrich, L. Podzielinski, L. Sokoll, A. Smith, J. R. van Nagell, and Z. Zhang. 2011. Effectiveness of a Multivariate Index Assay in the Preoperative Assessment of Ovarian Tumors. *Obstetrics and Gynecology* 117(6):1289-1297.

van de Vijver, M. J., Y. D. He, L. J. van't Veer, H. Dai, A. A. M. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteven, A. M. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E. T. Rutgers, S. F. Friend, and R. Bernards. 2002. A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* 347(25):1999-2009.

van't Veer, L. J., H. Dai, M. J. van de Vijver, Y. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Wittereveen, G. J. Schreiber, R. M. Kerkoven, C. Roberts, P. S. Linsley, R. Bernards, and S. F. Friend. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(31):530-536.

Vermillion. 2011. *Payor information*. http://ova-1.com/resources/payor-information (accessed October 11, 2011).

Visintin, I., Z. Feng, G. Longton, D. C. Ward, A. B. Alvero, Y. Lai, J. Tenthorey, A. Leiser, R. Flores-Saaib, H. Yu, M. Azori, T. Rutherford, P. E. Schwartz, and G. Mor. 2008. Diagnostic markers for early detection of ovarian cancer. *Clinical Cancer Research* 14(4):1065-1072.

Wagner, L. 2004. A test before its time? FDA stalls distribution process of proteomic test. *Journal of the National Cancer Institute* 96(7):500-501.

Wingrove, J. A., S. E. Daniels, A. J. Sehnert, W. Tingley, M. R. Elashoff, S. Rosenberg, L. Buellesfeld, E. Grube, L. K. Newby, G. S. Ginsburg, and W. E. Kraus. 2008. Correlation of peripheral-blood gene expression with the extent of coronary artery stenosis. *Circulation Cardiovascular Genetics* 1:31–38

Wittner, B. S., D. C. Sgroi, P. D. Ryan, T. J. Bruinsma, A. M. Glas, A. Male, S. Dahiya, K. Habin, R. Bernards, D. A. Haber, L. J. van't Veer, and S. Ramaswamy. 2008. Analysis of the MammaPrint breast cancer assay in a predominantly postmenopausal cohort. *Clinical Cancer Research* 14(10):2988-2993.

Wolff, A. C., M. E. Hammond, J. N. Schwartz, K. L. Hagerty, D. C. Allred, R. J. Cote, M. Dowsett, P. L. Fitzgibbons, W. M. Hanna, A. Langer, L. M. McShane, S. Paik, M. D. Pegram, E. A. Perez, M. F. Press, A. Rhodes, C. Sturgeon, S. E. Taube, R. Tubbs, G. H. Vance, M. van de Vijver, T. M. Wheeler, and D. F. Hayes. 2007a. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Journal of Clinical Oncology* 25(1):118-145.

Wolff, A. C., M. E. Hammond, J. N. Schwartz, K. L. Hagerty, D. C. Allred, R. J. Cote, M. Dowsett, P. L. Fitzgibbons, W. M. Hanna, A. Langer, L. M. McShane, S. Paik, M. D. Pegram, E. A. Perez, M. F. Press, A. Rhodes, C. Sturgeon, S. E. Taube, R. Tubbs, G. H. Vance, M. van de Vijver, T. M. Wheeler, and D. F. Hayes. 2007b. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Archives of Pathology and Laboratory Medicine* 131(1):18-43.

Wu, A. H., J. C. Drees, H. Wang, S. R. VandenBerg, A. Lal, W. D. Henner, and R. Pillai. 2010. Gene expression profiles help identify the tissue of origin for metastatic brain cancers. *Diagnostic Pathology* 5(26):10.1186/1746-1596-5-26.

Zhang, Z., and D. W. Chan. 2010. The road from discovery to clinical diagnostics: Lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiology, Biomarkers, and Prevention* 19(12):2995-2999.

Zhang, Z., R. C. Bast Jr., Y. Yu, J. Li, L. J. Sokoll, A. J. Rai, J. M. Rosenzweig, B. Cameron, Y. Y. Wang, X. Y. Meng, A. Berchuck, C. Van Haaften-Day, N. F. Hacker, H. W. de Bruijn, A. G. van der Zee, I. J. Jacobs, E. T. Fung, and D. W. Chan. 2004. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Research* 64(16):5882-5890.

# Appendix B
# Gene Expression–Based Tests Developed at Duke University and Used in Clinical Trials

The emergence of high-throughput omics technologies beginning around the mid-1990s led to development of new approaches for studying the dynamics of biological systems. Multidisciplinary collaborations were formed among molecular biologists, bioinformatics experts, and statisticians at many institutions to devise experimental strategies and statistical methods for the analysis and interpretation of these rich new sources of data. At Duke University, researchers were pursuing these new avenues of research. In 2000, Joseph Nevins and Mike West founded the Computational and Applied Genomics Program (CAGP), a multidisciplinary research program (Kornbluth and Dzau, 2010). The CAGP formed the basis for what later became the Center for Applied Genomics and Technology (CAGT). As one of the initial centers of the Duke Institute for Genome Science and Policy (IGSP), which was formed in 2003 (Kornbluth and Dzau, 2010), CAGT researchers used various types of genomic analyses to elucidate potential mechanisms of oncogenesis and to understand the complexity of cancer phenotypes. DNA microarray analysis became a powerful tool in the CAGP/CAGT for the study of regulatory pathways essential for cancer initiation and tumor growth, and researchers developed several gene expression–based tests to predict patient responses to chemotherapeutic agents and published the results. At a very early stage in the discovery research, such tests were taken into clinical trials. The primary publications were criticized for major problems in data presentation and statistical analysis. Eventually, concerns were raised by statisticians about the validity of the tests and about potential harm to patients enrolled in the trials.

The Institute of Medicine (IOM) committee's statement of task refers to three trials that were conducted at Duke University. Table B-1 outlines some information related to those trials.

**TABLE B-1** Clinical Trials Related to Duke University Gene Expression–Based Tests and the Clinical Trials Listed in the Institute of Medicine Committee's Statement of Task

| ClinicalTrials.gov Number | NCT00636441 | NCT00509366 | NCT00545948 |
|---|---|---|---|
| ID | BOP0801 | TOP0602 | TOP0703 |
| Description | A Randomized Phase II Trial Evaluating the Performance of Genomic Expression Profiles to Direct the Use of Preoperative Chemotherapy for Early Stage Breast Cancer | Phase II Prospective Study Evaluating the Role of Personalized Chemotherapy Regimens for Chemo-Naive Select Stage IIIB and IV Non-Small Cell Lung Cancer (NSCLC) in Patients Using a Genomic Predictor of Platinum Resistance to Guide Therapy | Phase II Prospective Study Evaluating the Role of Directed Cisplatin-Based Chemo With Either Vinorelbine or Pemetrexed for the Adj[uvant] T[reatment] of Early Stage NSCLC in Patients Using Genomic Expression Profiles of Chemo Sensitivity to Guide Therapy |
| Disease | Breast cancer | Lung cancer | Lung cancer |
| Start date | April 2008 | February 2007 | October 2007 |
| Trial listed in | March 2008 | July 2007 | October 2007 |

ClinicalTrials.gov

| | | | |
|---|---|---|---|
| Patient accrual | | | |
| Intended | 270 | 80 | 117 |
| Actual[a] | 56 | 47 | 24 |
| Sponsor | DoD | Eli Lilly/Duke/NCI | Eli Lilly/Duke |
| Principal Investigator(s) | Paul K. Marcom, M.D., Duke University | Gordana Vlahovic, M.D., M.H.S., Duke University | Neal Ready, Ph.D., M.D., Duke University Medical Center, Hematology/Oncology, Duke Comprehensive Cancer Center |
| Chemosensitivity predictor | Doxorubicin (Adriamycin) and docetaxel (prospective) | Cisplatin (prospective) | Pemetrexed and vinorelbine (prospective) |
| Termination date | 11/4/2010 | 11/4/2010 | 2/3/2011 |
| Citations in ClinicalTrials.gov | Potti et al. (2006a) | Bild et al. (2006); Potti et al. (2006a) | Potti et al. (2006a); Potti et al. (2007b); Potti et al. (2006b) |

NOTES: NSCLC = DoD = Department of Defense, NCI = National Cancer Institute, non-small cell lung cancer,.

[a]Personal communication from Michael Cuffe , Duke University School of Medicine, July 23, 2010.

This appendix provides a concise summary of the research objectives and the approaches taken in developing several of the gene expression–based chemosensitivity tests implemented in the three clinical trials in Table B-2, and presents findings that provide important insights about processes that were in place at Duke University, to enlighten the development of and to provide motivation for many of the IOM committee's recommendations that are intended to enhance the integrity of future omics-related research. Many of these findings are in key areas that include the responsibilities of investigators and institutions, conflict of interest issues, and the roles of funders, regulatory authorities, journals, and biostatistical collaborators.

## DEVELOPMENT AND EVALUATION PROCESS

Investigators are responsible for systematic and rigorous development of omics-based tests. Chapters 2, 3, and 4 explained the IOM committee's recommendations on omics-based test discovery, development, and evaluation for clinical use. These recommendations are meant to help establish a process, agreed on by all collaborating disciplines, for the discovery and development of omics-based tests with the goal of improving patient care and outcomes.

### Discovery and Test Validation Phases

Chapter 2 explained the technologies, statistical methods, computational methods, and bioinformatics methods that should be used in the discovery, confirmation, and validation of omics-based tests. Recommendation 1 presented critical steps in the discovery and confirmation of new omics-based test concepts. Recommendation 2 (Chapter 3) focused on omics-based test development and validation within the *Clinical Laboratory Improvement Amendments of 1988* (CLIA) setting, in preparation for use in patient management decisions in clinical trials or for eventual use in patient management decisions in medical care. These steps include the design, optimization, validation, and implementation of the locked-down test in single or multiple CLIA-

certified laboratories. Recommendation 2 also covers discussion with the Food and Drug Administration (FDA) prior to validation.

The sections below present facts from the discovery and validation phases of the gene expression–based tests developed at Duke University and used in the three clinical trials the committee was tasked to evaluate: (1) tests for docetaxel and doxorubicin (Adriamycin) sensitivity were used in the trial NCT00636441; (2) a test for cisplatin sensitivity was used in the trial NCT00509366; and (3) tests for pemetrexed and vinorelbine sensitivity were used in the trial NCT00545948. For each, a brief explanation of test discovery and validation is given, including information on the confirmation of the gene expression–based signatures; the availability of the data, metadata, code, and fully specified computational procedures used in the discovery and confirmation of the test; and whether the tests were locked down prior to progression to subsequent phases of test development.

Information regarding the CLIA laboratory and FDA aspects of test validation is general rather than specific for each of the tests discussed below. Communication with the FDA is discussed later in this appendix. The committee had little information relating to the design, optimization, validation, and implementation of the tests in the CLIA-certified laboratory. At the March 2011 meeting, Nevins informed the committee that, at the time of performance testing, the laboratory was CLIA-registered. A certificate of registration does not indicate CLIA compliance but only that a CLIA application was submitted to CMS, but does allow a laboratory to perform moderate and high complexity testing until an onsite survey is performed leading to CLIA certification if compliance to the regulatory standards is demonstrated. The laboratory became CLIA certified during the course of the trials. He stated that the investigators had implemented data quality control and security systems as well as an automated system for running the computational models that would ensure high-quality, reliable data (Nevins, 2011). The clinical trial protocols indicate that patient sample processing and microarray analyses were conducted in a CLIA-certified laboratory setting (Marcom, 2008; Ready, 2010; Vlahovic, 2010). It is not clear from the trial protocols where the computational models were performed on the data, but the Duke Clinical Genomics Studies Unit defined operational standards for "array data analysis through an automated system designed and controlled by a Duke Faculty biostatistician" (Kornbluth and Dzau, 2010, p. 5). Two of the protocols note that the data were available for quality assessment and analysis to a computational biologist (Ready, 2010; Vlahovic, 2010). As noted by Baggerly and by Lisa McShane, the computational models were not locked down when their performance was evaluated prior to use in the clinical trials (Baggerly, 2011; McShane, 2010a). As described in Chapter 3 and reflected in the committee's recommendations, this constitutes a serious flaw in the test development process.

### *Docetaxel and Doxorubicin Chemosensitivity Tests (Potti et al., 2006a) Used in Breast Cancer Trial NCT00636441*

The Duke researchers first published gene expression–based chemosensitivity tests for docetaxel and doxorubicin in the 2006 *Nature Medicine* paper (Potti et al., 2006a). This paper also presented chemosensitivity tests for five other chemotherapeutic drugs: paclitaxel, topotecan, 5-FU, cyclophosphamide, and etoposide. The drugs were chosen based on availability of gene expression microarray data and in vitro drug response (sensitivity) measures from the NCI-60 cell line panel from the National Cancer Institute, or NCI (Potti et al., 2006a).

A subsequent study conducted to evaluate the ability of the docetaxel and doxorubicin tests to predict patient response to a combination taxane chemotherapy regimen (docetaxel and epirubicin; abbreviated TET) or a non-taxane chemotherapy regimen (fluorouracil, epirubicin, and cyclophosphamide; abbreviated FEC), respectively, was published in 2007 (Bonnefoi et al., 2007). Both papers have now been retracted (Bonnefoi et al., 2011; Potti et al., 2011a).

The Duke researcher's general approach for identifying signatures for each of the drugs was to first identify cell lines from the NCI-60 panel that were the most sensitive and resistant to the drugs. Then, they used statistical methods to develop the gene expression–based signatures that would form the basis of the computational models in the tests. However, conflicting and confusing information in the papers and the cited references regarding the data and the statistical methods contributed to the inability of colleagues in the scientific community to understand and to replicate the generation of the computational models (Baggerly, 2011; McShane, 2010a; *Review of genomic predictors for clinical trials from Nevins, Potti, and Barry*, 2009). For example, the authors describe using Bayesian binary regression analysis, but the paper cited for this analysis (Pittman et al., 2004) presents a different statistical methodology for Bayesian binary prediction tree models. In addition, there were simple linear regression analyses reported in which p-values were stated to have been obtained by use of a log-rank test. The log-rank test is a statistical testing method applied for analysis of survival (time-to-event) data; its citation in the simple linear regression setting should have signaled a need for statistical review. The committee does not know if the paper was reviewed by a statistician either internally at Duke or during the *Nature Medicine* review process, but whatever statistical review occurred for this paper was inadequate. These instances point to the risks of relying on journal publication as the sole basis for judging the soundness of science, particularly when the results are poised for translation into the clinic.

Several datasets were used to confirm the gene expression signatures generated. Potti et al. (2006a) reported using leave-one-out cross-validation to confirm the docetaxel signature developed from NCI-60 breast cancer drug sensitivity data. The docetaxel test was reported to have been validated on several independent sets of data from ovarian and lung cancer cell lines and from clinical samples of breast and ovarian tumors; some of these data had been previously published and others were generated at Duke. The doxorubicin test also was reported to have been confirmed using leave-one-out cross-validation and then validated on independent gene expression datasets from breast, ovarian, and leukemia studies (Bonnefoi et al., 2007; Potti et al., 2006a).

Both the docetaxel and doxorubicin tests were used as part of computational models developed to predict response to multidrug chemotherapy regimens. Potti et al. (2006a) reported that when a predictor model for combined TFAC (paclitaxel, 5-FU, Adriamycin, and cyclophosphamide) was applied to gene expression data from 51 patients from a breast neoadjuvant treatment study, there was a statistically significant association between the predicted multiregimen response probability and response outcome. Similar statistically significant results were reported from a second collection of breast cancer specimens from patients who had received FAC (5-FU, Adriamycin, cyclophosphamide). Bonnefoi et al. (2007) reported good performance of multidrug predictors when tested on samples from the intergroup neoadjuvant therapy trial EORTC-10994/BIG-00-01, which randomized patients with estrogen receptor-negative breast tumors between treatment arms for TET (docetaxel for three cycles followed by epirubicin plus docetaxel) and FEC (fluorouracil, epirubicin, and

cyclophosphamide). The doxorubicin signature developed in Potti et al. (2006a) was used in lieu of an epirubicin signature. The reported successful extension of the predictor model methodology to multidrug regimens was seen as important because many cancer patients receive multidrug chemotherapy regimens.

Several aspects of the validations reported in Bonnefoi et al. (2011) and Potti et al. (2006a) raise questions about the rigor with which those validations were conducted. There was a lack of information about how the thresholds applied to the response probabilities generated by the predictor models were selected for the validations involving clinical samples in these studies (Bonnefoi et al., 2011; Potti et al., 2006a), and the reported use of different thresholds for the two tumor types (breast and ovarian) indicates that these two validation studies on clinical samples could not have been based on an appropriately locked down model (which must include locking down any threshold). In addition, neither paper states that the investigators were blinded to the response outcome data when they calculated the predicted response probabilities (Bonnefoi et al., 2011; Potti et al., 2006a). The Bonnefoi et al. (2007) paper states that several authors had full access to all of the raw data, but it is not known when in the course of the study they may have used that access.

The drug sensitivity measures and gene expression microarray data used to develop the docetaxel and doxorubicin predictors were publicly available in the database from the NCI-60 website.[1] Code used to generate signatures in Potti et al. (2006a) was available on a Duke website (Baggerly and Coombes, 2009). However, when Baggerly and Coombes attempted to assess the validity of the tests at the request of colleagues at MD Anderson Cancer Center who were interested in using the tests or the same approach to develop new tests, they found insufficient information to reproduce the published results, using the available data and the methods published in the *Nature Medicine* paper (Baggerly, 2011). Therefore, Baggerly and Coombes began corresponding with the principal authors at Duke to better understand the data and methodology. At first there was an exchange of questions and answers regarding the data, cell line labels, and gene lists. However, after multiple exchanges between November 2006 and June 2007, Baggerly and Coombes were still unable to reproduce the results and communications between the groups broke off (Baggerly, 2011). Statisticians Kevin Coombes, Jing Wang, and Keith Baggerly submitted correspondence to *Nature Medicine* outlining their unresolved concerns and questions. Their correspondence was published along with a reply (Coombes et al., 2007; Potti and Nevins, 2007). The concerns included inability to reproduce the selection of cell lines from sensitivity measures, errors in gene lists, incorrect figures, combining of training and test sets in developing predictors, and inability to produce the reported predictor performance results. Further communication between Baggerly and Coombes and the authors and journals is described in the section on journals later in this appendix. When the *Nature Medicine* paper was eventually retracted on January 7, 2011, corruption of additional validation datasets was noted, with an explicit statement that the authors had been "unable to reproduce certain crucial experiments showing validation of signatures for predicting response to chemotherapies, including docetaxel and topotecan" (Potti et al., 2011a, p. 135).

The clinical trial using this predictor, NCT00636441, titled *A Randomized Phase II Trial Evaluating the Performance of Genomic Expression Profiles to Direct the Use of Preoperative Chemotherapy for Early Stage Breast Cancer*, was listed in ClinicalTrials.gov on March 9, 2008. This trial was temporarily suspended from October 19, 2009, to February 12, 2010. The trial was

[1] See http://dtp.nci.nih.gov/docs/cancer/cancer_data.html (Potti et al., 2006a).

suspended again on July 23, 2010, and terminated on November 4, 2010. This trial and the following two clinical trials named in the Statement of Task are discussed in more detail later in this appendix (see section on evaluation for clinical use).

*Cisplatin Chemosensitivity Test (Hsu et al., 2007) Used in Lung Cancer Patients in NCT00509366*

The gene expression–based chemosensitivity test for cisplatin was published in the *Journal of Clinical Oncology* (Hsu et al., 2007), along with a chemosensitivity test for pemetrexed; this paper has now been retracted due to the "inability to reproduce the experiments demonstrating a capacity of a cisplatin response signature to validate in either a collection of ovarian cancer cell lines or ovarian tumor samples" (Hsu et al., 2010, p. 5229). The general statistical approach used to develop the models was similar to the one reported in Potti et al. in *Nature Medicine* (2006a); the authors had made code available on a Duke website. The cisplatin test was developed using publicly available gene expression microarray data and drug sensitivity data from a study published in the *International Journal of Cancer* (Gyorffy et al., 2006). Hsu et al. (2007) reported that the cisplatin test had been validated in two experiments. The first experiment used ovarian cancer cell lines for which the Duke investigators had performed drug sensitivity experiments and gene expression microarray profiling. A second experiment used clinical specimens from patients with ovarian cancer. There were no reported validation attempts using clinical tumor samples from patients with lung cancer, but the first trial in which the cisplatin test was used to guide therapy was the NCT00509366 trial for advanced lung cancer. As indicated in Figure S-1, the omission of such a validation step constitutes a critical flaw in the test development process.

Problems with posted data and figures were identified by Baggerly and Coombes for both the cisplatin and pemetrexed tests (Baggerly and Coombes, 2009). For example, they identified off-by-one errors in gene lists for both predictors, "outlier" genes reported for the cisplatin predictor that could not be reproduced from the data (even after accounting for the off-by-one error), and a reversal of sensitive/resistant labels in a data figure for the pemetrexed predictor. Baggerly and Coombes (2009) noted in their analysis: "one theme that emerges is that the most common errors are simple (e.g., row or column offsets); conversely, it is our experience that the most simple errors are common." The statisticians were particularly concerned that the four outlier genes (probesets) mistakenly reported for the cisplatin predictor were exactly those cited in Hsu et al. (2007) as providing biological plausibility for the model. Even with access to the publicly available primary data and code posted by the authors on a Duke website, Baggerly and Coombes were unable to reproduce the published results. Further information on Baggerly and Coombes's examination of the cisplatin and several other predictors is given later in this appendix.

The clinical trial, NCT00509366, titled *Phase II Prospective Study Evaluating the Role of Personalized Chemotherapy Regimens for Chemo-Naive Select Stage IIIB and IV Non-Small Cell Lung Cancer (NSCLC) in Patients Using a Genomic Predictor of Platinum Resistance to Guide Therapy*, began accruing patients in June 2007 (McShane, 2010b) and was listed in ClinicalTrials.gov on July 30, 2007, temporarily suspended from October 6, 2009, to February 12, 2010, resuspended on July 23, 2010, and terminated on November 4, 2010.

*Pemetrexed (Hsu et al., 2007) and Vinorelbine Chemosensitivity Tests Used in Clinical Trial of Lung Cancer Patients NCT00545948*

The gene expression–based chemosensitivity test for pemetrexed was published in the *Journal of Clinical Oncology* (Hsu et al., 2007); as mentioned in the previous section, this paper has now been retracted (Hsu et al., 2010). The gene expression–based chemosensitivity test for vinorelbine does not appear to have been published; the protocol for NCT00545948 cites Potti et al., *Nature Medicine* (2006a) as the relevant reference (Ready, 2010). The general statistical approach used to develop the model for pemetrexed was similar to that in Potti et al. (2006a). The pemetrexed test was developed using methods similar to those used to develop the cisplatin test, but the data source was different. This test was developed using the publicly available gene expression profiles and drug sensitivity data on the NCI-60 cell lines. Hsu et al. (2007) reported that the pemetrexed test had been validated using in vitro drug sensitivity data from an independent set of 17 NSCLC cell lines. This appears to have been the only validation study conducted before the pemetrexed test was used to direct patient therapy in the NCT00545948 clinical trial. In this trial, the pemetrexed test was used along with a similar gene expression–based test for vinorelbine sensitivity to determine which of those drugs should be coupled with cisplatin for adjuvant therapy.

As mentioned in the previous section, problems with posted data and figures were identified by Baggerly and Coombes for both the cisplatin and pemetrexed tests. They were able to detect these problems using the data that were available from the NCI-60 website and the same computer code mentioned in the previous two sections that was also used for this predictor (Baggerly and Coombes, 2009). Further information on their examination of the pemetrexed and several other predictors is given later in this appendix. No information is available relating to the vinorelbine predictor.

The clinical trial, NCT00545948, titled *Phase II Prospective Study Evaluating the Role of Directed Cisplatin Based Chemo With Either Vinorelbine or Pemetrexed for the Adj[uvant] T[herapy] of Early Stage NSCLC in Patients Using Genomic Expression Profiles of Chemo Sensitivity to Guide Therapy*, was listed in ClinicalTrials.gov on October 17, 2007, temporarily suspended from October 6, 2009, to February 11, 2010, suspended again on July 23, 2010, and terminated on February 3, 2011.

## Evaluation for Clinical Use Stage

Chapter 4 presented the committee's third recommendation, on steps important for taking a validated omics-based test into clinical trials. The decisions to move the tests into clinical trials and subsequent decisions about use of the tests to guide therapy in the clinical trials are described in greater detail in the next section on Roles and Responsibilities. The series of events following publication of the Baggerly and Coombes paper in the *Annals of Applied Statistics* (2009), as described below, applies to all three clinical trials and related tests (docetaxel and doxorubicin chemosensitivity tests used in NCT00636441, cisplatin chemosensitivity test used in NCT00509366, pemetrexed chemosensitivity test used in NCT00545948). No information is available about the vinorelbine signature.

In September 2009, NCI was in the process of reviewing the revised clinical trial protocol CALGB-30702, which was proposing to use six of the Duke chemosensitivity predictors in an advanced lung cancer trial, and the reviewers had noted serious discrepancies in the information presented in the protocol and a lack of validation of the predictors on human lung tumor samples.

NCI disapproved that protocol. However the protocol also mentioned several Duke trials already underway using several of the predictors. The concerns generated by this protocol, along with the publication of the Baggerly and Coombes paper (2009), led NCI to contact leadership at Duke University, and ultimately resulted in suspension of the trials and launch of the external review in early October 2009.

These events prompted NCI to further scrutinize another predictor developed by Nevins and Potti (but not one of the tests being studied in the three clinical trials listed in the committee's statement of task), the Lung Metagene Score (LMS), for which a clinical trial had already opened. In that trial, CALGB-30506, the LMS predictor was being used as a stratification factor for randomization. During the protocol review process for CALGB-30506, NCI decided that, while the LMS predictor appeared to have some promise, there were concerns that laboratory batch effects might influence its performance. Therefore, NCI insisted on a change in the originally proposed design of the trial so that the predictor would not be used to direct therapy in the trial. Although results of the predictor were kept blinded and were not being used to guide therapy in the trial, evaluation of the predictor was a co-primary aim of the trial. In November 2009, NCI's Cancer Therapy Evaluation Program (CTEP) made a request to the Cancer and Leukemia Group B (CALGB) for data and computer code to reevaluate that predictor and information that had been provided to CTEP during its original protocol review process for that trial 2 years earlier, when NCI did not have access to the data and computer code. With data and computer code in hand, NCI's reevaluation was able to identify a number of problems with the version of the LMS predictor that had been the basis for the trial approval and a supporting publication (Potti et al., 2006b, *New England Journal of Medicine*). The problems included an unstable computational algorithm and an inability to reproduce findings from a prevalidation exercise that had taken place during the trial approval process (McShane, 2010a). Eventually, the *New England Journal of Medicine* article was retracted due to "failure to reproduce results supporting the validation of the lung metagene model described in the article using a sample set from a study by the American College of Surgeons Oncology Group (ACOSOG) and a collection of samples from a study by the Cancer and Leukemia Group B (CALGB)" (Potti et al., 2011b, p. 1176).

In contrast to NCI's reviews, oversight committees at Duke did not recognize significant problems with the other Duke chemosensitivity predictors, and allowed them to be used to direct therapy selection in clinical trials. It is not known if the Cancer Protocol Review Committee (CPRC) and Duke Institutional Review Board (IRB), who were responsible for approving and overseeing the Duke trials, were fully aware of the extent of problems with the published papers or aware of contradictory statements being made about the validation status of some of the tests. For example, the IOM committee received conflicting information about validation of the pemetrexed test. Information supporting the lack of validation included correspondence between Potti and NCI. In Potti's submission of R01-CA131049-01A1 in March 2008, Potti stated: "we have only been able to validate the accuracy of the cisplatin test in independent patient samples…, not the pemetrexed test … it is probably a little bit premature to employ the pemetrexed test to stratify patients" (NCI, 2010a). Potti also mentioned the "premature" status of the pemetrexed test in his 4/14/10 response to NCI's letter dated 4/13/10 requesting information about his grant.[2] Information suggesting that the tests had been validated was included in the

[2] Communication from Anil Potti, Duke University, to William Timmer, National Cancer Institute, RE: R01CA131049-01A1 Information Request, April 14, 2010.

protocol for the TOP0703 that was using the pemetrexed and vinorelbine predictors. In Section 1.4.2 of the 4/21/08 version of the trial protocol, it is stated, "Using Affymetrix gene expression data with corresponding in vitro drug response data for vinorelbine and pemetrexed, our group has developed robust gene expression based models predictive of vinorelbine and pemetrexed sensitivity. These multigene models were validated with an accuracy of greater than 85% in independent in vitro studies of lung cell lines treated with vinorelbine and pemetrexed respectively." There is no mention of validation using clinical samples. It is possible this represents confusion between in vitro validation (i.e., cell lines) and validation on human tumor samples. Despite reservations expressed by Potti about use of the pemetrexed predictor for directing patient therapy in 2008-2009, and an apparent absence of published validation results for the vinorelbine predictor, TOP0703 was opened to accrual and was listed in Clinicaltrials.gov in October 2007. Both the pemetrexed and vinorelbine predictors were being used to select therapy in that trial.

## ROLES AND RESPONSIBILITIES

This section of the appendix explores the actions of the Principal Investigators (PIs), university, funders, and journals involved in the Duke case. It begins with a discussion of investigator responsibility, of Duke University's existing infrastructure and oversight during the launch and conduct of the three clinical trials mentioned in the IOM committee's statement of task (referred to as "the three clinical trials" hereinafter), and of the University's response to the scientific controversy. Topics include a discussion on oversight of research, the need for an Investigational Device Exemption (IDE), conflict of interest (COI) management, the whistleblowing system, the investigation into the controversy, and the nature of biostatistical collaboration. Subsequent sections address the role of the funder in responding to scientific controversies and the role of journals in responding to credible concerns about published manuscripts.

### Investigator Responsibility

First and foremost, investigators are responsible for the accuracy of their data, for the fairness of their conclusions, and for responding appropriately to criticism. Reproducibility—based on transparency—is a central component of the system of science. In the Duke story, there were not only inaccuracies in data, but also a lack of transparency by the investigators related to journals, to other investigators, and to the university's external review committee (discussed more below).

Second, investigators have responsibility that clinical studies being conducted have appropriate scientific justification and approval of relevant review bodies. At Duke, it appeared that in some instances, genomic predictors were being used for patient management in studies, while they simultaneously were being tested in other "preliminary" studies for their ability to predict results. This kind of problem arguably should have been apparent to—and avoided by—the PIs and their clinician colleagues.

The lead Duke investigators were not responsive to the queries of external investigators wanting to learn from and duplicate these methods on their own data, particularly after serious questions were raised in the medical literature. In addition, none of the coinvestigators in this series of publications originating from Duke raised concerns about the predictors. As reported by

Robert Califf in August 2011, Duke eventually surveyed 162 investigators involved in 40 papers co-authored by Potti, half of whom were by then at other institutions. Two thirds of these papers, he testified, will be partially or fully retracted, with others pending evaluation. Yet in no instance did anyone make any inquiries or call for retractions until contacted by Duke. This experience suggests the need for coauthors to have more shared responsibility for the integrity of the published research.

When the Duke leadership was interviewed by the IOM committee on August 22, 2011, they stated that it is essential to be able to trust the PI because no audit system can totally overcome a fundamental lack of trust. In retrospect, they said that PIs must develop an appropriate culture with an accountability plan that must have (1) trust, (2) a system where dissent is encouraged, (3) appropriate data management systems, and (4) appropriate biostatistics collaboration.

### Institutional Responsibility

Universities arguably have some of the most important responsibilities, as a "responsible party," for assuring the soundness of science. Universities evaluate and hire researchers; have high standards for faculty appointment, promotion, and tenure decisions; and the university's name is inevitably associated with the work of its faculty. Universities are responsible for establishing oversight structures, such as IRBs, COI management, and other review committees, and for providing "safe environments" for reporting irregularities, to help ensure soundness of science and protection of patient-participants. Last, universities are directly charged with being the "oversight" bodies when specific questions or challenges arise, for example, in investigating questions of misconduct, or simply in investigating instances of "soundness of science" as Duke University was asked to do by NCI in fall 2009. For scientists at companies or stand-alone research institutions, the same institutional responsibilities apply.

### Institutional Culture

Institutional "culture" includes expectations of behavior, achievement, and integrity that are transmitted by the institution and modeled by its leadership. Institutional culture starts with the dean, more senior leaders, and members of their team stating how research is to be conducted, with integrity and transparency, and with clarity that shortcuts will not be tolerated and that dishonesty is the basis for dismissal.

### Role of Institutional Structure

In his opening remarks to the IOM panel on August 22, 2011, Califf outlined the organizational context. He specified that the Statement of Task trials were conducted in the context of Duke University, under general supervision by its Board of Trustees, President, Vice President, and Provost. The Chancellor for Health Affairs is responsible for the Duke University Health System, the Duke University School of Medicine, and the Duke University School of Nursing (integrated as "Duke Medicine"). The Dean of the School of Medicine and several Vice Chancellors—for science and technology, for clinical research, and for global health—report to the Chancellor. There are 20 departments and about 13 major centers and institutes, whose chairs and directors report to the Dean of the School of Medicine. Directors of several campus-wide institutes, including the Institute for Genome Sciences and Policy (IGSP), report jointly to the

Provost and the Dean. Califf leads the Duke Translational Medicine Institute (DTMI), which has six major components, including the Duke Clinical Research Institute (DCRI). DCRI conducts multisite clinical research; it does not have accountability or authority for research done on patients in the Duke University Health System, unless part of a multisite trial. Dr. John Falletta is the Senior Chair of the Duke University Health System IRB and Dr. Michael Kelley chairs the Cancer Protocol Committee.

### Institute for Genomic Science and Policy

Duke University launched the IGSP in 2003 to bring together multidisciplinary teams of researchers. The Center for Applied Genomics and Technology, in which Nevins and Potti worked, was embedded within the IGSP. At the time of the three clinical trials, Duke had an extensive clinical trials infrastructure within the Duke Cancer Center, which normally would have been responsible for oversight and data stewardship in oncology trials conducted at Duke. These three oncology trials were not subject to Cancer Center oversight because of the collaboration with IGSP to manage the implementation of the genomic predictors. This led to IGSP staff becoming involved in data entry and data management.

The genomics work was permitted to operate outside the established structures for review and supervision of clinical research, such as the Duke Cancer Center or the DTMI. This was explained by the fact that work spanned both basic sciences, including research in animals or banked specimens, and clinical research in people. The consequence was that a "separate pathway" had been created within the university that ultimately did not provide the normal "checks and balances" in clinical research for storage of data, for blinding where appropriate, for providing locked-down protocols and plans for analysis, let alone an openness to critical reviews of protocols and publications. The IGSP created another separate infrastructure for gene expression–based clinical trials without the experience or expertise of the Cancer Center or the DTMI. It established the Duke Clinical Genomics Studies Unit (CGSU) to develop standards for genomics research. It also created a Data and Safety Monitoring Board "Plus" to monitor the safety of human subjects and the validity and integrity of data in ongoing genomic trials (Kornbluth and Dzau, 2011). However, this monitoring committee was not totally independent of the investigator team.

The IOM committee concluded that many of the problems that occurred would have been detected early or prevented entirely if routine structures and checks and balances had been in place or if the full infrastructure of the cancer center had been used. According to Califf, "there were numerous missed signals," any one of which might have helped to avert the problems that emerged. Moreover, "there was ambiguity" in the lines of authority and oversight in the IGSP during the conduct of the three clinical trials. As Califf stated, the IGSP was supposed to be consultative with other research groups within the health sciences, but things got jumbled up. Geoff Ginsburg, a leader in the IGSP, added that in 2005, the IGSP set about to create a core function, which was to provide expertise fundamental to conducting clinical research in genomics: for example, biospecimen procurement and quality assessment, sample management and processing, biobanking, and statistical analysis of genomic data. Ginsburg indicated that as time went on and the trials evolved, the sheer workload of the trials increased, and some of the genomics research coordinators began assisting in data entry and data management for these trials. The institution has since gained an appreciation of the need to clearly separate clinical trials activities from the genomics aspects of these studies and has created the systems to

maintain this clarity; thus, the IGSP's Clinical Genomics Studies Unit no longer participates in activities directly related to the performance of clinical trials. Instead, it consults, provides a resource, or collaborates with anyone doing genomics–based research, including investigators from other units who may be leading clinical trials.

### Duke Cancer Center

For most NCI-sponsored cancer centers, such as the Duke Cancer Center, there is a requirement for a highly structured process for conducting clinical trials. Components of this system include a CPRC (Clinical Protocol Review Committee) that provides a thorough scientific review regarding detailed scientific justification for the protocol, adequate trial design including sample size and an analysis plan, a clinical trials management system, and an independent data monitoring committee (IDMC) watching for early evidence of harm or benefit and quality of trial conduct. In this case, the Cancer Center CPRC reviewed the protocols for their scientific merit, apparently relying on the journals' decisions to accept the relevant scientific publications. Moreover, as noted above, the IGSP rather than the Cancer Center had taken responsibility for conducting and overseeing the trial. The protocols were reviewed by the Duke IRB to assess whether human safety, privacy, and autonomy were protected and appropriate informed consent procedures were in place. After these initial reviews, the trials were executed by the investigators and overseen by the IGSP (Kornbluth and Dzau, 2011). Ultimately, the IRB is responsible, but must rely on the CPRC for scientific review and the IDMC for expertise on the conduct of trials.

### Duke Response

The University did not institute extra oversight or launch formal investigations of the genomic trials during the first 3 years after the original publications triggered widely known controversy about the scientific claims and after concerns started to develop about the possible premature early initiation of clinical trials. *The Cancer Letter* began covering the story in 2009, and intense activity with NCI and the external community occurred during 2010. In 2010, the University formed a Translational Medicine Quality Framework (TMQF) committee to make recommendations to University leadership on appropriate oversight policies for future omics research being tested in clinical trials (TMQF Committee, 2011a,b). Their recommendations address lines of authority, oversight, and accountability.

### IDE Requirement

There was significant ambiguity about whether an IDE was required at the onset of the three clinical trials. FDA has an oversight role in late-stage research before a test (or drug) is applied to patients. At the time that the Duke investigators were developing their clinical trial protocols, FDA was in the process of clarifying the requirements for when an application for an IDE must be made for genomic predictors. The Duke IRB's understanding was that computational models were not considered devices by the FDA and, even if a computational model were a device, the omics-based tests were not a significant risk to the patient-participants' health, safety, and welfare because the tests were being used to direct choices among standard therapies (Falletta, 2011; FDA, 2011). However, the FDA sent a letter to the investigators in 2009 stating that the omics-based tests being studied in the three clinical trials named in the

statement of task needed to go through the IDE process (Chan, 2009). The investigators made some changes to the protocol of the studies in response to this letter and contacted the FDA for further clarification about whether an IDE was still required (FDA, 2011; Potti, 2009). When the FDA failed to respond to this letter[3], the Duke IRB determined that an IDE was not needed (FDA, 2011). In retrospect, and with FDA guidance in 2010, the Duke IRB chair recognized that an IDE should have been obtained for the omics-based tests used in the trials because the tests were used to direct patient management in the clinical trials (Falletta, 2011; FDA, 2011).

If an IDE had been sought at the initial stages, the FDA process of review would likely have asked for validation data and a locked-down data file and algorithm. While it cannot be known, it is likely that an FDA review would have uncovered some of the validation issues that would have prevented the use of these predictors in clinical trials. Duke has since instituted new controls to ensure compliance with FDA's IDE requirement (Califf, 2011).

## Conflict of Interest Management

The Duke COI committee was charged with overseeing and managing investigators' COIs while the three clinical trials were being conducted. The Duke IRB routinely asks investigators to answer questions about their COI when submitting protocols for review (Ginsburg, 2011). If the answers trigger concerns about COI, the IRB notifies the COI committee. The COI committee also requires all investigators to complete an annual reporting form intended to identify financial COIs. If a COI is identified through either of these channels, the COI committee works with the individuals to create management plans to address their COI. McKinney informed the IOM committee that individual investigators could be required to disclose their COI on all consent forms for a clinical trial or be prohibited from acting as PIs in a clinical trial in which there had a substantial financial stake (McKinney, 2011).

The IOM committee reviewed information regarding the potential for COI for investigators, members of IRBs, the Data Safety Monitoring Board (called the DSMB-plus), and other oversight bodies for the trials at Duke. There is evidence that some of those involved in the design, conduct, analysis, and reporting of the three clinical trials and related trials involving Duke genomic predictors had either financial or intellectual/professional conflicts of interest that were not disclosed. Specifically, some investigators involved in the three clinical trials were evaluating omics-based tests for which they held a patent, or had a financial relationship with Expression Analysis Inc. and/or CancerGuideDx Inc.,[4] laboratory and bioinformatics companies that were established quite early in the process to market the Duke omics-based tests. According to Califf, there was great deal of confusion within the University at this time about when a patent and intellectual property interest qualified as a conflict. Some investigators believed a conflict developed when a patent application was filed or a patent was issued; others believed it was when a relationship was formed with a commercial company or when a marketable product was produced (Califf, 2011). At the meeting with the IOM committee on August 22, 2011, Califf acknowledged that the COI process had not identified that there was an important COI of a member of the DSMB-plus for the three clinical trials, which resulted from the member's

[3] According to the FDA website, CDER has no record of receiving the Dec. 2009 letter from Dr. Anil Potti discussing an IND-exemption for the trial that received pre-IDE review. The letter was brought to FDA's attention during their 2011 inspection of the Duke IRB and clinical investigators (see http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm289100.htm).
[4] NOTE: This company no longer exists.

previous substantive collaboration with some lead investigators for these three trials on research closely related to that in those trials.

In addition to individual COI, the potential for institutional COIs is important. Such COI not only is financial (i.e., the universities get a portion of profits from patent licenses and from spin-off companies), but also arises from interest to protect the reputations of the institution and respected colleagues. At the IOM committee meeting in March 2011, Califf acknowledged the university's concern for its reputation in its handling of controversial issues of all kinds and the need to be especially careful in assessing the work of an esteemed senior faculty member. Managing situations in which both an investigator and the institution have potential COIs is particularly challenging.

Some investigators at Duke had intellectual property (IP) and equity interests in Expression Analysis and CancerGuideDx. Duke had no institutional interest in Expression Analysis, but did have a license agreement in place with CancerGuideDx as of January 2010; license negotiations began in early 2009. Duke's vice dean for research indicated she was not aware of these conflicts at the time, and that better communication is needed because all parties should be aware of such issues (Kornbluth, 2011).

Duke leadership indicated that, whenever IP is filed, the institution, the COI committees, and the IRBs should be informed, but this was not routine procedure during the time of the design and conduct of the three clinical trials. Duke leadership acknowledged the possibility that COIs would have been considered to be private information, so that people working in the same groups or as coauthors would not necessarily know about some potentially important conflicts of interests of their colleagues. It was also reported that there was a lack of an institutional process that provided insight about how to address conflicts across the continuum of IP generation and development, from planning to file IP, filing IP, and forming relationships with a company. However, the senior chair for the Duke IRB process confirmed that PIs now are expected to disclose IP in their IRB submissions (Falletta, 2011).

### Whistleblowing

Duke University follows a "just culture" model, which does not hold individuals accountable for system failings and errors over which they have no control; it only holds individuals accountable for mistakes that disregard patient safety or involve gross misconduct (Zuiker, 2008). Under a just culture model, it is expected that anyone at any level can criticize the scientific methods of a study in a protected environment. At the time of the three clinical trials, Duke University used both anonymous and non-anonymous reporting systems. It also had a compliance hotline through which individuals could report breaches of the rules and regulations governing clinical research (Cuffe, 2011). However, the problems with the three clinical trials were not brought to the attention of the appropriate individuals within the university leadership through any of these whistleblowing channels. According to Vice Dean for Research Sally Kornbluth, a number of people came forward after the university undertook its investigation and said they "were glad [the university was] reviewing things carefully" (Kornbluth, 2011). Why no one came forward earlier, or perhaps any such concern was not forwarded appropriately, is not known, but the fact that these problems were not brought forward earlier may be an indication of the discomfort or lack of confidence that faculty and staff may have with these systems.

Duke has taken steps to improve its whistleblowing system in response to what occurred in this case. The TMQF plan requires that every site-based research unit's accountability plan must include a strategy to encourage individuals to discuss concerns about methods of research and to report suspected breaches in appropriate research practices. The goal is to create a culture of "dissent and discussion" (Califf, 2011). The university has made its Post-Doctoral Office more robust with the intent to make it a place within the university where postdoctoral students can report their concerns. It has also added an ombudsperson for faculty and students (i.e., an individual whom faculty and students can approach with concerns, if they feel uncomfortable confronting a superior) (Kornbluth, 2011). Everyone acknowledges that raising such challenges leads to anxiety.

### Responding to Scientific Controversies

University leadership originally believed that the controversy surrounding the use of the omics-based tests in the three clinical trials involved disagreement about arcane scientific methodology. According to Kornbluth, "it was not presented or recognized as a criticism or implication of underlying data corruption" (Kornbluth, 2011). This position changed in 2009 when Baggerly and Coombes published the article in the *Annals of Applied Statistics* (Baggerly and Coombes, 2009), which stated that the omics-based tests did not work and were potentially endangering patient safety by incorrectly directing therapy (Kornbluth and Dzau, 2011). The potential for the predictors to incorrectly direct therapy existed, among other reasons, because discrepancies identified in the data included reversal of some sensitive/resistant labels.

At this time, NCI also had begun the process of reviewing protocol CALGB-30702 that had been submitted to NCI's CTEP. The proposed trial would have used six of the chemosensitivity predictors to guide therapy in an advanced lung cancer trial. The NCI reviewers noticed substantial differences between the trial protocol's descriptions of the test and the way in which the tests were described in the validation studies. NCI disapproved the CALGB-30702 protocol, but the protocol mentioned that several genomic predictors were already guiding therapy in some Duke University trials. NCI staff conducted a search of ClinicalTrials.gov and became concerned when several clinical trials at Duke were identified using omics-based tests with similar methodologies that were developed by the Nevins/Potti group. NCI contacted Duke in September 2009 regarding these trials (McShane, 2010a).

In response, Duke suspended the three trials and the Duke IRB initiated an investigation.[5] The IRB was designated as the appropriate university entity to conduct the investigation because the focus was on patient-participant safety concerns (Kornbluth, 2011). The Duke IRB formed an external peer review committee composed of two statisticians to conduct an independent evaluation of the data and three clinical trials. The reviewers' identities were protected under a confidentiality agreement to encourage the reviewers to act objectively and without fear of reprisal (TMQF Committee, 2011b). The original intent was to give reviewers "unfettered access to all the data, software, and analyses. They also could request any other information needed from Nevins and Potti" (Kornbluth and Dzau, 2011, p. 16). The charge to the external reviewers consisted of two questions:

[5] The Office of Research Integrity's policy on responding to scientific controversies is limited to addressing scientific misconduct.

1. "Have the methodology errors originally communicated by the MD Anderson Cancer Center researchers, Baggerly and Coombes, been adequately addressed by the Duke researchers?
2. Do the methods as originally developed and as applied in the context of these trials remain valid?" (Kornbluth and Dzau, 2011).

In concluding their review, the external statisticians stated they were "able to show with an independent analysis that the approaches used in the Duke clinical predictors are viable and likely to succeed" (*Review of genomic predictors for clinical trials from Nevins, Potti, and Barry,* 2009, p. 1). On the basis of this report, the university resumed the three trials (Kornbluth and Dzau, 2011).

However, several revelations raise substantive concerns about this review process. The external statistical reviewers explicitly noted, given the data they had to review, that they were "unable to identify a place where the statistical methods were described in sufficient detail to independently replicate the findings of the papers." They stated, "The one area [in which] they [the Duke investigators] have not been fully responsive and really need to do so is in clearly explaining and laying out [sic] the specific statistical steps used in developing the predictors and the prospective sample assignments" (*Review of genomic predictors for clinical trials from Nevins, Potti, and Barry,* 2009).

The integrity of the external review may also have been influenced by the involvement of Nevins. In the 12/22/2009 report of the external statistical reviewers, a reference is made to the pemetrexed predictor, "In addition, we agree with Nevins and Potti that since the profile is *not used in any of the clinical trials* patients are not being endangered." This statement contradicts the fact that the pemetrexed predictor was being used to guide the choice of treatment in the Statement of Task trial NCT00545948 that opened in October 2007 even though, as noted previously, Potti had stated in his submission of R01-CA131049-01A1 in March 2008 that the accuracy of the pemetrexed predictor had not been validated in independent patient samples. This suggests that the independent review process permitted the PIs to be in direct contact with the external independent statistical reviewers, allowing the PIs to provide misleading information to these reviewers.

The external committee's review also was influenced by lack of access to important relevant information. In the first week of November 2009, while the external review was in progress, new data about the cisplatin and pemetrexed predictors (the subject of Hsu et al., 2007) were posted to a Duke website. Baggerly examined the new data and found additional errors, noting in particular that all of the samples used for validation were mislabeled. He forwarded a report and raw data to Duke officials on 11/9/09. That material, however, was never forwarded to the external statistical reviewers due to the university leadership's concerns that it might "bias" the committee's review (Kornbluth and Dzau, 2011).

In summary, concerns have emerged about three areas: (1) whether the external IRB reviewers were encouraged to delve deeply because 4-6 weeks of intensive work likely would have been needed to do so, (2) the comprehensiveness of the information provided to the external statistical reviews, and (3) whether the information provided to the reviewers was substantially influenced by Nevins and Potti. These issues speak to the challenge of responsible parties' oversight, resulting from institutional "conflicts" that could compromise the integrity of such oversight. In retrospect, Duke's leadership recognizes that if the reviewers "had been explicitly

sent to McShane or sent to Baggerly and [instructed to consult with them during their analysis], there would have been a different outcome" (Kornbluth, 2011). According to Kornbluth, "One of the chief lessons learned is there's a balance between trusting investigators [who] have a very long track record with an institution and also thinking about what is necessary to ensure an adequate review" (Kornbluth, 2011).

### Biostatistical Collaboration and Data Provenance Issues

As mentioned above, the IGSP was set up to conduct multidisciplinary research and included staff with diverse expertise, such as biostatistics, bioinformatics, clinical trials, pathology, and laboratory science. Various individuals with biostatistical expertise were involved in the development of the omics-based tests used in the three clinical trials, but there was not continuity in personnel. Numerous errors identified in the statistical methodology and analyses (Baggerly and Coombs, 2009; McShane, 2010a,b) suggest there was insufficient statistical expertise involved in the studies for which published papers have now been retracted. The Duke TMQF committee recommends involving biostatistical expertise in all translational research projects (TMQF Committee, 2011b) and recognized the need to increase education and training of statisticians (TMQF Committee, 2011a). These experiences at Duke also motivate the IOM committee to emphasize the importance of involving senior-level biostatisticians who are coinvestigators and co-owners of responsibility, and who are intellectually independent, preferably reporting to an independent mentor or department chair.

Duke also attributes the lack of "sustained statistical collaboration" as a contributing factor to the research team's failure to follow proper data management practices (Kornbluth and Dzau, 2011). In its discussion of data provenance, the TMQF committee recognizes the importance to the integrity of confirmatory trials of maintaining confidentiality of interim data, stating, "secure database management systems are used to store and interrogate data for quality assurance; persons with vested interests (such as clinical investigators) are blinded to, and independent from, data and analyses" (TMQF Committee, 2011b). This principle for having secure databases with a firewall between the interim data and the trial investigators is, for many reasons, of key importance to the integrity of confirmatory trials, including ensuring that the main hypotheses being addressed by the trial are not influenced by the data from the trial (see Chapters 2, 3, and 4).

There are indications that clinical databases for the three clinical trials were not adequately secure, in contrast to the principles stated above by the later Duke TMQF document. McShane stated that she had been contacted by someone who raised allegations that there were problems with how the data were being handled in some of the prospective Duke trials. In the 8/22/11 meeting with the IOM committee, Califf indicated some of the data from those trials could have been accessed some of the time, that the endpoint information was not going into secure databases. Thus, in essence, the investigators could know case by case what the outcomes were, and they could have reconstructed the data. He also agreed that it is possible that such access to emerging data could potentially lead to inappropriate actions such as reformulation of hypotheses.

## THE ROLE OF THE FUNDER IN RESPONDING TO SCIENTIFIC CONTROVERSIES[6]

NCI's involvement was limited in Duke's external peer review of the three clinical trials. It provided Duke with the names of statisticians who could potentially serve as peer reviewers and assisted Duke with the initial contact to the reviewers. However, according to McShane, "due to the incomplete information in the [Duke external reviewers'] report and the fact that NCI had no access to the data provided to the reviewers, NCI could not make a judgment on whether the concerns about the [tests] used in the Duke trials had been adequately addressed" (McShane, 2010a).

When NCI staff determined in April 2010 that NCI was providing partial funding through an R01 grant for the Duke trial using the cisplatin predictor (NCT00509366), it requested the data and computer code for the cisplatin *and* pemetrexed predictors that were the basis for the primary aims of the grant work (McShane, 2010a). The investigators then provided NCI with the data and computer code for the cisplatin test, but not for the pemetrexed test. NCI staff evaluated the cisplatin test and were unable to reproduce the results. The analyses by both the NCI and Duke's external reviewers relied on the data provided by the Duke investigators. NCI did not believe the external review was adequate. On June 29, 2010, NCI met with the Duke investigators and the Duke leadership. NCI asked Duke to produce the original raw data that would reproduce the findings in the papers. On October 22, 2010, Duke notified NCI that multiple validation datasets associated with the cisplatin predictor were corrupted. For example, William Barry explained in his August 22, 2011, testimony to the IOM committee that when he found the original source data for the ovarian cancer cell line drug sensitivity experiments, he was able to determine that the drug sensitivity measurements supplied to NCI for their evaluation of the cisplatin predictor differed from the true source data. The data corruption had the effect that the sensitivity predictions produced by the genomic test showed a significant association with the incorrect sensitivity measures, but the association disappeared completely when the correct sensitivity data were used. Thus, the PIs and the Duke leadership agreed to terminate the trials (Kornbluth and Dzau, 2011). The process was also initiated to retract the paper by Hsu et al. that had been published in 2007 in the *Journal of Clinical Oncology.* This retraction (Hsu et al., 2010) was the first of several retractions from the investigators.

Nevins reported during his March 30, 2011, testimony that findings of corruption had been observed for multiple datasets compiled by his team for purposes of validating the various chemotherapy sensitivity signatures (Kornbluth and Dzau, 2011). These included data derived not only from Duke sources, but also publicly available data. As an example, a dataset of 133 samples from a neoadjuvant breast cancer study at MD Anderson involving patients treated with the combined regimen TFAC was used for validation of an Adriamycin signature. The clinical annotation that was assumed to be used by Potti et al. included 34 responders and 99 non-responders, the same distribution as reported by MD Anderson. However, a detailed comparison of the two datasets revealed that the response information was reversed for 24 cases, with 12 labeled incorrectly in each direction. In this case, the corrupted data yielded positive validation

[6] The Department of Defense (DoD) Breast Cancer Program funded the NCT00636441 trial and Eli Lilly was a sponsor of the NCT00509366 and NCT00545948. However, the committee did not interact with DoD or Eli Lilly and does not have information on any steps they took to investigate the scientific controversies surrounding the trials.

results whereas the accurate data did not provide evidence for validation. Similar findings of data corruption in key validation datasets were observed in other instances.

There was a lack of substantive interaction between Duke and NCI about details of the charge to the external review committee and about details of the conduct of the investigation (e.g., regarding what material the committee had access to). Duke did not ask for any detailed help or comment from the NCI, and the NCI seemed to think it was not appropriate to try to provide specific direction unless it was "invited" by the university. The IOM committee recommended that federal funders of omics-based translational research should have authority to exercise the option of investigating any research being conducted by a funding recipient after requesting an investigation by the institution. For example, NCI might consider, in the future, having a more active supervisory role, regarding adequacy of the "charge" or of the work of an independent review committee. Indeed, Dr. Kornbluth stated that Duke leadership wished they had understood early on that part of NCI's concern stemmed from their inability to reproduce the exact published data. This became clear when she accompanied Drs. Nevins and Potti and colleagues to visit the NCI after the external review (Kornbluth, 2011).

In this case, NCI, through McShane, who had invested many months of time in pursuing the process and statistical issues, had great insight into the problems independent of the Baggerly and Coombs efforts. But NCI did not pursue this until later when the Duke external review appeared inadequate and after NCI determined that it was supplying partial funding for one of the three Duke clinical trials through an R01 grant to Potti. It was apparently only after discovering the funding tie through the grant that NCI believed it was justified in taking a more active role in the investigation and in requesting data and computer code to evaluate the predictors. When NCI asked for data and computer code for both the cisplatin and pemetrexed predictors that were being studied in the grant, the Duke investigators declined to provide the necessary data and computer code for the pemetrexed predictor on grounds that it was not being used in the trial linked to the grant. However, NCI funding supported much of the relevant foundational research for the trials. Perhaps a more active role by NCI earlier on might have avoided some mistakes in the external review process. By June 2010, however, Duke officials came to NCI and McShane laid out the issues in great detail. As a result, Kornbluth said Duke leadership clearly understood the seriousness of the concerns. At the August 2011 meeting with the IOM committee, both Califf and Kornbluth clearly stated that they viewed McShane's work as critical. Moving forward, it is important to ensure that universities (presuming that universities continue to be the major "responsibility party" for this kind of work) can get the detailed expertise and advice necessary to conduct a proper "evaluation process." Kornbluth suggested during the August 22, 2011, meeting that for some problems a university might want to obtain "outside help," perhaps a sister university in a consortium, either because of lack of expertise in-house or because of institutional conflict.

## THE ROLE OF JOURNALS IN RESPONDING TO CREDIBLE CONCERNS ABOUT PUBLISHED MANUSCRIPTS

As described earlier in this appendix, Baggerly and Coombes pursued correspondence with the authors of the *Nature Medicine* paper, both directly and through letters to the editor. Baggerly and colleagues pursued communication directly with Nevins and colleagues from November 2006 to June 2007. Shortly after, communications between the groups broke off. In June 2007, Baggerly and Coombes submitted correspondence to *Nature Medicine* outlining their

unresolved concerns and questions about the omics-based tests (Baggerly, 2011). In November 2007, the correspondence was published along with a reply (Coombes et al., 2007; Potti and Nevins, 2007). As noted earlier in the section "Development and Evaluation Process," the letter from Coombes et al. expressed five major concerns about the paper, including: (1) their inability to reproduce the selection of cell lines from the sensitivity measures, (2) errors in the gene lists, (3) incorrect figures, (4) the use of combined training and test sets in the development process, and (5) their inability to reproduce the reported test performance results. In their reply, Potti and Nevins acknowledged some errors in the posted data, figures, and gene lists. However, they countered that Baggerly and Coombes had used different analytic methods and they disagreed with Baggerly and Coombes' objection to combining the training and test sets to develop the computational model. Furthermore, the authors stated that they had been able to successfully validate the tests in independent datasets, as reported in the papers by Hsu et al. (2007) and Bonnefoi et al. (2007). Two corrections to the Potti et al. *Nature Medicine* paper were published in November 2007 and August 2008, and the authors indicated that corrections had been made to the supplementary information posted online (Potti et al., 2007a, 2008). The Duke investigators said that, with only a few exceptions, the errors in posted data, figures, and gene lists were clerical errors that had no impact on the actual tests developed or the reported test performance results.

Baggerly and Coombes also corresponded with the authors and journal editors regarding the papers by Dressman et al. and Hsu et al., published in the *Journal of Clinical Oncology* (JCO). JCO published their letter and a reply from the authors regarding the Dressman et al. article (Baggerly et al., 2008; Dressman et al., 2008, but declined to publish the letter regarding the Hsu et al. article. According to Baggerly, when he tried to correspond with Potti et al. regarding the Bonnefoi et al. article published in the *Lancet Oncology* (Bonnefoi et al., 2007), Potti was no longer willing to engage in a discussion. The *Lancet Oncology* rejected their letter (Baggerly, 2011). Meanwhile, the papers were used and cited by hundreds of other investigators.[7]

Ultimately, the *Nature Medicine* paper was retracted on January 7, 2011, based on the NCI's recommendation for a full review of all data associated with all of the predictors in the key papers that had been questioned or that had been used in clinical trials. The retraction cites the corruption of the validation datasets and explicitly states that the authors were "unable to reproduce certain crucial experiments showing validation of signatures for predicting response to chemotherapies, including docetaxel and topotecan" (Potti et al., 2011a, p. 135). The papers by Hsu et al. (2007) and Bonnefoi et al. (2007) were retracted in November 2010 and February 2011, respectively. The Potti et al. (2006b) *New England Journal of Medicine* paper likewise was retracted in March 2011. The Dressman et al. (2007) *Journal of Clinical Oncology* paper was retracted in January 2012 (JCO, 2012). In addition, Duke leadership has identified 40 papers in which Potti was a coauthor and the study involved original data analysis. Duke has contacted all 162 of his coauthors and asked whether they support the veracity of their work. Based on this dialogue, two thirds of the papers are being partially or fully retracted; one third were still considered valid by Duke leadership as of August 2011 (Califf, 2011).

[7] The Potti et al. (2006a) article was cited 306 times, the Hsu et al. (2007) article 60 times, the Dressman et al. (2008) article 111 times, the Bonnefoi et al. (2007) article 95 times, and the Potti et al. (2006b) article 350 times in Scopus (all as of October 28, 2011).

## USE OF CHEMOSENSITIVITY TESTS AT OTHER INSTITUTIONS

One of the motivations for encouraging transparency and open scientific discourse is that scientific progress is built on the foundations of past work. The retracted papers were cited dozens or hundreds of times before they were retracted, and many grants were awarded based on such work. Thus, the committee sought information on whether clinical trials had been initiated at other institutions, based on the now-retracted work from Duke University investigators. The committee's concerns focused on the following questions:

- Have genomic signatures either developed or validated on the Nevins/Potti data been used for patient management decisions in clinical trials other than those named in this committee's statement of task?
- Are investigators at other universities or cancer centers involved in the design and conduct of trials using genomic signatures linked to the work of Nevins and Potti and colleagues?
- If so, who sponsored these trials? Are these sponsors fully informed about possible integrity issues with the genomics-based tests used in these trials?
- Have NCI and other sponsors conducted appropriately comprehensive investigations into any suspected integrity issues?

Through a search on ClinicalTrials.gov and on the NIH RePORTER, the committee identified a clinical trial at the Moffitt Cancer Center & Research Institute (MCC), NCT00720096, "A pilot prospective trial of genomic directed salvage chemotherapy with either liposomal doxorubicin or topotecan in recurrent or persistent ovarian cancer within 12 months of platinum-based chemotherapy" (ClinicalTrials.gov, 2011d) and an NIH R33 grant, 5R33CA110499-05, "Molecular profiling to predict response to chemotherapy" (Lancaster, 2008). According to its ClinicalTrials.gov entry, the trial was initiated in July 2008 and was terminated in October 2009, with four patients accrued. The R33 grant turned out to be a continuation of the grant 1R21CA110499-01A2, acknowledged in the 2007 *Journal of Clinical Oncology* paper authored by Dressman et al. (2007).

The committee requested copies of the MCC clinical trial protocol and informed consent documents (H. Lee Moffitt Cancer Center & Research Institute, 2007, 2008a,b, 2009), and sent a letter to Moffitt Cancer Center & Research Institute Director William Dalton requesting further information about the trial and grant. Moffitt provided copies of the protocol and informed consent documents. Dalton responded to the letter from the committee,[8] providing important additional insights into the history of the trial and grant. The letter also described some interactions with NCI, which prompted a response to the IOM committee from NCI in the form of a letter from NCI statistician McShane.[9]

The Moffitt ovarian cancer clinical trial was assessing omics-based tests developed to predict sensitivity to liposomal doxorubicin and to topotecan. Treatment in the trial was to be

[8] Personal communication from William Dalton, Moffitt Cancer Center & Research Institute, to Gilbert S. Omenn, University of Michigan, RE: Response to questions - Genomic-directed salvage chemotherapy with either liposomal doxorubicin or topotecan, September 28, 2011.

[9] Personal communication from Lisa McShane, National Cancer Institute, to Gilbert S. Omenn. University of Michigan, RE: Moffitt response to questions on the trial *Genomics-directed salvage chemotherapy with either liposomal doxorubicin or topotecan*, October 23, 2011.

directed by the genomic predictors. Regarding the source of the predictors, the protocol states, "The predictive models for Doxil [liposomal doxorubicin] and topotecan as defined in our previous work [Potti et al., 2006a] will be implemented to assess the predictive response of a clinical trial sample." Sponsors of the trial were Moffitt and DoD, and the PI of the trial was MCC's Robert M. Wenham. Jonathan Lancaster, currently an MCC investigator, was identified by Dalton as a coinvestigator on the trial. Lancaster was formerly at Duke University and a coauthor with Nevins and Potti on three of the papers retracted to date: the 2006 *Nature Medicine* paper and two 2007 *Journal of Clinical Oncology* papers (Dressmann et al., 2007; Hsu et al., 2007; Potti et al., 2006a).

Dalton's letter, which he describes as representing the inputs of Wenham and Lancaster, states that the predictors were developed at Moffitt as part of the NCI-funded R21 grant work and were undergoing prospective validation in the R33 grant, where they were not being used to guide therapy in that NCI-funded grant. The Moffitt trial, which was using the predictors to guide therapy, was running concurrently with the R33 validation study. McShane's response indicates that the NCI did not view the predictors as being ready for use in directing therapy, whereas Moffitt viewed it as acceptable for the predictors to be used to guide therapy in the context of a feasibility trial. The Moffitt trial had undergone institutional scientific and IRB review as well as DoD review where it was deemed acceptable, just as in the case of the Duke predictors and trials. This experience may point to inconsistencies in standards required by different funding agencies and institutions for when an omics predictor is ready for use in directing therapy, or it may reflect insufficient information provided to the funding agency. The committee makes several recommendations (in Chapter 5 on Responsible Parties) to guide these determinations and promote more consistent standards to protect patients and avoid waste of resources that could result when predictors are put into clinical studies prematurely.

NCI stated that it had concerns about the information initially presented on the predictors during the R21 grant transition review. Problems with clarity and consistency of information presented about the omics-based tests were identified by NCI transition reviewers, and NCI questioned whether the predictors were appropriately locked down and ready for validation in the R33. This caused NCI to conduct a more extensive review than usual, involving direct interaction in early 2008 between NCI statistician McShane and Moffitt statistician Steven Eschrich and the provision of some example data and computer code to NCI to allow NCI to assure lockdown of one of the five predictors reported to have been developed in the R21. However, the predictor examined was not one of those used in the Moffitt trial. NCI states that it was not aware until October 2009 of the Moffitt trial running concurrently with the R33 grant. Whether the trial cosponsor DoD or Moffitt's review bodies were aware of the concerns NCI had about readiness of the predictors for use in directing therapy at the time they approved the trial is unknown. When NCI did find out about the Moffitt trial, it contacted Lancaster with its concerns, and the Moffitt trial was terminated shortly afterward. Termination of the Moffitt trial was also mentioned by the October 23, 2009, *Cancer Letter*, where it was reported that a Moffitt spokesperson indicated the closure was unrelated to the controversy concerning the three Duke University clinical trials named in the statement of task (Goldberg, 2009b). This suggests that DoD and the Moffitt IRB might not have been aware of NCI's concerns even when the trial was terminated. The IOM committee's recommendations encourage better communications among funders so that important information about research and omics predictors developed in the

course of that research is shared to more fully inform decisions about the readiness of predictors for use in clinical trials or clinical care.

Ambiguity remains on exactly what predictors were used in the Moffitt trial and whether IDEs had been obtained. The Dalton letter states that the predictors used in the trial were derived at Moffitt and not at Duke. NCI expressed its uncertainty about the source of the predictors because the Moffitt trial protocol identifies Potti et al. (2006a) as that source. The answer to this question is important because the retraction notice for the Potti et al. (2006a) paper specifically states an inability to reproduce the validation results for the topotecan predictor as one of the reasons for the retraction. FDA oversight, such as through an IDE review process, might avoid some of these types of confusion and ensure that predictors are locked down and identifiable. Understanding of CLIA and FDA requirements has evolved over recent years, and the recommendations from the IOM committee should be helpful in promoting better understanding and appreciation for these federal regulations. The Dalton letter supports the committee's emerging recommendation that institutions seek FDA guidance and meet FDA IDE requirements.

The events that occurred at Moffitt support the notion that many of the problems identified at Duke are probably not unique to Duke. Learning about these multiple situations informed the development of the IOM committee's recommendations.

## CONCLUDING REMARKS

A time line summarizing events related to the material in this appendix is presented in Table B-2.

The committee identified several overarching themes in design, conduct, and oversight of omics research from the Duke case study. Among these themes, transparency and open communication remain important principles of the conduct of science, whether reporting data and code, disclosing conflicts of interest, or reporting potential breaches in scientific procedures. Institutions play an important role in establishing a culture that includes expectations of behavior, achievement, and integrity, and providing safe environments for reporting irregularities. Oversight processes that will maintain integrity even in the presence of institutional conflicts of interest may be especially important in achieving this goal.

Regarding development of gene expression–based chemosensitivity tests, validation requires steps to lock down important features, including hypotheses, predictors, and analysis plan. In order to protect patients from harm due to use of a faulty predictor, it is essential to follow the kind of scheme presented in Figure S-1 to confirm and then validate biomarkers and classifiers before launching clinical trials or offering them commercially for clinical use. It is important to involve appropriate expertise of biostatisticians and bioinformatics scientists in design, analysis, and oversight. It is also important for all members of a research team to understand the aims and many details of the collaborative study and for coauthors of a publication to keep each other informed about constructive criticism of the work and ways to improve the publications and ongoing research. Even before a predictor is considered for use to direct patient care, these good science practices should be followed to avoid wasted effort and resources. These themes are reflected in the detailed recommendations presented in Chapters 2-5.

While further pursuit of the questions raised about the clinical trials and omics-based tests discussed in this appendix may be undertaken separately from the work of the IOM committee, the need for availability of data and code, the need to follow a rigorous test

development and evaluation process prior to use of an omics-based test in clinical trials, and the responsibilities of investigators, institutions, journals, and funding agencies are clear lessons.

**TABLE B-2** Time Line of Events Surrounding the Duke Gene Expression–Based Tests

| Date | Event |
|---|---|
| 2000 | Computational and Applied Genomics Program (CAGP) at Duke University founded by Joseph Nevins and Mike West (Kornbluth and Dzau, 2011). |
| 2003 | Creation of Duke University Institute for Genome Sciences and Policy (IGSP) CAGP becomes the new IGSP Center for Applied Genomics and Technology (CAGT) (Kornbluth and Dzau, 2011). |
| | Anil Potti completes medical residency in North Dakota and begins fellowship at Duke in lab of Thomas Ortel; in 2004, he joins Nevins laboratory (Kornbluth and Dzau, 2011). |
| 2006 | Potti hired by CAGT to establish an independent lab focused on gene expression–based research (Kornbluth and Dzau, 2011). |
| August 2006 | *New England Journal of Medicine* (*NEJM*) publishes of "A genomic strategy to refine prognosis in early-stage non-small cell lung cancer" (Potti et al., 2006b). |
| October 2006 | *Nature Medicine* publishes online "Genomic signatures to guide the use of chemotherapeutics" (Potti et al., 2006a). |
| November 2006 | Keith Baggerly and colleagues begin correspondence about the *Nature Medicine* and subsequent publications with Potti and colleagues (Baggerly, 2011). Communication continues through June 2007. |
| 2007 | Clinical Genomics Studies Unit established at Duke University (Kornbluth and Dzau, 2011). |
| January 2007 | Letters to the editor and author reply related to the *NEJM* paper (Potti et al., 2006b) published (Larsen et al., 2007; Potti et al., 2007b; Singh and Dhindsa, 2007; Sun and Yang, 2007). |
| | Correction to Potti et al. (2006b) published in *NEJM* (Correction, 2007). |
| February 2007 | *Journal of Clinical Oncology* publishes "An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer" (Dressman et al., 2007). |
| April 2007 | William Barry joins IGSP (Kornbluth and Dzau, 2011). |
| July 2007 | Study Using a Genomic Predictor of Platinum Resistance to Guide Therapy in Stage IIIB/IV Non-Small Cell Lung Cancer (TOP0602) entered on ClinicalTrials.gov (Identifier NCT00509366). |
| October 2007 | *Journal of Clinical Oncology* publishes "Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer" (Hsu et al., 2007). |
| | Adjuvant Cisplatin With Either Genomic-Guided Vinorelbine or Pemetrexed for Early Stage Non-Small Cell Lung Cancer (TOP0703) entered on ClinicalTrials.gov (Identifier NCT00545948). |
| November 2007 | Publication of letter by Coombes, Wang, and Baggerly in *Nature Medicine* critiquing (Potti et al., 2006a), together with a rebuttal (Coombes et al., 2007; Potti and Nevins, 2007). |
| | Baggerly et al. submit letter, "Pharmacogenomic strategies may not provide a rational approach to the treatment of cisplatin-resistant patients with advanced lung cancer," to *Journal* |

| | |
|---|---|
| | *of Clinical Oncology*. It is rejected (Baggerly, 2011). |
| December 2007 | *Lancet Oncology* publishes "Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: A substudy of the EORTC 10994/BIG 00-01 clinical trial" (Bonnefoi et al., 2007). |
| March 2008 | Trial to Evaluate Genomic Expression Profiles to Direct Preoperative Chemotherapy in Early Stage Breast Cancer entered on ClinicalTrials.gov (Identifier NCT00636441). |
| | Potti et al. submit revised R01 grant proposal, "Prospective Validation of Genomic Signatures of Chemosensitivity in NSCLC" (CA131049-01A1), which is linked to a Phase II trial using the cisplatin chemosensitivity predictor to direct therapy for advanced-stage lung cancer patients. The trial was later identified as NCT00509366, which began enrolling patients in June 2007 (McShane, 2010b). |
| | Publication of letter to the editor by Baggerly, Coombes, and Neeley, "Run batch effects potentially compromise the usefulness of genomic signatures for ovarian cancer" (Baggerly et al., 2008), a comment on Dressman et al. (2007), and an author reply in the *Journal of Clinical Oncology* (Dressman et al., 2008). |
| May 2008 | Baggerly and Coombes submit a letter to the editor of *Nature Medicine,* "Microarrays: Retracing Steps (Again)"[a] |
| | Baggerly and Coombes submit a letter to the editor of *Lancet Oncology*, "Have gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy been validated?" (Baggerly, 2011). |
| June 2008 | *Nature Medicine* requests that Baggerly and Coombes 5/08 letter be sent to Potti and coauthors.[b] |
| | *Nature Medicine* rejects letter.[c] *Lancet Oncology* rejects letter.[d] |
| July 2008 | Genomic Directed Salvage Chemotherapy With Either Liposomal Doxorubicin or Topotecan entered on ClinicalTrials.gov (Identifier NCT00720096). |
| July 2009 | Cancer and Leukemia Group B (CALGB) submits revised CALGB-30702 protocol (Genome-Guided Chemotherapy for Untreated and Treated Advanced Stage Non-Small Cell Lung Cancer: A Limited Institution, Randomized Phase II Study).[e] |
| | *Current Oncology Reports* publishes "Translating genomics into clinical practice: Applications in lung cancer" (Jolly Graham and Potti, 2009). |
| September 2009 | *Annals of Applied Statistics* publishes online: "Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology" (Baggerly and Coombes, 2009). |
| | The National Cancer Institute (NCI) contacts Duke to ask that the university carefully consider the validity of the work and its extrapolation to clinic (McShane, 2010a). |
| October 2009 | 10/2 – *The Cancer Letter* first covers the story; Nevins asserts that the approach has been shown to work in a blinded validation by Bonnefoi et al. (2007) (Goldberg, 2009a). |
| | The Data Safety Monitoring Board and Duke Cancer Protocol Review Committee conclude that issues raised by Baggerly and Coombes (2009) presented no immediate increased risks to study patients already on therapy (Kornbluth and Dzau, 2011). |
| | Enrollment in the three trials is suspended (Duke University, 2007a,b, 2008). Patients on trials are informed of the controversy and reconsented (Kornbluth and Cuffe, 2010). |
| | Duke IRB commissions an independent, external two-person review of the scientific methodology in question. NCI provides assistance in identifying potential external experts |

| | |
|---|---|
| | (Kornbluth and Dzau, 2011; McShane, 2010a). |
| | Baggerly and Coombes' data analysis and questions from the *Annals of Applied Statistics* paper were shared with Duke IRB and Principal Investigators of the three clinical trials (Kornbluth and Dzau, 2011). |
| | 10/23 – *The Cancer Letter* reports statements from coauthors of the *Lancet Oncology* study that the validation was never blinded (Goldberg, 2009b). |
| November 2009 | 11/9 – Baggerly sends report highlighting problems with data posted on a webpage on the cisplatin and pemetrexed tests to Kornbluth at Duke. This report was shared with Nevins, who asked that it be withheld from the external reviewers; Duke leadership decided to honor Nevins' request (Kornbluth and Dzau, 2011). |
| | 11/9 – Claudio Dansky Ullmann of NCI submits review of revised CALGB-30702 protocol (Genome-Guided Chemotherapy for Untreated and Treated Advanced Stage Non-Small Cell Lung Cancer: A Limited Institution, Randomized Phase II Study) to NCI's Cancer Therapy Evaluation Program (CTEP) Protocol and Information Office and forwards the review and disapproval letter to CALGB.[f,g] |
| | 11/16 – Lisa McShane and Jeffrey Abrams of NCI contact CALGB requesting re-evaluation of the Lung Metagene Score (LMS) predictor for CALGB-30506.[h] |
| | Ullmann and McShane contribute to erratum published in *Current Oncology Reports* to Jolly Graham and Potti (2009). |
| December 2009 | External review finds that "In summary we believe the predictors are scientifically valid and with a few additions can be fully responsive to the comments of Baggerly and Coombes" (*Review of genomic predictors for clinical trials from Nevins, Potti, and Barry*, 2009). |
| January 2010 | Letter submitted to NCI on 1/7/2010, accompanied by report from external reviewers (Kornbluth and Dzau, 2011; *Review of genomic predictors for clinical trials from Nevins, Potti, and Barry*, 2009; McShane, 2010a). |
| | Duke restarts the three trials (NCT00545948, NCT00509366, and NCT00636441) (ClinicalTrials.gov, 2011a,b,c). |
| February 2010 | NCI completes reevaluation of supporting data for the CALGB-30506 trial (NCI, 2010b). |
| March 2010 | Nevins, Potti, and colleagues send letter to McShane in response to some of her concerns about the LMS used in CALGB-30506.[i] |
| | McShane and Abrams reply with the conclusions of their analysis of the LMS in the CALGB-30506 clinical trial: The test should not remain as a stratification factor, and the coprimary aim to evaluate its performance should be removed from the study.[j] |
| April 2010 | CTEP requests data and computer code from Potti regarding R01 grant CA131049-01A1 entitled "Prospective validation of genomic signatures of chemosensitivity in NSCLC" (cisplatin and pemetrexed predictors).[k] |
| | Potti responds to CTEP.[l] |
| | *The Cancer Letter* obtains a copy of Duke University's external review report from NCI via a *Freedom of Information Act* request and publishes the document (Goldberg, 2010a). |
| May 2010 | CTEP sends follow-up questions to Potti regarding their response to the April 2010 request regarding the cisplatin and pemetrexed predictors. Potti responds.[m] |
| June 2010 | NCI completes reevaluation of the cisplatin chemosensitivity predictor (McShane, 2010c). |
| | NCI hosts Duke researchers to discuss genomic predictors developed at Duke. NCI states that it is not satisfied, and directs Potti and Nevins to conduct a search of their labs to supply the |

| | |
|---|---|
| | data and code reproducing the results in Hsu et al. (2007) and justifying the trials underway. Duke statistician William Barry is tasked with checking the cisplatin/pemetrexed predictors and verifying the data (NCI, 2010a; Kornbluth and Dzau, 2011; TMQF Committee, 2011b). |
| July 2010 | 7/16 – *The Cancer Letter* reports that Anil Potti incorrectly stated his credentials. Duke places Potti on administrative leave while University investigates allegations of inaccuracies in his curriculum vitae and in the research with Nevins (Goldberg, 2010b). |
| | 7/19 – 31 biostatisticians and bioinformatics experts from around the world send letter, "Concerns about prediction models used in Duke clinical trials," to NCI director Harold Varmus. This letter is later signed by two additional statisticians (Baron et al., 2010). |
| | 7/23 – *Lancet Oncology* issues an expression of concern for "Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: A substudy of the EORTC 10994/BIG 00-01 clinical trial" (Bonnefoi et al., 2007). |
| | NCT00545948, NCT00509366, and NCT00636441 trials suspended a second time (ClinicalTrials.gov, 2011a,b,c). |
| | 7/30 – NCI and Duke request assistance from the Institute of Medicine (IOM) in assessing the scientific foundation of the three clinical trials and identifying appropriate evaluation criteria for future tests based on omics technologies. |
| August 2010 | 8/27 – Duke completes review of Potti's credentials; identifies issues of substantial concerns resulting in corresponding sanctions. Potti remains on administrative leave (Duke Today, 2010). |
| October 2010 | 10/22 – Duke officials inform NCI they have determined that several datasets that were reported to have been used to validate the cisplatin predictor were found to be flawed. The Hsu et al. (2007) paper would be retracted. Investigation into other datasets was ongoing (McShane, 2010a). |
| November 2010 | NCT00545948, NCT00509366, and NCT00636441 trials terminated in ClinicalTrials.gov (ClinicalTrials.gov, 2011a,b,c). |
| | 11/16 – *Journal of Clinical Oncology* retracts "Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer" (Hsu et al., 2007; Hsu et al., 2010) |
| | 11/19 – Anil Potti resigns from his position at Duke (DukeHealth.org, 2010), later taking a position as an oncologist in South Carolina (*The Cancer Letter,* 2010) with strong endorsement from some Duke faculty (Duke.Fact.Checker, 2011). |
| December 2010 | 12/20 – McShane describes to the IOM committee NCI interactions pertaining to the genomic signatures, and supplies documentation to the committee. This is the first public explanation of why the NCI thought problems with the LMS were severe enough to warrant pulling it from CALGB 30506. This publicly calls the *NEJM* paper into question. In addition, she reveals that NCI had discovered that it had been providing partial funding to the trial NCT00509366 through an R01 grant awarded to Anil Potti. She describes her unsuccessful attempts to reproduce the results reported in the Hsu et al. (2007) paper for the cisplatin predictor and how that eventually led to discovery of several corrupted datasets (McShane, 2010a). |
| January 2011 | IGSP Center for Applied Genomics and Technology dissolved (Goldberg, 2011; Havele, 2011). |
| | *Nature Medicine* retraction (Potti et al., 2011a). |
| | 1/31 – The Food and Drug Administration (FDA) conducts an inspection at Duke University to detemune the rationale for the IRB's initial non-significant risk decision regarding an Investigational Device Exemption (IDE) (FDA, 2011). |

| | |
|---|---|
| February 2011 | *Lancet Oncology* retraction (Bonnefoi et al., 2011). |
| March 2011 | *NEJM* retraction (Potti et al., 2011b). |
| | Draft document, *A framework for the quality of translational medicine with a focus on human genomic studies: Principles from the Duke Medicine Translational Medicine Quality Framework [TMQF] committee,* released. Final draft is released in May 2011. |
| July 2011 | Duke sends the IOM committee a list of identified problems, missed signals, and proposed solutions based on the work of the TMQF committee (TMQF Committee, 2011b). |
| August 2011 | 8/22 Duke representatives meet with IOM committee: Robert Califf, Sally Kornbluth, Michael Cuffe, Ross McKinney, John Falletta, Geoff Ginsburg, Michael Kelley, and William Barry. |
| January 2012 | 1/25 The FDA posts documents on its website indicating that it informed Duke in 2009 that an IDE should have been obtained for the three trials (Chan, 2009; FDA, 2011; Potti, 2009) |
| | *Journal of Clinical Oncology* retracts "An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer" (Dressman et al, 2007; JCO, 2012). |

[a]Communication from Michael Burns, *Nature Medicine*, to Keith Baggerly, MD Anderson Cancer Center. Receipt of NMED-LE40837, May 30, 2008.
[b]Communication from Alison Farrell, *Nature Medicine*, to Keith Baggerly, MD Anderson Cancer Center. NMED-LE40837, June 2, 2008.
[c]Communication from Alison Farrell, *Nature Medicine*, to Keith Baggerly, MD Anderson Cancer Center. Decision on NMED-LE40837, June 11, 2008.
[d]Communication from David Collingridge, *Lancet Oncology,* to Keith Baggerly, MD Anderson Cancer Center. Your submission to the *Lancet Oncology*, September 6, 2008.
[e]Communication from Olwen Hahn, CALGB, to Michael Montello, National Cancer Institute. RE: CALGB 30702, July 28, 2009.
[f]Communication from Claudio Dansky Ullmann, National Cancer Institute, to CTEP Protocol and Information Office. Consensus review of revised protocol CALGB 30702: Genome-guided chemotherapy for untreated and treated advanced stage non-small cell lung cancer: A limited institution, randomized phase II study, November 9, 2009.
[g]Communication from Claudio Dansky Ullmann, National Cancer Institute, to Richard Schilsky, CALGB. Reference number PCALBG-30702#R01PDISAPP01, November 9, 2009.
[h]Communication from Jeffrey Abrams and Lisa McShane, National Cancer Institute, to Richard Schilsky, CALGB. Important computer code and data request for CALGB-30506, November 16, 2009.
[i]Communication from Joseph R. Nevins, Anil Potti, William Barry, and David Harpole, Duke University. Response to the NCI re-evaluation of supporting data for the CALGB-30506 trial, March 8, 2010.
[j]Communication from Lisa McShane and Jeffrey Abrams, National Cancer Institute, to Joseph R. Nevins, Anil Potti, William Barry, and David Harpole, Duke University. RE: Nevins, Potti, Barry, and Harpole response to the NCI re-evaluation of supporting data for the CALGB-30506 trial, March 26, 2010.
[k]Communication from William C. Timmer, National Cancer Institute, to Anil Potti, Duke University. RE: R01CA131049-01A1 information request, April 13, 2010.
[l]Communication from Anil Potti, Duke University, to William C. Timmer, National Cancer Institute. RE: R01CA131049-01A1 information request, April 29, 2010.
[m]Communication from Lisa McShane, National Cancer Institute, to Anil Potti, Duke University. RE: R01CA131049-01A1 information request, May 17, 2010.

## REFERENCES

Baggerly, K. A. 2011. Forensics bioinformatics. Presentation to the Workshop of the IOM Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30-31.

Baggerly, K. A., and K. R. Coombes. 2009. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Annals of Applied Statistics* 3(4):1309-1334.

Baggerly, K. A., K. R. Coombes, and E. S. Neeley. 2008. Run batch effects potentially compromise the usefulness of genomic signatures of ovarian cancer. *Journal of Clinical Oncology* 26(7):1186-1187.

Baron, A. E., K. Bandeen-Roche, D. A. Berry, J. Bryan, V. J. Carey, K. Chaloner, M. Delorenzi, B. Efron, R. C. Elston, D. Ghosh, J. D. Goldberg, S. Goodman, F. E. Harrell, S. Galloway Hilsenbeck, W. Huber, R. A. Irizarry, C. Kendziorski, M. R. Kosorok, T. A. Louis, J. S. Marron, M. Newton, M. Ochs, J. Quackenbush, G. L. Rosner, I. Ruczinski, S. Skates, T. P. Speed, J. D. Storey, Z. Szallasi, R. Tibshirani, and S. Zeger. 2010. Letter to Harold Varmus: Concerns about prediction models used in Duke clinical trials. Bethesda, MD, July 19.

Bild, A. H., G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck, J. A. Olson, Jr., J. R. Marks, H. K. Dressman, M. West, and J. R. Nevins. 2006. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439(7074):353-357.

Bonnefoi, H., A. Potti, M. Delorenzi, L. Mauriac, M. Campone, M. Tubiana-Hulin, T. Petit, P. Rouanet, J. Jassem, E. Blot, V. Becette, P. Farmer, S. Andre, C. R. Acharya, S. Mukherjee, D. Cameron, J. Bergh, J. R. Nevins, and R. D. Iggo. 2007. Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: A substudy of the EORTC 10994/BIG 00-01 clinical trial. *Lancet Oncology* 8(12):1071-1078.

Bonnefoi, H., A. Potti, M. Delorenzi, L. Mauriac, M. Campone, M. Tubiana-Hulin, T. Petit, P. Rouanet, J. Jassem, E. Blot, V. Becette, P. Farmer, S. Andre, C. Acharya, S. Mukherjee, D. Cameron, J. Bergh, J. R. Nevins, and R. D. Iggo. 2011. Retraction: Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: A substudy of the EORTC 10994/BIG 00-01 clinical trial. *Lancet Oncology* 12(2):116.

Califf, R. M. 2011. Discussion at Discovery of Process Working Group Meeting with representatives of Duke faculty and administration, Washington, DC, August 22.

*The Cancer Letter.* 2011. In the cancer centers. 37(22):1.

Chan, M. M. 2009. *Letter to Division of Medical Oncology, Duke University Medical Center*. http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289102.pdf (accessed February 9, 2012).

ClinicalTrials.gov. 2011a. *History of NCT00509366.* http://clinicaltrials.gov/archive/NCT00509366. (accessed December 12, 2011).

ClinicalTrials.gov. 2011b. *History of NCT00545948*. http://clinicaltrials.gov/archive/NCT00545948 (accessed December 12, 2011).

ClinicalTrials.gov. 2011c. *History of NCT00636441.* http://clinicaltrials.gov/archive/NCT00636441 (accessed December 12, 2011).

ClinicalTrials.gov. 2011d. *Genomic directed salvage chemotherapy with either liposomal doxorubicin or topotecan*. http://clinicaltrials.gov/ct2/show/NCT00720096?term=NCT00720096&rank=1 (accessed October 11, 2011).

Coombes, K. R., J. Wang, and K. A. Baggerly. 2007. Microarrays: Retracing steps. *Nature Medicine* 13(11):1276-1277.

Correction. 2007. *New England Journal of Medicine* 356(2):201-202.

Cuffe, M. 2011. Discussion at Discovery of Process Working Group Meeting with representatives of Duke faculty and administration, Washington, DC, August 22.

Dressman, H. K., A. Berchuck, G. Chan, J. Zhai, A. Bild, R. Sayer, J. Cragun, J. Clarke, R. S. Whitaker, L. Li, G. Gray, J. Marks, G. S. Ginsburg, A. Potti, M. West, J. R. Nevins, and J. M. Lancaster. 2007. An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *Journal of Clinical Oncology* 25(5):517-525.

Dressman, H. K., A. Potti, J. R. Nevins, and J. M. Lancaster. 2008. In reply. *Journal of Clinical Oncology* 26(7):1187-1188.

Duke.Fact.Checker. 2011. *Texts of letters of recommendation for Dr. Anil Potti.* http://dukefactchecker.blogspot.com/2011/06/texts-of-letters-of-recommendation-for.html (accessed December 12, 2011).

DukeHealth.org. 2010. *Duke accepts Potti resignation; retraction process initiated with* Nature Medicine. http://www.dukehealth.org/health_library/news/duke-accepts-potti-resignation-retraction-process-initiated-with-nature-medicine (accessed December 12, 2011).

Duke Today. 2010. *Duke updates response to Potti allegations.* http://today.duke.edu/2010/08/pottiresponse.html (accessed December 12, 2011).

Duke University. 2007a. *Adjuvant cisplatin with either genomic-guided vinorelbine or pemetrexed for early stage non-small-cell lung cancer (TOP0703)*. http://clinicaltrials.gov/ct2/show/NCT00545948?term=nct00545948&rank=1 (accessed November 23, 2011).

Duke University. 2007b. *Study using a genomic predictor of platinum resistance to guide therapy in stage IIIB/IV non-small cell lung cancer (TOP0602)*. http://clinicaltrials.gov/ct2/show/NCT00509366?term=nct00509366&rank=1 (accessed November 23, 2011).

Duke University. 2008. *Trial to evaluate genomic expression profiles to direct preoperative chemotherapy in early stage breast cancer*. http://clinicaltrials.gov/show/NCT00636441 (accessed November 22, 2011).

Falletta, J. 2011. Discussion at Discovery of Process Working Group Meeting with representatives of Duke faculty and administration, Washington, DC, August 22.

Food and Drug Administration (FDA). 2011. *FDA establishment inspection report, Duke University Medical Center*. http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289106.pdf (accessed February 9, 2012).

Ginsburg, G. S. 2011. Discussion at Discovery of Process Working Group Meeting with representatives of Duke faculty and administration, Washington, DC, August 22.

Goldberg, P. 2009a. A biostatistic paper alleges potential harm to patients in two Duke clinical studies. *The Cancer Letter* 35(36):1-5.

Goldberg, P. 2009b. Duke halts third trial; coauthor disputes claim that data validation was blinded. *The Cancer Letter* 35(39):1-4.

Goldberg, P. 2010a. NCI raises new questions about Duke genomics research, cuts assay from trial. *The Cancer Letter* 36(18):1-7.

Goldberg, P. 2010b. Prominent Duke scientist claimed prizes he didn't win, including Rhodes Scholarship. *The Cancer Letter* 36(27):1-7.

Goldberg, P. 2011. FDA auditors spend two weeks at Duke; Nevins loses position in reorganization. *The Cancer Letter* 37(4):1-2, 4-5.

Gyorffy, B., P. Surowiak, O. Kiesslich, C. Denkert, R. Schafer, M. Dietel, and H. Lage. 2006. Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. *International Journal of Cancer* 118(7):1699-1712.

H. Lee Moffitt Cancer Center & Research Institute. 2007. NCT00720096 Protocol version 10, October 19.

H. Lee Moffitt Cancer Center & Research Institute. 2008a. NCT00720096 Protocol version 13, January 10.

H. Lee Moffitt Cancer Center & Research Institute. 2008b. NCT00720096 Protocol version 14, July 14.

H. Lee Moffitt Cancer Center & Research Institute. 2009. NCT00720096 Protocol version 15, July 26.

Havele, S. 2011. IGSP reviews organization, future plans. *The Chronicle* January 21. http://dukechronicle.com/article/igsp-reviews-organization-future-plans (accessed January 13, 2012).

Hsu, D. S., B. S. Balakumaran, C. R. Acharya, V. Vlahovic, K. S. Walters, K. Garman, C. Anders, R. F. Riedel, J. Lancaster, D. Harpole, H. K. Dressman, J. R. Nevins, P. G. Febbo, and A. Potti. 2007. Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *Journal of Clinical Oncology* 25(28):4350-4357.

Hsu, D. S., B. S. Balakumaran, C. R. Acharya, V. Vlahovic, K. S. Walters, K. Garman, C. Anders, R. F. Riedel, J. Lancaster, D. Harpole, H. K. Dressman, J. R. Nevins, P. G. Febbo, and A. Potti. 2010. Retraction: Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *Journal of Clinical Oncology* 28(35):5229.

JCO (Journal of Clinical Oncology). *An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer: Retraction* http://jco.ascopubs.org/content/25/5/517/suppl/DC1 (accessed January 30, 2012).

Jolly Graham, A., and A. Potti. 2009. Translating genomics into clinical practice: Applications in lung cancer. *Current Oncology Reports* 11(4):263-268.

Kornbluth, S. 2011. Discussion at Discovery of Process Working Group Meeting with representatives of Duke faculty and administration, Washington, DC, August 22.

Kornbluth, S. A. and M. Cuffe. 2010. *Preliminary accounting of events at Duke University*. Durham, NC: Duke University.

Kornbluth, S. A., and V. Dzau. 2011. *Predictors of chemotherapy response: Background information: Draft.* Duke University.

Lancaster, J. M. 2008. *Molecular profiling to predict response to chemotherapy, 5R33CA110499-05*. http://projectreporter.nih.gov/project_info_description.cfm?aid=8101020&icde=10060731 (accessed October 11, 2011).

Larsen, J. E., K. M. Fong, and N. K. Hayward. 2007. To the editor: Refining prognosis in non-small-cell lung cancer. *New England Journal of Medicine* 356(2):190.

Marcom, P. K. 2008. *A randomized Phase II trial evaluating the performance of genomic expression profiles to direct the use of preoperative chemotherapy for early stage breast cancer.* Durham, NC: Duke Institute for Genome Sciences and Policy.

McKinney, R. 2011. Discussion at Discovery of Process Working Group Meeting with representatives of Duke faculty and administration, Washington, DC, August 22.

McShane, L. M. 2010a. *NCI address to Institute of Medicine committee convened to review omics-based tests for predicting patient outcomes in clinical trials*. Presentation at Meeting 1: Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, December 20.

McShane, L. M. 2010b. Notes from June 29 meeting with Duke.

McShane, L. M. 2010c. *Re-analysis report for cisplatin chemosensitivity predictor.* Bethesda, MD: National Cancer Institute.

NCI (National Cancer Institute). 2010a. Discussion of genomic predictors developed at Duke University. Presented at the National Cancer Institute, Rockville, MD June 29.

NCI. 2010b. Executive summary: NCI re-evaluation of supporting data for the CALGB-30506 trial. Bethesda, MD: National Cancer Institute.

Nevins, J. 2011. *Genomic strategies to address the challenge of personalizing cancer therapy.* Presentation at the Workshop of the IOM Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Washington, DC, March 30-31.

Pittman, J. E. Huang, J. Nevins, Q. Wang, M. West. 2004. Bayesian analysis of binary prediction tree models for retropsectively sampled outcomes. *Biostatistics* 5(4):587-601.

Potti, A. 2009. *Letter to FDA's CDER from Division of Medical Oncology, Duke University Medical Center.* http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/UCM289103.pdf (accessed February 9, 2012).

Potti, A., and J. R. Nevins. 2007. Potti et al. reply. *Nature Medicine* 13(11):1277-1278.

Potti, A., H. K. Dressman, A. Bild, R. F. Riedel, G. Chan, R. Sayer, J. Cragun, H. Cottrill, M. J. Kelley, R. Petersen, D. Harpole, J. Marks, A. Berchuck, G. S. Ginsburg, P. Febbo, J. Lancaster, and J. R. Nevins. 2006a. Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine* 12(11):1294-1300.

Potti, A., S. Mukherjee, R. Petersen, H. K. Dressman, A. Bild, J. Koontz, R. Kratzke, M. A. Watson, M. Kelley, G. S. Ginsburg, M. West, D. H. Harpole, and J. R. Nevins. 2006b. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *New England Journal of Medicine* 355(6):570-580.

Potti, A., H. K. Dressman, A. Bild, R. F. Riedel, G. Chan, R. Sayer, J. Cragun, H. Cottrill, M. J. Kelley, R. Petersen, D. Harpole, J. Marks, A. Berchuck, G. S. Ginsburg, P. Febbo, J. Lancaster, and J. R. Nevins. 2007a. Corrigendum: Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine* 13(11):1388.

Potti, A., D. Harpole, and J. R. Nevins. 2007b. The authors reply: Refining prognosis in non-small-cell lung cancer. *New England Journal of Medicine* 356(2):190-191.

Potti, A., H. K. Dressman, A. Bild, R. F. Riedel, G. Chan, R. Sayer, J. Cragun, H. Cottrill, M. J. Kelley, R. Petersen, D. Harpole, J. Marks, A. Berchuck, G. S. Ginsburg, P. Febbo, J. Lancaster, and J. R. Nevins. 2008. Corrigendum: Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine* 14(8):889.

Potti, A., H. K. Dressman, A. Bild, G. Chan, R. Sayer, J. Cragun, H. Cottrill, M. J. Kelley, R. Petersen, D. Harpole, J. Marks, A. Berchuck, G. S. Ginsburg, P. Febbo, J. Lancaster, and J. R. Nevins. 2011a. Retraction: Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine* 17(1):135.

Potti, A., S. Mukherjee, R. Petersen, H. K. Dressman, A. Bild, J. Koontz, R. Kratzke, M. A. Watson, M. Kelley, G. S. Ginsburg, M. West, D. H. Harpole, Jr., and J. R. Nevins. 2011b. Retraction: A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *New England Journal of Medicine* 364(12):1176.

Ready, N. 2010. *Phase II prospective study evaluating the role of directed cisplatin based chemotherapy with either vinorelbine or pemetrexed for the adjuvant treatment of early stage non-small cell lung cancer (NSCLC) in*

*patients using genomic expression profiles of chemotherapy sensitivity to guide therapy.* Durham, NC: Duke University Medical Center.

*Review of genomic predictors for clinical trials from Nevins, Potti, and Barry.* 2009. Durham, NC: Duke University.

Singh, T., and J. Dhindsa. 2007. To the editor: Refining prognosis in non-small-cell lung cancer. *New England Journal of Medicine* 356(2):190.

Sun, Z., and P. Yang. 2007. To the editor: Refining prognosis in non-small-cell lung cancer. *New England Journal of Medicine* 356(2):189-190.

TMQF Committee. 2011b. *A framework for the quality of translational medicine with a focus on human genomic studies: Principles from the Duke Medicine Translational Medicine Quality Framework Committee*. Durham, NC: Duke University.

Translational Medicine Quality Framework (TMQF) Committee. 2011a. *Draft table of categories and areas of improvement related to TMQF.* Durham, NC: Duke University.

Vlahovic, V. 2010. *Phase II prospective study evaluating the role of personalized chemotherapy regimens for chemo-naive select Stage IIIB and IV non-small cell lung cancer (NSCLC) in patients using a genomic predictor of platinum-resistance to guide therapy.* Durham, NC: Duke University Medical Center.

Zuiker, A. 2008. *Building a just culture.* http://inside.duke.edu/article.php?IssueID=183&ParentID=17859 (accessed November 22, 2011).

# Appendix C
# Introduction to Biomarkers

The purpose of this appendix is to introduce readers to biomarkers, uses of biomarkers, and requirements for determining clinical utility of biomarkers. This material is intended for readers who wish to learn more about general biomarker concepts. This appendix explains how specific terms are used in the report and also illustrates several common misconceptions about biomarkers and the terminology used to describe them.

## BIOMARKERS

Investigators seek to discover new biomarkers for many purposes, including for uses where they would guide clinical decision-making regarding how best to select agents to treat individual patients. The scientific literature provides definitions of the term *biomarker* as well as some of the principal uses of biomarkers. A widely used definition of a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n]…intervention" (Biomarkers Definitions Working Group, 2001). A recent IOM report on *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease* provided the following description of biomarkers (IOM, 2010): Biomarkers can be measurements of macromolecules (DNA, RNA, proteins, lipids), cells, or processes that describe a normal or abnormal biological state in an organism.

Biomarkers may be detected and analyzed in tissue, in circulation (blood, lymph), and in secretions (urine, stool, sputum, breast nipple aspiration, etc.). Biomarkers may be identified in cells of concern (such as pre-malignant or even existing cancer cells) or in tissues surrounding the area of interest (such as evidence of neo-vascularization or inflammation surrounding a cancer). Biomarkers may be in exfoliated cells, or they may be soluble or suspended molecules (for example proteins, DNA, microRNA) in circulation or in secretions. Finally, inherited germline biomarkers can be evaluated from circulating leukocytes or exfoliated cells from easily accessible tissues, such as from a cheek swab.

## CLINICAL USES OF BIOMARKERS

Biomarkers have many important potential roles in clinical research and in clinical practice (IOM, 2010). See Tables 1-1 and 1-2 in Chapter 1. These include prognosis, prediction of response to therapy (effect modifiers), prediction of clinical outcome (surrogate endpoints), risk assessment, screening, diagnosis, pharmacogenetics, and patient monitoring during and after treatment. Although these uses are applicable to most if not all disease processes, this appendix refers to oncologic examples because most of the case studies for this report arose from the field of oncology.

### Prognostic Factors and Effect Modifiers

Also see the discussion in Chapter 1.

*Prognostic Factors*

Prognostic factors, used to estimate the risk of or time to clinical outcomes such as disease recurrence or progression, may be useful even though these biomarkers are simply correlated with the causal mechanisms of the disease process (Fleming, 2005). In oncology, biomarkers may have both roles as prognostic factors and effect modifiers, and, in fact, these may be mixed. For example, HER2 overexpression and/or amplification portend a poor prognosis in breast cancer patients who do not receive any adjuvant systemic therapy. Moreover, this same biomarker appears to be associated with poorer response to endocrine therapy in patients with ER positive breast cancer (compared to those who have ER positive, HER2 negative cancers), but it has been associated with higher response rates to various chemotherapies (anthracyclines, taxanes) and it is especially related to benefit from anti-HER2 therapies, such as trastuzumab and lapatinib (Wolff et al., 2007). Therefore, any study addressing the prognostic role of a biomarker needs to take into account the specific, intended clinical use of the biomarker and the potential confounding effects of how patients in the study cohort are treated. Perhaps the best example of a successful omics-based test for prognosis is the development of Onco*type*Dx. The case study for this omics-based test is described in great detail in Chapter 5.

*Effect Modifiers*

An effect modifier may relate to a class of therapy (such as chemotherapy in general) or a specific agent within the class (such as an anthracycline, or, more specifically, doxorubicin). Indeed, the genesis of this report stems from development and application of omics-based tests designed to be effect modifiers for chemotherapy in several types of cancers, including lung, breast, and ovarian cancer (see Chapter 6).

With the advent of targeted cancer therapies that are directed towards a specific, somatic molecular abnormality, biomarkers may provide direct guidance for selection of individual agents. Examples of these types of biomarkers include the expression status of the estrogen receptor and of HER2 for selection of endocrine or anti-HER2 therapies, respectively, in breast cancer (Hammond et al., 2010; Wolff et al., 2007), KRAS mutations for selection of antibody therapy against the epidermal growth factor receptor (Allegra et al., 2009), and ALK mutations for selection of crizotinib in lung cancer (Shaw et al., 2010).

## Surrogate Endpoints

A third important category of biomarkers is that of surrogate endpoints. A surrogate endpoint is "a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence" (Biomarkers Definitions Working Group, 2001). In standard clinical practice, a surrogate endpoint may provide the clinician either increased certainty that an event is already, or is likely to be, occurring. For example, if the level of a tumor biomarker in the bloodstream is rising in a patient with previously established cancer, that may be an indication of an impending relapse, and might guide earlier intervention than if the clinician waits for the relapse to be detectable by other means. Such a surrogate may have clinical/biological validity but may or may not have clinical utility. This issue is described in greater detail below.

To illustrate the higher level of complexity to validate a biomarker as a surrogate or as an effect modifier, the fact that patients with tumor response is correlated with longer life does not allow one to conclude that a treatment that induces tumor response will also induce an increase in survival duration. In other words, "a correlate does not a surrogate make" (Fleming and DeMets, 1996). In turn, when using biomarkers in the development of omics-based tests in order to guide decisions about when to use available agents, it is not enough to establish that the biomarker is a prognostic factor. It is important to recognize that "a prognostic factor does not an effect modifier/predictive factor make" (Fleming and Powers, In Press).

### Risk Assessment

Biomarkers can be used to assess a patient's risk for a future diagnosis of disease. Risk assessment is particularly valuable if preventive measures and/or early detection and intervention have been shown to effectively reduce morbidity or mortality. For example, prophylactic surgery and/or chemoprevention with the selective estrogen receptor modulators (SERMs; tamoxifen, raloxifene) have been shown to reduce the incidence and, in the case of surgery, associated mortality of breast cancer (Newman and Vogel, 2007). However, application of these strategies is only applied to subjects with appropriately high risk: in particular, women with either an inherited genetic risk for breast cancer development, such as those who harbor mutations in the BRCA1 and 2 genes, or those with a sufficiently high calculated risk of a new breast cancer using well validated instruments such as the Breast Cancer Risk Assessment Tool of the National Cancer Institute (NCI) (Gail et al., 2007).

### Screening

For several diseases, screening permits diagnosis of disease at an earlier, more treatable point. Screening strategies have been implemented for a number of cancers, including breast, colorectal, lung, prostate, and cervical cancer, and have been proposed for others, such as ovarian cancer. Of these, reduction of mortality has only been demonstrated for breast (USPSTF, 2009), colorectal (AHRQ, 2012; Zauber et al., 2012), lung (Aberle et al., 2011) and cervical cancers (Gates, 2001; NCI, 2012). Radiological or endoscopic methods have been employed in the first three examples, while cytologic examination—and more recently, detection of human papilloma virus (HPV) infection—is used to screen for cervical cancer. For colorectal cancer screening, biomarker stool tests are also used for screening (IOM, 2008). Only single biomarkers have thus far been accepted as effective screening approaches.

### Diagnosis

Diagnostic biomarkers are used to confirm whether a patient has a particular disease. In an oncology setting, a patient may present with clinical, radiographic, and even histologic findings that are strongly suggestive, or confirm, the diagnosis of malignancy, but the primary source of the cancer is not easily discernible. Biomarkers might be used to distinguish either a diagnosis of cancer vs. benign, or the site of origin of the cancer (for example, lymphoma vs. solid malignancy, or one type of epithelial cancer, such as colon, vs. another, such as breast). Indeed, one of the case studies presented in Chapter 5 provides an example of an omics-based test, the Pathworks Tissue of Origin test, which may effectively identify the primary site of a malignancy of unknown origin.

### Pharmacogenetics

In many diseases, inherited germline DNA sequence variants, known as single nucleotide polymorphisms (SNPs), may determine individual differences in drug distribution within the body, metabolism, or effect on target tissues. SNPs can be used as biomarkers, and the field of study of these individual differences is known as pharmacogenetics. Such biomarkers may be used to predict particular susceptibility to drug toxicities and/or activities (Wang et al., 2011; Weinshilboum, 2003).

### Monitoring

Biomarkers may be used to periodically monitor subjects for an impending, but at the present time occult, event. Although tissue-based biomarkers might be monitored, such a strategy is invasive, and a strategy making use of biomarkers that are secreted or circulating in the bloodstream is preferable and more frequently used. Monitoring biomarkers might be used for monitoring previously effected individuals who have been rendered disease free, to detect an event earlier than what might be possible with standard clinical approaches.

## OMICS-BASED BIOMARKERS

In the past, most single-factor biomarkers have been generated and studied because of a pre-conceived biological association between them and the associated disease. For example, ER was pursued as an effect modifier because of the known beneficial effect of endocrine therapy (ovarian ablation, pharmacologic estrogen therapy, tamoxifen) on some but not all breast cancers (Hammond et al., 2010; McGuire et al., 1975). On the other hand, discovery of associations between omics-based tests and disease biology and/or clinical outcomes may be more likely to be a result of mathematical correlations between large numbers of high order factors (such as a 15,000 gene array expression chip). The resulting signature, which must then be developed into a clinically applicable omics-based test, requires data intensive methods to provide an evidence-based validation of the use of the biomarker. While still desirable and beneficial, it is not always possible to link omics-based biomarkers to biological rationale.

### Technologies Enabling Omics-Based Research

Technologies enabling omics research include gene expression microarrays, multiplex quantitative reverse transcriptase polymerase chain reaction (qRT-PCR), sequence analysis of DNA, RNA, and proteins, and multiple mass spectrometry techniques. On a gene expression microarray, thousands of oligonucleotides are arranged on a surface and hybridize with corresponding nucleic acid sequences in complex biological samples from tissues or plasma. In qRT-PCR, fluorescent-labeled oligonucleotide probes are used to amplify, detect, and quantify the presence of multiple genetic sequences and track gene transcription into RNA. Analysis of the sequences of DNA, RNA, and proteins enables better understanding of fundamental biological function and interaction. Multiple mass spectrometry techniques enable analysis of complex biological samples to identify and quantify proteins and their numerous modified isoforms. Other mass spectrometry or molecular magnetic resonance imaging instruments are important for metabolomics studies. These tools potentially open the door to development of omics-based tests that may help to improve treatment efficacy and help patients avoid adverse side effects of therapies.

## BIOMARKERS IN CLINICAL TRIALS

The use of prognostic biomarkers and effect modifiers in clinical trials is particularly relevant to the statement of task that led to this report. A biomarker might be used for direct assignment of patients to different treatment regimens within a clinical trial. For example, enrolling only patients who are positive for a biomarker that is known or presumed to be associated with a higher risk of subsequent events, or with a higher possibility of responding to a specific type of therapy, could reduce the size or duration of a clinical trial. In this case, the biomarker is used to enrich accrual to the trial with patients most likely to benefit from the treatment.

## REFERENCES

Aberle, D. R., C. D. Berg, et al. 2011. The National Lung Screening Trial: overview and study design. *Radiology* 258(1):243-253.

AHRQ (Agency for Healthcare Research and Quality). 2012. "Colorectal Cancer Screening." Retrieved February 13, 2012, from http://www.ahrq.gov/clinic/colorsum.htm.

Allegra, C. J., J. M. Jessup, M. R. Somerfield, S. R. Hamilton, E. H. Hammond, D. F. Hayes, P. K. McAllister, R. F. Morton, and R. L. Schilsky. 2009. American Society of Clinical Oncology provisional clinical opinion: Testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *Journal of Clinical Oncology* 27(12):2091-2096.

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 69(3):89-95.

Fleming, T. R. 2005. Surrogate endpoints and FDA's accelerated approval process. *Health Affairs* 24(1):67-78.

Fleming, T. R., and D. L. DeMets. 1996. Surrogate end points in clinical trials: Are we being misled? *Annals of Internal Medicine* 125(7):605-613.

Fleming, T. R., and J. H. Powers. In Press. Biomarkers and surrogate endpoints in clinical trials. *Statistics in Medicine.* In Press.

Gail, M. H., J. P. Costantino, D. Pee, M. Bondy, L. Newman, M. Selvan, G. L. Anderson, K. E. Malone, P. A. Marchbanks, W. McCaskill-Stevens, S. A. Norman, M. S. Simon, R. Spirtas, G. Ursin, and L. Bernstein. 2007. Projecting individualized absolute invasive breast cancer risk in African American women. *Journal of the National Cancer Institute* 99(23):1782-1792.

Gates, T. J. (2001). Screening for cancer: evaluating the evidence. *American Family Physician* 63(3): 513-522.

Hammond, M. E., D. F. Hayes, M. Dowsett, D. C. Allred, K. L. Hagerty, S. Badve, P. L. Fitzgibbons, G. Francis, N. S. Goldstein, M. Hayes, D. G. Hicks, S. Lester, R. Love, P. B. Mangu, L. McShane, K. Miller, C. K. Osborne, S. Paik, J. Perlmutter, A. Rhodes, H. Sasano, J. N. Schwartz, F. C. Sweep, S. Taube, E. E. Torlakovic, P. Valenstein, G. Viale, D. Visscher, T. Wheeler, R. B. Williams, J. L. Wittliff, and A. C. Wolff. 2010. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Journal of Clinical Oncology* 28(16):2784-2795.

Henry, N. L., and D. F. Hayes. 2006. Uses and abuses of tumor markers in the diagnosis, monitoring, and treatment of primary and metastatic breast cancer. *Oncologist* 11(6):541-552.

Hodgkinson, V. C., G. L. Eagle, P. J. Drew, M. J. Lind, and L. Cawkwell. 2010. Biomarkers of chemotherapy resistance in breast cancer identified by proteomics: Current status. *Cancer Letters* 294(1):13-24.

IOM (Institute of Medicine). 2010. *Evaluation of biomarkers and surrogate endpoints in chronic disease*. Washington, DC: The National Academies Press.

McGuire, W. L., G. C. Chamness, M. E. Costlow, and N. J. Richert. 1975. Steroids and human breast cancer. *Journal of Steroid Biochemistry* 6(5):723-727.

McGuire, W. L., A. K. Tandon, D. C. Allred, G. C. Chamness, and G. M. Clark. 1990. How to use prognostic factors in axillary node-negative breast cancer patients. *Journal of the National Cancer Institute* 82(12):1006-1015.

NCI (National Cancer Institute). (2012). "Cervical Cancer Screening (PDQ®)." Retrieved February 13, 2012, from http://www.cancer.gov/cancertopics/pdq/screening/cervical/HealthProfessional/page2.

Newman, L. A., and V. G. Vogel. 2007. Breast cancer risk assessment and risk reduction. *Surgical Clinics of North America* 87(2):307-316.

Nordgren, A., J. Schoumans, S. Soderhall, M. Nordenskjold, and E. Blennow. 2001. Interphase fluorescence in situ hyridization and spectral karyotyping reveals hidden genetic aeberrations in children with acute lymphoblastic leukaemia and a normal banded karyotype. *British Journal of Haematology* 114(4):786-793.

O'Connor, C. 2008. Human chromosome translocations and cancer. *Nature Education* 1(1).

Osin, P. P., and S. R. Lakhani. 1999. The pathology of familial breast cancer: Immunohistochemistry and molecular analysis. *Breast Cancer Research* 1(1):36-40.

Shaw, A. T., B. Y. Yeap, B. J. Solomon, G. J. Riely, J. Gainor, J. A. Engelman, G. I. Shapiro, D. B. Costa, S. H. Ou, M. Butaney, R. Salgia, R. G. Maki, M. Varella-Garcia, T. C. Doebele, Y. J. Bang, K. Kulig, P. Selaru, Y. Tang, K. D. Wilner, E. L. Kwak, J. W. Clark, A. J. Iafrate, and D. R. Camidge. 2010. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: A retrospective analysis. *Lancet Oncology* 12(11):1004-1012.

USPSTF (U.S. Preventive Services Task Force). 2009. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* 151(10): 716-726, W-236.

Wang, L., H. L. McLeod, and R. M. Weinshilboum. 2011. Genomics and drug response. *New England Journal of Medicine* 364(12):1144-1153.

Weinshilboum, R. 2003. Inheritance and drug response. *New England Journal of Medicine* 348(6):529-537.

Wolff, A. C., M. E. Hammond, J. N. Schwartz, K. L. Hagerty, D. C. Allred, R. J. Cote, M. Dowsett, P. L. Fitzgibbons, W. M. Hanna, A. Langer, L. M. McShane, S. Paik, M. D. Pegram, E. A. Perez, M. F. Press, A. Rhodes, C. Sturgeon, S. E. Taube, R. Tubbs, G. H. Vance, M. van de Vijver, T. M. Wheeler, and D. F. Hayes. 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Journal of Clinical Oncology* 25(1):118-145.

Zauber, A. G., S. J. Winawer, M. J. O'Brien, I. Lansdorp-Vogelaar, M. van Ballegooijen, B. F. Hankey, W. Shi, J. H. Bond, M. Schapiro, J. F. Panish, E. T. Stewart, and J. D. Waye. 2012. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* 366(8):687-696.

# Appendix D
# Reporting Guidelines

When preparing results of an omics-based test evaluation study, authors should clearly report the methods and results. Comprehensive reporting allows independent scientists and other stakeholders (journal editors, reviewers, clinicians, funding agencies, etc.) to assess the strength of a study and the findings. Inadequate reporting may impair the ability of the field to determine whether research findings are reproducible, accurate, and well founded. In addition, scientific research depends on understanding and building on previous findings. Thus, incomplete reporting can impede future research and new scientific discovery.

Evidence shows that many published biomarker research studies inadequately document important aspects of the scientific process. Brundage et al. (2002) noted that some studies do not provide the level of detail necessary to understand what methodologies were used and to understand important variations that arose, further complicating interpretations in the scientific literature. Studies can also fail to adequately address statistical analyses and ineffectively describe survival statistics (Riley et al., 2003). Missing information on covariate data, which was observed in 81 of 100 articles in one review, has the potential to introduce bias and interfere with the development of lead prognostic models (Burton and Altman, 2004).

In recent years, a number of international, multidisciplinary groups have developed reporting guidelines to improve the quality of published health research studies. A working definition of a reporting guideline is: "A checklist, flow diagram, or explicit text to guide authors in reporting a specific type of research, developed using explicit methodology" (Moher et al., 2010b). A seminal reporting guideline is the Consolidated Standards of Reporting Trials (CONSORT), which was first published in 1996 and updated in 2001 and 2010, to provide a checklist and flow diagram intended to improve the reporting of randomized controlled trials (RCTs) (Moher et al., 2010a). Other familiar reporting guidelines include:

- STARD (STAndards for the Reporting of Diagnostic accuracy studies)
- STROBE (STrengthening the Reporting of OBservational studies in Epidemiology)
- REMARK (REporting recommendations for tumor MARKer prognostic studies)
- MIAME (Minimum Information About a Microarray Experiment)
- PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)
- BRISQ (Biospecimen Reporting for Improved Study Quality)

Reporting guidelines are distinct from prescribing the methods used in the conduct of research. Although reporting focuses on describing the methods used in a specific study, it does not necessarily mean that a well-reported study is a high-quality study. However, there is some thought that reporting guidelines, by transparently and thoroughly describing the methods employed and results obtained, can help investigators assess the quality of a study.

## THE NEED FOR MULTIPLE REPORTING GUIDELINES IN OMICS

In the 2007 Institute of Medicine (IOM) report, *Cancer Biomarkers: The Promises and Challenges of Improving Detection and Treatment*, the committee noted that different sets of reporting guidelines would need to be developed depending on the technologies involved in a study, whether a single biomarker versus panels or patterns of biomarkers is being investigated, and the intended applications of the study. Table D-1 lists several reporting guidelines relevant to omics-based research. In addition, in June 2011, a group of 20 methodologists, clinicians, and journal editors from around the world convened to develop reporting guidelines for studies to develop and/or validate multivariable prediction models. The intended output of this meeting included the publication of a reporting guideline checklist that describes the guideline development process, as well as a longer explanatory publication that mirrors those produced for other reporting guidelines (Collins, 2011). Box D-1 provides the REMARK guideline as an example of a reporting guideline checklist. REMARK, first published in 2005, was developed to address a number of inadequacies of reporting prevalent in tumor marker prognostic studies (McShane et al., 2005). Altman et al. (In press) expands upon the REMARK checklist in order to improve its use and effectiveness by better clarifying the intent of each item and why the information is important to report. Each checklist item is explained and accompanied by published examples of good reporting, as well as relevant empirical evidence on the quality of reporting. The explanation and elaboration document highlights the REMARK profile, a suggested tabular format for summarizing key study details, and serves as a reference for the many issues to consider when designing, conducting, and analyzing tumor marker studies and prognostic studies in medicine in general.[1]

The current patchwork of reporting guidelines relevant to omics-based research may leave a number of gaps and overlaps in coverage (IOM, 2007). The 2007 IOM biomarker report concluded that current efforts to create reporting guidelines are piecemeal and primarily nonbinding, involving a number of professional organizations and other groups. To facilitate the development of coherent guidelines, the committee called on government agencies to work together with pharmaceutical and diagnostic industries, academia, and healthcare payors. The committee concluded: "[T]here is a great need for a coherent strategy to make the biomarker development and adoption process more transparent, to remove inconsistency and uncertainty, and to elevate the standards and oversight applied to biomarker tests. No federal agency currently takes responsibility for ensuring the clinical validity of biomarkers, but oversight and ownership of the process are key to developing strategies and making effective and efficient progress in the field" (IOM, 2007, p. 83). To provide leadership to this effort, the committee strongly urged an appropriate federal agency to coordinate and oversee interagency efforts in the development of standards and guidelines, and suggested that the National Institute of Standards and Technology may be an option.

No government agency has assumed this leadership role. However, in 2006, the National Knowledge Service of the U.K. National Health Service established the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network to foster coordination and collaboration in the development of reporting guidelines (EQUATOR Network, 2011) (see Box D-2). Similarly, a multidisciplinary group developed the MIBBI Project (The Minimum Information for Biological and Biomedical Investigators) in 2008 to harmonize the development of minimum information checklists for biological and biomedical investigations.

[1] Personal communication, Lisa McShane, NCI, January 9, 2012.

**TABLE D-1** Reporting Standards Used in Omics-Based Studies

| System | Date | Study Type | Structure | Adoption by Journals |
|---|---|---|---|---|
| REMARK | 2005 | Tumor marker prognostic studies | — A 20-item checklist with the headings: Introduction, Materials and Methods, Results, and Discussion<br>— NOTE: Item 12 states that a diagram may be helpful to illustrate the numbers of individuals included at different stages of a study | Mentioned in instructions to authors:<br>— *Clinical Cancer Research*<br>— *Breast Cancer Research and Treatment*<br>— *Journal of Clinical Oncology*<br>— *Journal of the National Cancer Institute*<br>— *Journal of Pathology* |
| CONSORT | 2010 (updated the 2001 statement) | Randomized controlled trials | — A 25-item checklist with the headings: Title, Abstract, Introduction, Methods, Results, Discussion, and Other Information<br>— A flow diagram depicting the passage of participants through a trial (enrollment, intervention allocation, follow-up, and analysis) | CONSORT is endorsed by over 50% of the core medical journals listed in the *Abridged Index Medicus* on PubMed |

| System | Date | Study Type | Structure | Adoption by Journals |
|---|---|---|---|---|
| MIAME | 2001 | Microarray-based gene expression experiments | Six critical elements:<br>1. Experimental design: the set of hybridization experiments as a whole<br>2. Array design: each array used and each element (spot, feature) on the array<br>3. Samples: samples used, extract preparation, and labeling<br>4. Hybridizations: procedures and parameters<br>5. Measurements: images, quantification, and specifications<br>6. Normalization controls: types, values, and specifications | Over 50 journals require MIAME-compliant data as a condition for publishing microarray-based papers |
| BRISQ | 2011 | Studies that use human biospecimens | Three tiers of data elements that should be considered for reporting, *if known or applicable*:<br>— First tier: Items recommended to report (e.g., organ[s] or anatomical site from which the biospecimens were derived)<br>— Second tier: Items beneficial to report (e.g., time from biospecimen excision/acquisition to stabilization)<br>— Third tier: Additional items to report (e.g., environmental factors to which patients were exposed) | None |

| System | Date | Study Type | Structure | Adoption by Journals |
|---|---|---|---|---|
| STARD | 2003 | Diagnostic accuracy | — A 25-item checklist with the headings: Title/Abstract/Keywords, Introduction, Methods, Results, and Discussion<br>— A flow diagram depicting the method of recruitment of patients or samples, the order of test execution, and the number of patients undergoing the test under evaluation and the reference test | More than 200 biomedical journals encourage the use of the STARD statement in their instructions for authors |

SOURCES: Bossuyt et al. (2004); Brazma et al. (2001); McShane et al. (2005); Moore et al. (2011); Schulz et al. (2010).

**BOX D-1**
**Example of a Reporting Guideline Checklist:**
**The REMARK Checklist**

**INTRODUCTION**

1. State the marker examined, the study objectives, and any prespecified hypotheses.

**MATERIALS AND METHODS**

**Patients**

2. Describe the characteristics (e.g., disease stage or comorbidities) of the study patients, including their source and inclusion and exclusion criteria.
3. Describe treatments received and how chosen (e.g., randomized or rule based).

**Specimen characteristics**

4. Describe type of biological material used (including control samples) and methods of preservation and storage.

**Assay methods**

5. Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.

**Study design**

6. State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g., by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.

7. Precisely define all clinical endpoints examined.
8. List all candidate variables initially examined or considered for inclusion in models.
9. Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.

**Statistical analysis methods**

10. Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.
11. Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.

**RESULTS**

**Data**

12. Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined, report the numbers of patients and the number of events.
13. Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.

**Analysis and presentation**

14. Show the relation of the marker to standard prognostic variables.
15. Present univariate analyses showing the relation between the marker and outcome, with the estimated effect (e.g., hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan–Meier plot is recommended.
16. For key multivariable analyses, report estimated effects (e.g., hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.
17. Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.
18. If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.

**DISCUSSION**

19. Interpret the results in the context of the prespecified hypotheses and other relevant studies; include a discussion of limitations of the study.
20. Discuss implications for future research and clinical value.

SOURCE: McShane et al. (2005).

**BOX D-2**
**The EQUATOR Network**

The EQUATOR Network is the most exhaustive source of reporting guidelines, containing an up-to-date list of health research reporting guidelines. Acting as an umbrella organization, the EQUATOR Network convenes researchers, medical journal editors, peer reviewers, developers of reporting guidelines, research funding bodies, and other collaborators with the goals of promoting transparency and accurate reporting of research studies and monitoring progress in health research reporting.

The first project of the Network was to identify all of the available guidelines for the reporting of health research studies and survey the reporting guideline authors to ascertain more information about their development methodology, dissemination and implementation strategies, problems they encountered, and impact of the guidelines. Investigators identified 37 reporting guidelines that met their inclusion criteria, of which they received 30 survey responses (81 percent response rate). The majority (73 percent) of reporting guidelines were developed by an international, multidisciplinary group, with most groups including statisticians, journal editors, clinicians, and epidemiologists. The most cited reason for developing a reporting guideline was the poor quality of reporting (87 percent), followed by the influence of other guidelines (30 percent). Survey respondents noted a dearth of funding for the development of guidelines, which may be problematic because guidelines will likely require additional work as updates are needed. Additional observations from the survey suggested that there is a need to harmonize the methodology used to develop reporting guidelines and to determine their impact on the field.

SOURCES: EQUATOR Network, 2011; Simera et al., 2008; Verbeek, 2008.

## REPORTING GUIDELINES AND THE QUALITY OF REPORTING

The evidence that the use of reporting guidelines leads to improvements in reporting is limited. Plint and colleagues (2006) conducted a systematic review to assess whether the CONSORT checklist improves the quality of RCT reporting. This review suggested that journals that adopted CONSORT had significantly better reporting of method of sequence generation (risk ratio 1.67; 95% confidence interval 1.19-2.33); allocation concealment (risk ratio 1.66; 95% confidence interval 1.37-2.00); and overall number of CONSORT items compared with non-adopters (standardized mean difference 0.83; 95% confidence interval 0.46-1.19). CONSORT adoption was not associated with improved reporting of participant flow, blinding of participants, or data analysis. A study by Smith and colleagues (2008) analyzed the quality of reporting and adherence to a modified CONSORT statement in four nursing journals. Investigators found no difference in quality of reporting among the four journals, and found that the quality of reporting of RCTs improved significantly in only one journal (*Nursing Research*) with the adoption of the CONSORT guideline (Smith et al., 2008).

Investigators have also studied whether diagnostic accuracy reporting improved after the publication of the STARD reporting guidelines, and found that the mean number of reported STARD items was 11.9 (range 3.5-19.5) in 2000, before publication of STARD, and 13.6 (range 4.0-21.0) in 2004, after the publication of STARD. Investigators reported that this was a significant increase of 1.81 items (95% confidence interval 0.61-3.01) (Smidt et al., 2006). However, a more recent analysis of the STARD reporting guideline found that the quality of

reporting diagnostic accuracy studies has remained relatively constant (Wilczynski, 2008). This analysis searched six journals adhering to STARD guidelines and six journals that did not for the years 2001, 2002, 2004, and 2005. The change in the mean total score of the modified STARD checklist (which included 13 of 25 STARD items) was analyzed using covariance. The covariance analysis found that the interaction between the two independent factors (STARD or non-STARD journal and year of publication) and the dependent variable (mean total STARD score) was not significant ($F = 0.664$, $df = 3$, partial eta(2) = 0.009, $p = 0.58$).

In the IOM (2011) report *Finding What Works in Health Care: Standards for Systematic Reviews*, the committee recommended three related standards for documenting a systematic review process that draws largely on the PRISMA reporting guideline. However, the committee emphasized that the evidence that reporting guidelines improve the quality of reporting is weak. The committee noted that the few observational studies that have evaluated this question have serious flaws, and that no controlled studies to assess whether PRISMA has improved the reporting of systematic reviews have been undertaken (IOM, 2011).

## Adherence to Reporting Guidelines

There is little agreement on who should be responsible for monitoring researchers' adherence to reporting guidelines and if adherence should be voluntary or mandatory (Vendenbroucke, 2009). The developers of the SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines have encouraged studying the implementation of reporting guidelines, stating that "the question of how publication guidelines can be used most effectively appears to us to be an empirical one, and therefore we strongly encourage editors and authors to collect, analyze, and report their experiences" (Davidoff et al., 2008, p. 675).

The 2007 IOM biomarkers committee recommended that federal funding should stipulate that researchers comply with reporting guidelines. Other possible monitors of adherence include journal editors or peer reviewers.

The International Committee of Medical Journal Editors (ICMJE) is an organization of medical journal editors (including *Annals of Internal Medicine*, *New England Journal of Medicine*, *Journal of the American Medical Association*, and others) who collaborate to produce and update the Uniform Requirements for Manuscripts (ICMJE, 2011). The ICMJE encourages authors to "consult reporting guidelines relevant to their specific research design," noting that "[r]esearch reports frequently omit important information" that reporting guidelines are meant to correct. It refers authors to the EQUATOR Network to find applicable guidelines (ICMJE, 2010).

However, the extent that journals either encourage or mandate adherence to guidelines is variable. An individual journal's policies can be located in the instruction to authors section. For example, *Science's* instruction to authors states: "We encourage compliance with MIBBI guidelines," but do not mandate their use (*Science*, 2011). The *British Medical Journal*'s (*BMJ*'s) instructions explicitly list the reporting guidelines that its authors must follow and also note that authors need to submit accompanying checklists and flowcharts in compliance with the guidelines. Its justification for this policy is "to improve *BMJ* papers' reporting and increase reviewers' understanding" of the elements that must be included in a report of a health research study (*BMJ*, 2011). In describing journals' roles in publication, Dr. Veronique Kiermer of *Nature* said, "We cannot necessarily dictate how people do research, but we can dictate how they report it, or at least we can play an important role there. Ensuring proper document[ation], based on community-driven standards and guidelines, is very important" (Kiermer, 2011).

## REFERENCES


Altman, D. G., L. M. McShane, W. Sauerbrei, and S. E. Taube. In press. Reporting recommendations for tumor marker prognostic studies (REMARK): Explanation and elaboration. *BMC Medicine*.

*BMJ* (*British Medical Journal*). 2011. *Article requirements*. http://resources.bmj.com/bmj/authors/article-submission/article-requirements (accessed May 25, 2011).

Bossuyt, P. M., J. B. Reitsma, D. E. Bruns, C. A. Gatsonis, P. P. Glasziou, L. M. Irwig, J. G. Lijmer, D. Moher, D. Rennie, H. C. de Vet, and f. t. S. group. 2004. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Family Practice* 21(1):4-10.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat Genet* 29(4):365-371.

Brundage, M. D., D. Davies, and W. J. Mackillop. 2002. Prognostic factors in non-small cell lung cancer: A decade of progress. *Chest* 122(3):1037-1057.

Burton, A., and D. G. Altman. 2004. Missing covariate data within cancer prognostic studies: A review of current reporting and proposed guidelines. *British Journal of Cancer* 91(1):4-8.

Collins, G. 2011. Consensus-based guidelines for transparent reporting. Place Published. http://blogs.bmj.com/bmj/2011/08/03/gary-collins-opening-up-multivariable-prediction-models/ (accessed December 15, 2011).

Davidoff, F., P. Batalden, D. Stevens, G. Ogrinc, S. Mooney, and SQUIRE Development Group. 2008. Publication guidelines for improvement studies in health care: Evolution of the SQUIRE project. *Annals of Internal Medicine* 149(9):670-676.

EQUATOR Network. 2011. *Welcome to the EQUATOR Network website--The resource centre for good reporting of health research studies*. http://www.equator-network.org/home/ (accessed May 19, 2011).

ICMJE (International Committee of Medical Journal Editors). 2010. *Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. Updated April 2010*. http://www.icmje.org/urm_full.pdf (accessed May 25, 2011).

ICMJE. 2011. *About the International Committee of Medical Journal Editors*. http://www.icmje.org/about.html (accessed May 25, 2011).

IOM (Institute of Medicine). 2007. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

IOM. 2011. *Finding what works in health care: Standards for systematic reviews*. Washington, DC: National Academies Press.

Kiermer, V. 2011. Publication of research involving large datasets and 'omics' technologies. Paper read at Workshop on the review of omics-based tests for predicting patient outcomes in clinical trials, March 20, Washington, DC.

McShane, L. M., D. G. Altman, W. Sauerbrei, S. E. Taube, M. Gion, and G. M. Clark. 2005. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Journal of the National Cancer Institute* 97(16):1180-1184.

Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gotzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman. 2010a. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology* 63(8):e1-e37.

Moher, D., K. Schulz, I. Simera, and D. G. Altman. 2010b. Guidance from developers of health research reporting guidelines. *PLoS Medicine* 7(2):e1000217.

Moore, H. M., A. B. Kelly, S. D. Jewell, L. M. McShane, D. P. Clark, R. Greenspan, D. F. Hayes, P. Hainaut, P. Kim, E. A. Mansfield, O. Potapova, P. Riegman, Y. Rubinstein, E. Seijo, S. Somiari, P. Watson, H.-U. Weier, C. Zhu, and J. Vaught. 2011. Biospecimen Reporting for Improved Study Quality (BRISQ). *Cancer Cytopathology* 119(2):92-101.

Plint, A. C., D. Moher, A. Morrison, K. F. Schulz, D. G. Altman, C. Hill, and I. Gaboury. 2006. Does the CONSORT checklist improve the quality of reports of randomized controlled trials: A systematic review. *The Medical Journal of Australia* 185(5):263-267.

Riley, R. D., K. R. Abrams, A. J. Sutton, P. C. Lambert, D. R. Jones, D. Heney, and S. A. Burchill. 2003. Reporting of prognostic markers: Current problems and development of guidelines for evidence-based practice in the future. *British Journal of Cancer* 88(8):1191-1198.

Schulz, K. F., D. G. Altman, and D. Moher. 2010. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *PLoS Medicine* 7(3):e1000251.

*Science.* 2011. *General information for authors*. http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml (accessed May 25, 2011).

Simera, I., D. G. Altman, D. Moher, K. F. Schulz, and J. Hoey. 2008. Guidelines for reporting health research: The EQUATOR Network's survey of guideline authors. *PLoS Medicine* 5(6):e139.

Smidt, N., A. W. Rutjes, D. A. van der Windt, R. W. Ostelo, P. M. Bossuyt, J. B. Reitsma, L. M. Bouter, and H. C. de Vet. 2006. The quality of diagnostic accuracy studies since the STARD statement: Has it improved? *Neurology* 67(5):792-797.

Smith, B. A., H. J. Lee, J. H. Lee, M. Choi, D. E. Jones, R. B. Bausell, and M. E. Broome. 2008. Quality of reporting randomized controlled trials (rcts) in the nursing literature: Application of the consolidated standards of reporting trials (consort). *Nursing Outlook* 56(1):31-37.

Vendenbroucke, J. P. 2009. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK, and CONSORT: For whom does the guideline toll? *Journal of Clinical Epidemiology* 62(6):594-596.

Verbeek, J. 2008. MOOSE CONSORT STROBE and MIAME STARD REMARK or how can we improve the quality of reporting studies. *Scandinavian Journal of Work, Environment, and Health* 34(3):165-167.

Wilczynski, N. L. 2008. Quality of reporting of diagnostic accuracy studies: No change since STARD statement of publication--before-and-after study. *Radiology* 248(3):817-823.

# Appendix E
# Committee Member Biographies

**Gilbert S. Omenn, M.D., Ph.D.** (*Chair*), is a professor of Medicine, Genetics, Public Health, and Computational Biology at the University of Michigan, where he also serves as director of the Center for Computational Medicine & Bioinformatics and the Proteomics Alliance for Cancer Research. From 1997 to 2002, Dr. Omenn was chief executive officer (CEO) and executive vice president for Medical Affairs of the University of Michigan Health System. His research interests include cancer proteomics, chemoprevention of cancers, public health genetics, science-based risk analysis, and health policy. Dr. Omenn served as associate director, Office of Science and Technology Policy, and associate director, Office of Management and Budget, in the Executive Office of the President in the Carter Administration. He was president of the American Association for the Advancement of Science in 2006. He received his M.D. magna cum laude from Harvard Medical School and his Ph.D. in Genetics from the University of Washington. His advice is sought from non-profit organizations, pharmaceutical companies, and biotechnology companies. Dr. Omenn serves on scientific advisory boards and boards of directors for several companies in the pharmaceutical/biotechnology realm.

**Catherine D. DeAngelis, M.D.,** is editor in chief emerita of the *Journal of the American Medical Association* and a professor at the Johns Hopkins University School of Medicine (Pediatrics) and the Johns Hopkins University School of Public Health (Health Services Administration). From 1990 to 1999 she was vice dean for Academic Affairs and Faculty, Johns Hopkins University School of Medicine, and from 1994 to 2000, she was editor of *Archives of Pediatrics and Adolescent Medicine*. She also has been a member of numerous journal editorial boards. She has authored or edited 11 books on Pediatrics and Medical Education and has published more than 250 peer-reviewed articles, chapters, and editorials. Most of her recent publications have focused on conflict of interest in medicine, on professionalism and integrity in medicine, on women in medicine, and on medical education. Dr. DeAngelis is a former council member of the National Academy of Sciences (NAS), Institute of Medicine (IOM); a Fellow of the American Association for the Advancement of Science; and a Fellow of the Royal College of Physicians, She has served as an officer of numerous national academic societies, including past chair of the American Board of Pediatrics and chair of the Pediatric Accreditation Council for Residency Review Committee of the American Council on Graduate Medical Education. She currently serves on the Advisory Board of the U.S. Government Accountability Office. She received her M.D. from the University of Pittsburgh's School of Medicine, her M.P.H. from the Harvard Graduate School of Public Health (Health Services Administration), and her pediatric specialty training at the Johns Hopkins Hospital. She also has been awarded seven honorary Doctorate degrees and has received numerous awards for humanitarianism and medical excellence.

**David L. DeMets, Ph.D.,** is currently professor in the Department of Biostatistics and Medical Informatics at the University of Wisconsin–Madison. Since receiving his Ph.D. from the University of Minnesota, he has been very active in the design, conduct, and analysis of clinical trials in several disease areas. Following a postdoctoral appointment at the National Institutes of Health,

or NIH (1970-1972), he spent 10 years (1972-1982) at the National Heart, Lung, and Blood Institute, where he became chief of the Biostatistics Research Branch. In 1982, he joined the University of Wisconsin School of Medicine, where he founded and chaired the Department of Biostatistics and Medical Informatics until 2009. He is currently assistant director for Biostatistics and Medical Informatics for the University of Wisconsin Institute for Clinical and Translational Research, funded in part by the NIH. He has coauthored or edited four texts, including, *Fundamentals of Clinical Trials*, now in its fourth edition, as well as texts on statistical methods for clinical trials, principles for data monitoring committees, and case studies for data monitoring. Dr. DeMets is a recognized international leader in statistical research and methods for the analysis of clinical trials. He has collaborated in the development of statistical methods for the sequential analysis of outcome data and the design of clinical trials. He has extensive national and international clinical trial experience and has served on and chaired numerous NIH and industry-sponsored Data Safety and Monitoring Committees for clinical trials in diverse disciplines. He served on the Board of Scientific Counselors of the National Cancer Institute (NCI) and Board of Directors of the American Statistical Association, and has been president of the Society for Clinical Trials and president of the Eastern North American Region of the Biometric Society. He is a fellow of the American Statistical Association, International Statistics Institute, Society for Clinical Trials, American Association for Medical Informatics, and American Association for the Advancement of Science.

**Thomas Fleming, Ph.D.,** is professor of Biostatistics and Statistics and a full member in the Division of Public Health Sciences at the Fred Hutchinson Cancer Research Center. He chaired the Department of Biostatistics from 1993 to 2006, and until 2007 directed the Biostatistics/Epidemiology Core in the University of Washington Center for AIDS Research. He also served as director of the Statistical Center for the HIV/AIDS Prevention Trial Network of the National Institute of Allergy and Infectious Diseases (NIAID) from 1993 to 2007. For 25 years, Dr. Fleming has been a Special Government Employee with the U.S. Food and Drug Administration (FDA). He has served as chair or a member of data monitoring committees for approximately 200 industry- and government-sponsored clinical trials. He was involved in the development and coordination of NIAID's national clinical trials program for the prevention and treatment of HIV infection and AIDS and is a member of the Therapeutics Research Working Group in the Office of AIDS Research.

Chief among Dr. Fleming's research interests are survival analysis, sequential analysis, and the design and analysis of clinical trials, with a special interest in regulatory issues. In 1987 he was elected a Fellow of the American Statistical Association. He has received numerous awards, including American Public Health Association's Spiegelman Award, the FDA Commissioner's Special Citation Award, the School of Public Health's Outstanding Teacher Award, and was the 2007 Greenberg Lecturer at the University of North Carolina. Dr. Fleming received his Master's Degree and Ph.D. from the University of Maryland, College Park. He has authored or coauthored several books and more than 200 research articles in peer-reviewed journals.

**Gail Geller, Sc.D., M.H.S.,** is a professor in the Department of Medicine at Johns Hopkins University, with joint appointments in the Department of Pediatrics and the Bloomberg School of Public Health's Departments of Health, Behavior & Society and Health Policy & Management. She is also co-deputy director of the Greenwall Postdoctoral Fellowship Program in Bioethics &

Health Policy in the Berman Institute of Bioethics, and is affiliated with the McKusick-Nathans Institute of Genetic Medicine.

Dr. Geller's research has centered on the ethical and psychosocial implications of genetics and genomics. She has been a member of NIH's Cancer Genetics Studies Consortium and Informed Consent Consortium, and cochaired the Task Force on Informed Consent for Cancer Susceptibility Testing. Her focus on informed consent for genetic testing stems from a broader interest in communication and decision making under conditions of uncertainty and the intrapersonal, interpersonal, cultural, and institutional forces that affect it. This overarching theme is reflected in Dr. Geller's other areas of scholarship, including the use of complementary and alternative medicine and the care of children, young adults, and families affected by chronic, life-limiting disorders (recently, she received one of the NIH "challenge" grants to explore this issue). Dr. Geller also has longstanding interests in research ethics and integrity (she is a member of the ethics core of the Johns Hopkins Clinical and Translational Science Awards program) and in medical education. Dr. Geller has served on numerous panels, including the scientific review panel for the Ethical, Legal and Social Issues Program at the NIH's National Human Genome Research Institute; and the Board of Directors of the American Society for Bioethics & Humanities. She is a Fellow of the Hastings Center. She received her B.S. from Cornell University and her Doctorate from the Johns Hopkins School of Public Health with concentrations in Bioethics and Social and Behavioral Sciences.

**Joe W. Gray, Ph.D.,** received undergraduate training in Engineering Physics from the Colorado School of Mines and a Ph.D. in Nuclear Physics from Kansas State University. He joined the Biomedical Sciences Division of the Lawrence Livermore National Laboratory, where he became increasingly active in the development of a broad range of analytic techniques useful in the study of human and model cancers. Dr. Gray became a professor of Laboratory Medicine and Radiation Oncology at the University of California, San Francisco (UCSF) in 1991 to develop clinical applications of these tools. He was interim director at the UCSF Helen Diller Family Comprehensive Cancer Center from 1995 to 1997 and served as program leader for Cancer Genetics and Breast Oncology. Dr. Gray was associate laboratory director for Life Sciences and director of the Life Sciences Division at the Lawrence Berkeley National Laboratory (LBNL) from 2003 to 2010. He joined Oregon Health and Science University (OHSU) in 2011 as the Gordon Moore Endowed Chair, where he serves as chair of the Department of Biomedical Engineering, director of the OHSU Center for Spatial Systems Biomedicine, and associate director for Translational Cancer Research in the Knight Cancer Institute. He continues as a visiting senior scientist at the LBNL and emeritus adjunct professor of Laboratory Medicine at UCSF. He is a member of the NCI Board of Scientific Advisors and the NAS Nuclear and Radiation Studies Board. Dr. Gray's current research program explores mechanisms by which genomic, transcriptional, and proteomic abnormalities occur in selected cancers; elucidates how these abnormalities contribute to cancer pathophysiology; and assesses the ways in which cells carrying these abnormalities interact with the microenvironment to influence responses to gene-targeted therapies.

Dr. Gray's work is described in more than 390 publications and in 68 U.S. patents. He has won many distinguished awards, including the United States Department of Defense's Innovator Award, Susan G. Komen Foundation's Brinker Award for Scientific Distinction, and the American Association for Cancer Research's (AACR's) Team Science Award.

**Daniel F. Hayes, M.D.,** is the clinical director of the Breast Oncology Program at the University

of Michigan Comprehensive Cancer Center (UM CCC), where he is the Stuart B. Padnos Professor of Breast Cancer Research and a professor in the Department of Internal Medicine, Division of Hematology Oncology at the University of Michigan School of Medicine. Dr. Hayes' professional training and career have been directed toward bridging the gap between laboratory and clinical research. In 1992, he assumed the role as the medical director of the Breast Evaluation Center at Dana-Farber Cancer Institute (DFCI) until 1996, when he moved to Georgetown University and spent the next 5 years establishing a successful collaboration with Dr. Marc E. Lippman. In 2001, both Drs. Lippman and Hayes joined the UM CCC to continue their translational science. Dr. Hayes serves on scientific advisory boards for biotechnology companies.

Dr. Hayes has been influential in both clinical and laboratory studies of the diagnosis and treatment of breast cancer. With his colleague, Dr. Donald Kufe, Dr. Hayes published the first reports concerning the development of the CA15-3 blood test, which is currently used world-wide to evaluate patients with breast cancer. He has become an internationally recognized leader in the use of this and other tumor markers, such as HER2. Dr. Hayes and his colleagues have also reported ground-breaking results regarding circulating tumor cells in metastatic breast cancer and the pharmacogenomics of tamoxifen. He is considered to be an expert in the field of clinical research of breast cancer, especially new hormonal and chemotherapeutic treatments, and lectures and publishes on the management of patients with breast cancer. Dr. Hayes has been chair of the Solid Tumor Correlative Sciences Committee of the Cancer and Leukemia Group B (CALGB), and now holds similar positions in the Southwest Oncology Group and the U.S. Breast Cancer Intergroup. He cochairs the Expert Panel for Tumor Marker Practice Guidelines for the American Society of Clinical Oncology (ASCO), and he is on the editorial boards of several leading cancer journals. He received a Bachelor's Degree in Biology and a Master's Degree in Biochemistry at Indiana University. He received his M.D. from the Indiana University School of Medicine, followed by a residency in Internal Medicine at the University of Texas Health Science Center at Dallas and a Fellowship in Medical Oncology at Harvard's DFCI.

**I. Craig Henderson, M.D.,** spent the first 25 years of his career conducting clinical and translational research, first at the NCI, then at DFCI from 1973 to 1992, where he founded the Breast Evaluation Center (1980), and then at UCSF, as a professor of Medicine and Chief of Medical Oncology (1992-1995). Currently, he is an adjunct professor at UCSF, where he sees patients, teaches, and participates in the Breast Cancer Specialized Program of Research Excellence (SPORE). In 1995 he became chair and CEO of SEQUUS Pharma. When it merged with Alza in 1999 he became a member of the Alza Board and a senior consultant to the company until 2001, when Alza was purchased by Johnson & Johnson. In 2000 he created a new biotechnology company, Access Oncology, Inc. (AOI) , which was merged with Keryx Biopharmaceuticals in 2004. He continued as president of the combined entity until 2008.

Highlights of Dr. Henderson's career include: principal investigator of the national trial that established paclitaxel as a treatment for early breast cancer; chair of the Early Breast Cancer Trialists Group ("Oxford Overviews") for 10 years; and chair of the Breast Committee of the CALGB from 1989 to 1995. He founded the Bay Area Breast Cancer Translational Research Program, which has received multiple SPORE awards. From 1989 to 1992, he was a member and chair of the FDA's Oncologic Drug Advisory Committee. Dr. Henderson has served on the Blue Cross Blue Shield Association Medical Advisory Panel for more than 19 years and the Medicare Coverage Advisory Committee for 3 years. During his tenure at SEQUUS, the FDA approved

two drugs, Amphotec and Doxil; Doxil was the first drug application submitted under accelerated approval regulations. AOI and Keryx in-licensed four oncology and four non-oncology drugs between 2000 and 2006, including a first-in-class Akt inhibitor, perifosine. Dr. Henderson serves on data monitoring committees for government- and industry- sponsored clinical trials; he also serves on scientific and medical advisory boards for organizations conducting technology assessments in health care. Dr. Henderson has published nearly 300 books and papers.

**Larry Kessler, Sc.D.,** is professor and chair, Department of Health Services, University of Washington. Dr. Kessler served as director of the Office of Surveillance and Biometrics, Center for Devices and Radiological Health, FDA, until 2008. From 1984 to 1995, he headed and developed the Applied Research Branch at the NCI. In 2001, he spent a year as a visiting scientist at the Fred Hutchinson Cancer Research Center and served on the faculty of the Department of Health Services. His current research focus in on cost-effectiveness and diagnostic value of medical technology in screening for cancer and other diseases. He is a collaborator on a number of national projects, including CanCORS, studies of outcomes and patterns of care of both colorectal cancer and lung cancer. The goal is to examine what kinds of treatment certain types of patients receive and how the process of obtaining the best cancer care can be improved. Dr. Kessler received his B.S. in Mathematics from Boston University and his Sc.D. in Health Services Administration from Johns Hopkins University.

**Stanley N. Lapidus, B.S.E.E.,** currently serves as president of SynapDx Corporation of Southborough, MA, a company he founded in 2009. SynapDx is developing an assay for early detection of autism based on measuring changes in RNA expression patterns derived from peripheral blood cells. In 2003, he cofounded Helicos BioSciences Corp. of Cambridge, MA, and served as CEO until 2008 and chair of its board of directors until 2010. He continues to serve as a member of the board. Helicos developed the world's first commercial single-molecule DNA sequencer. In 1995, he founded EXACT Sciences Corp. He served as president and CEO of EXACT until 2000 and then served as chair of its board of directors through 2005. EXACT develops non-invasive, DNA-based methods for early detection of colorectal cancer and its precursor lesions. In 1987 he founded Cytyc Corporation, now part of Hologic Corporation. He served as president and CEO until 1994. Cytyc developed and markets the leading Pap test for cervical cancer. Mr. Lapidus currently serves as a board member of Daktari, T2 BioSystems, and Precision Therapeutics, Inc. All are diagnostics companies. Mr. Lapidus holds more than 30 issued U.S. patents, primarily in the field of early detection and diagnosis of cancer. In addition to his entrepreneurial activities, Mr. Lapidus holds academic appointments in the Pathology Department at Tufts University Medical School and the Harvard/Massachusetts Institute of Technology Division of Health Sciences Technology. He earned a B.S.E.E. from Cooper Union in New York City in 1970. He has served as a trustee of Cooper Union since 2002.

**Debra G. B. Leonard, M.D., Ph.D.,** is professor, vice chair for Laboratory Medicine, director of the Clinical Laboratories and director of the Pathology Residency Training Program in the Department of Pathology and Laboratory Medicine at Weill Cornell Medical College and New York-Presbyterian Hospital. In 2009, Dr. Leonard was appointed chief diversity officer for Weill Cornell Medical College. She is an expert in molecular diagnostics for genetic, cancer, and infectious diseases. She is certified by the American Board of Pathology in Anatomic Pathology,

and by the American Boards of Pathology and Medical Genetics in Molecular Genetic Pathology. Currently, Dr. Leonard is a member of the IOM Roundtable on Translating Genomic-based Research for Health and chair of the Diagnostic Applications Working Group of this Roundtable. Dr. Leonard is a past member of the Secretary's Advisory Committee on Genetics, Health, and Society to the U.S. Department of Health and Human Services' Secretary Michael O. Leavitt, and a past president of the Association for Molecular Pathology. She has spoken widely on various molecular pathology test services, the future of molecular diagnostics, and the impact of gene patents on molecular pathology practice and patient access to molecular diagnostic testing services. Dr. Leonard occasionally provides advice to pharmaceutical and medical instrument companies on specific projects. She is the editor of two Molecular Pathology textbooks. Most recently, Dr. Leonard received the 2009 Leadership Award from the Association for Molecular Pathology. Her M.D. and Ph.D. were completed at the New York University School of Medicine, where she also did her postgraduate clinical training in Anatomic Pathology, including a Surgical Pathology Fellowship.

**Harold Moses, M.D.,** is the director emeritus of the Vanderbilt-Ingram Cancer Center; Hortense B. Ingram Professor of Molecular Oncology; professor of Cancer Biology, Medicine, and Pathology; and the founding and current director of the Frances Williams Preston Laboratories. After residency training in Pathology at Vanderbilt and postdoctoral research training at the NIH, he spent 5 years as a faculty member in Pathology at Vanderbilt and 12 years at the Mayo Clinic, the last 6 of which were as chair of the Department of Cell Biology. He returned to Vanderbilt 25 years ago as professor and chair of the Department of Cell Biology in the School of Medicine. Fifteen years ago he became the Founding Director of the Vanderbilt Cancer Center with a concurrent appointment as the B.F. Byrd, Jr., Professor of Clinical Oncology. He resigned as chair of the Department of Cell Biology in 1998 to devote more time to the Cancer Center, now named the E. Bronson Ingram Cancer Center. At the end of 2004, he became director emeritus of the Vanderbilt-Ingram Cancer Center and the Hortense B. Ingram Professor of Medical Oncology. Dr. Moses graduated from Berea College and then earned an M.D. from Vanderbilt University School of Medicine.

**William Pao, M.D., Ph.D.,** is associate professor of Medicine, Cancer Biology, and Pathology/Microbiology/Immunobiology at Vanderbilt University and the Ingram Associate Professor of Cancer Research, director of the Division of Hematology/Oncology, and director of Personalized Cancer Medicine at Vanderbilt-Ingram Cancer Center. He has developed a basic and translational research program that made seminal contributions to the understanding of molecular mechanisms in lung cancer pathogenesis. His work has identified new molecular mechanisms of sensitivity and resistance of lung cancers to EGFR tyrosine kinase inhibitors and has yielded important insights into a molecular understanding of lung adenocarcinoma in never smokers. Based on these discoveries, he has developed and successfully tested new anticancer therapies in animal models and humans. His work has helped change the standard of care in lung cancer. More recently, his laboratory has established a high-throughput screen to identify kinase fusions in cancers using next-generation sequencing technologies. Dr. Pao has received multiple honors and awards, including an ASCO Young Investigator Award, a Clinical Scientist Development Award from the Doris Duke Charitable Foundation, a V Foundation grant, the Hope Now Award from the Joan's Legacy Foundation, and a Stand Up To Cancer Innovative Grant Award from the AACR. He was elected to the American Society for Clinical Investigation

in 2011. Dr. Pao obtained his M.D. and Ph.D. at Yale University, did his residency training in Internal Medicine at New York Presbyterian Hospital-Weill Cornell Campus, and completed his medical oncology Fellowship training at Memorial Sloan-Kettering Cancer Center.

**Rebecca D. Pentz, Ph.D.,** is professor of Hematology and Oncology in Research Ethics and professor of Pediatrics at the Emory School of Medicine, as well as Faculty for the Emory Ethics Center in Atlanta. She does empirical ethics research on issues such as informed consent, Phase I research (first use of a drug in humans), and genetic confidentiality, as well as helping researchers with their protocols to make them ethically sound. Dr. Pentz embeds ethics companion studies in most of the major grants from the Winship Cancer Center at Emory and is working closely with Emergency Medicine in research that uses the Exception from informed consent. She also is active in pediatrics, having recently completed a multisite NCI-funded study on family decision making in pediatric bone marrow transplant, and serving on multiple Children's Oncology Group committees. Her extensive empirical research forms the basis of her international speaking and professional publications. Dr. Pentz sits on several key committees at Emory, including the Scientific Review Committee for all cancer research and the Faculty Advisory Committee to the Dean of the Medical School. She is actively involved in education at the Cancer Institute as the Course Director for Fundamentals of Clinical Research for second year oncology Fellows and mentor of the postbaccalaureate Ethics Fellow. In 2000, she moved to Atlanta from Houston, where she directed the Clinical Ethics Program at the MD Anderson Cancer Center. As a clinical ethicist, she worked closely with patients and families, offering help for those struggling with end-of-life issues.

Dr. Pentz chairs the Informed Consent Task Force for the HHS Secretary's Advisory Committee on Blood Stem Cell Transplantation. She is on the External Scientific Panel of the NIH Genotype-Tissue Expression project. She is a member of both the St. Jude and the Bone Marrow Transplant Clinical Trials Network's Data Safety Monitoring Boards and a member of a CDC IRB. She is coeditor of the Ethics Section of *The Oncologist.* She was on the Ethical, Legal, and Social Issues Subgroup in support of the cancer Human Biobank.

**Nathan Price, Ph.D.,** is an associate professor at the Institute for Systems Biology (ISB) in Seattle, WA. He is also an affiliate associate professor in the Departments of Bioengineering and Computer Science & Engineering at the University of Washington, where he advises graduate students as a member of the Graduate College. Prior to moving to ISB, he was an assistant professor at the University of Illinois at Urbana-Champaign from 2007 to 2011, where he continues to hold adjunct appointments in the Department of Chemical and Biomolecular Engineering and the Institute for Genomic Biology. In 2006, Dr. Price was named one of the inaugural "Tomorrow's PIs" as a "rising young investigator" in systems biology by Genome Technology, and in 2008 was the recipient of the NIH Howard Temin Pathway to Independence Award in Cancer Research. In 2009, he received a National Science Foundation CAREER Award to use system biology approaches to guide genome-scale synthetic biology efforts. In 2010, he received the Young Investigator Award from the Roy J. Carver Charitable Trust for his work to build genome-scale biomolecular network models of human glioblastoma (brain cancer). In 2011, he was one of 13 chemical scientists to be named a Camille-Dreyfus Teacher-Scholar by the Dreyfus Foundation. Dr. Price served on the steering committee of the Mayo Clinic-University of Illinois Alliance for Technology-Driven Medicine from 2010 to 2011. He now serves on the scientific advisory board of TetraVitae Bioscience, Inc., as well as

on the Board of Directors and Scientific Advisory Board of the P4 Medicine Institute. He is an associate editor of *BMC Systems Biology* and *Biotechnology Journal*, and a deputy editor in chief of *PLoS Computational Biology*.

**John Quackenbush, Ph.D.,** earned his Ph.D. in Theoretical Particle Physics from the University of California, Los Angeles, then completed a postdoctoral Fellowship in experimental high-energy physics. After receiving a Fellowship from the National Center for Human Genome Research, he worked with Glen Evans on the physical mapping of human chromosome 11, and later with Richard Myers and David Cox on large-scale DNA sequencing of chromosomes 21 and 4. In 1998 he joined the faculty at The Institute for Genomic Research, where his work focused on the use of genomic and computational methods for the study of human disease. In 2005 he joined DFCI, where his work has increasingly focused on the analysis of women's cancers, although the methods he and his group develop can be broadly applied. In 2009 he launched the Center for Cancer Computational Biology, a Dana-Farber Strategic Plan Center focused on providing computational support more broadly to the DFCI research community. Dr. Quackenbush occasionally provides advice to biotechnology companies on specific projects.

**Elda Railey** is cofounder of Research Advocacy Network, a non-profit group dedicated to advancing patient-focused research. She serves as editor of the e-newsletter, *Network News*, and director of the Advocate Institute™, an innovative concept and learning environment for patient advocates. Research Advocacy Network has published a tutorial series for advocates, including: Genomics in Cancer, Biomarkers in Cancer, Pathology and Tissue Research, and Understanding Clinical Trial Design. She served for 12 years with Susan G. Komen for the Cure in various capacities, primarily as director of Grants and Sponsored Programs, where she led one of the largest programs for private funding of breast cancer research.

She has reviewed for Komen, Avon-NCI Partners in Progress, and the CDC, and is a member of the Intercultural Cancer Council and the External Advisory Board of the UCSF Breast Cancer SPORE. She has served on the National Steering Committee for Redes En Acción: National Latino Cancer Research Network and was the team leader for the Insurance team for EDICT (Eliminating Disparities in Clinical Trials). She was honored with the Art of Advocacy award from the Genetic Alliance in 2005. As a consultant with the Coalition of Cooperative Groups, she was program coordinator and contributing author on *Cancer Research: A Guide to Clinical Trials Modules on Drug Development, Surgical and Radiation Therapies and Tissue and Its Use for the Advocate Training Program.* She has served on the Patient Advocate Committee for American College of Surgeons Oncology Group. Currently, she serves on NCI's Early Detection Research Network and Clinical Proteomics Technologies for Cancer. She is also a patient advocate for the Lung SPORE at UT Southwestern/MD Anderson Cancer Center.

**David F. Ransohoff, M.D.,** is professor of Medicine and clinical professor of Epidemiology at the University of North Carolina (UNC) at Chapel Hill. He graduated from Harvard College and Case Western Reserve School of Medicine and did his internship and residency at Dartmouth-Hitchcock. As a Fellow in the then-new field of clinical epidemiology, the field concerned with methods of clinical research to study diagnosis and prognosis, he studied with Alvan Feinstein in Yale's Robert Wood Johnson Clinical Scholars Program and wrote a paper in the *New England Journal of Medicine* that helped establish the field of methods to evaluate diagnostic tests. After

a gastroenterology Fellowship at the University of Chicago, Dr. Ransohoff served on the faculties of Case Western Reserve University, Yale University, and then UNC, conducting research in the diagnosis and management of gastrointestinal and other clinical problems. About 15 years ago, Dr. Ransohoff extended his work in clinical epidemiology and diagnosis to include molecular markers for cancer and now collaborates with a number of NCI groups, including the Early Detection Research Network and the Clinical Proteomic Technology Assessment for Cancer. At UNC, Dr. Ransohoff has directed the Robert Wood Johnson Clinical Scholars Program and the NIH-funded K30 clinical research curriculum faculty development program.

**E. Albert Reece, M.D., Ph.D., M.B.A.,** is vice president for Medical Affairs, University of Maryland, Baltimore and the John Z. and Akiko K. Bowers Distinguished Professor and Dean of the University of Maryland School of Medicine. He is also a professor in the departments of Obstetrics and Gynecology, Medicine, and Biochemistry & Molecular Biology. He completed an internship and residency in obstetrics and gynecology at Columbia University/Presbyterian Hospital, and a postdoctoral fellowship in Maternal-Fetal Medicine at Yale University School of Medicine. He remained as faculty at Yale for nearly 10 years, achieving accelerated promotion to associate professor in 1987. In 1990, went to Temple University to serve as the Abraham Roth Professor and Chair of the Department of Obstetrics, Gynecology and Reproductive Sciences. Between 2001 and 2006, he served as vice chancellor for Medical Sciences at the University of Arkansas and dean of the College of Medicine. He moved to his current in 2006.

He has published extensively in the scientific literature: 11 books; 5 monographs; and more than 500 articles, chapters, and abstracts. He is a member of many national scientific organizations. He served as chair of the Association of American Colleges' Council of Deans. He serves or has served on many governmental and civic organizations and committees, such as the FDA, the IOM, the NIH, the HHS Secretary's Committee on Infant Mortality, The March of Dimes Birth Defects Foundation, the Massachusetts General Hospital Scientific Advisory Committee, the Board (chair) of the Nelly Berman Classical Music Institute, and the Agnes Irwin School for Girls. He has received numerous special recognitions and awards, including the Distinguished Leadership Award in 2009 and the 2010 Berson Medical Alumni Achievement Award in Health Sciences from his alma mater, New York University School of Medicine. Originally from Jamaica, West Indies, Dr. Reece completed a B.S. magna cum laude from Long Island University; an M.D. from New York University School of Medicine; a Ph.D. in Biochemistry from the University of the West Indies, Kingston, Jamaica; and an M.B.A. from the Fox School of Business & Management of Temple University.

**Daniela M. Witten, Ph.D.,** is an assistant professor in the Department of Biostatistics at the University of Washington, an adjunct assistant professor in the Department of Statistics at the University of Washington, and an affiliate investigator at the Fred Hutchinson Cancer Research Center. Her research involves the development of statistical tools for the analysis of large-scale biological datasets, such as gene expression, DNA copy number, and DNA sequence data. As a graduate student, Dr. Witten held a National Defense Science and Engineering Graduate Fellowship (2006-2009) and won the Gertrude Cox Scholarship from the American Statistical Association (2008). She has held the Genentech Endowed Professorship in Biostatistics at the University of Washington (2010-2011). Recent awards include the David Byar Young Investigator Award from the American Statistical Association (2011) and the NIH Director's Early Independence Award (2011-2016). Dr. Witten currently serves as an associate editor for the *Journal of Computational and Graphical Statistics*. Dr. Witten did her training at Stanford

University, where she received a Bachelor's Degree in Mathematics and Biology with Honors and Distinction, and a Doctorate in Statistics.

# Appendix F
# Information Gathering Sessions and Speakers

### COMMITTEE MEETING 1 – DECEMBER 20, 2010

**ROBERT L. BECKER, JR.,** Chief Medical Officer, Office of In Vitro Diagnostic Device Evaluation and Safety, Center for Devices and Radiological Health, Food and Drug Administration

**LISA MCSHANE,** Mathematical Statistician, Biometric Research Branch, National Cancer Institute

### COMMITTEE MEETING 2 – MARCH 30 - 31, 2011

**KEITH BAGGERLY,** Associate Professor of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center

**NED CALONGE,** Chair, Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group, and President and CEO, The Colorado Trust

**KATRINA KELNER,** Editor, *Translational Medicine*, and Managing Editor, Research Journals, *Science Magazine*

**VÉRONIQUE KIERMER,** Executive Editor and Head of Researcher Services, Nature Publishing Group

**SUMITHRA MANDREKAR,** Associate Professor of Biostatistics, Division of Biomedical Statistics and Informatics, Mayo Clinic

**JOSEPH NEVINS,** Barbara Levine University Professor of Breast Cancer Genomics, Institute for Genome Sciences and Policy, Duke University

**HAROLD PAZ,** Chief Executive Officer, Penn State Milton S. Hershey Medical Center, and Senior Vice President for Health Affairs and Dean, College of Medicine, Pennsylvania State University

**CHARLES PEROU,** Professor of Genetics and Pathology, and Member, Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill

**PETER PRONOVOST,** Director, Division of Adult Critical Care Medicine, Director, Quality and Safety Research Group, and Medical Director, Center for Innovations in Quality Patient Care, Johns Hopkins University

**STEVEN SHAK,** Chief Medical Officer, Genomic Health

**RICHARD SIMON,** Chief, Biometric Research Branch, National Cancer Institute

**LAURA VAN'T VEER,** Director, Applied Genomics, and Program Leader, Breast Oncology, Helen Diller Family Comprehensive Cancer Center

**SCOTT ZEGER,** Professor of Biostatistics and Vice Provost for Research, Bloomberg School of Public Health, Johns Hopkins University

## COMMITTEE MEETING 3 – JUNE 30 - JULY 1, 2011

**LISA MCSHANE,** Mathematical Statistician, Biometric Research Branch, National Cancer Institute

## TELECONFERENCE 1 – AUGUST 4, 2011

**WILLIAM SELLERS,** Vice President and Global Head of Oncology, Novartis Institutes for BioMedical Research

## TELECONFERENCE 2 – AUGUST 15, 2011

**JEFFREY DRAZEN,** Editor in Chief, *New England Journal of Medicine*

## TELECONFERENCE 3 – AUGUST 19, 2011

**ALBERTO GUTIERREZ,** Director, Office of In Vitro Diagnostic Device Evaluation and Safety, Center for Devices and Radiological Health, Food and Drug Administration

## PANEL DISCUSSION WITH DUKE UNIVERSITY FACULTY AND ADMINISTRATION – AUGUST 22, 2011

**WILLIAM BARRY,** Assistant Professor of Biostatistics and Bioinformatics

**ROBERT CALIFF,** Vice Chancellor for Clinical Research and Director, Duke Translational Medicine Institute

**MICHAEL CUFFE,** Vice President for Ambulatory Services, Chief Medical Officer for Duke University Health System

**JOHN FALLETTA,** Professor of Pediatrics and Senior Chair, Institutional Review Board

**GEOFFREY GINSBURG,** Director of Genomic Medicine, Institute for Genome Sciences and Policy

**MICHAEL KELLEY,** Associate Professor of Medicine, Oncology, and Cancer Protocol Review Committee

**SALLY KORNBLUTH,** James B. Duke Professor of Pharmacology and Cancer Biology and Vice Dean for Research, Duke University Medical Center

**ROSS MCKINNEY,** Director, Clinical and Translational Research Ethics, Law, and Policy, Professor of Pediatrics, and Director, Trent Center for Bioethics, Humanities, and Medical History

# Abbreviations and Acronyms

| | |
|---|---|
| 21-Gene RS | 21-Gene Recurrence Score |
| ACR | acute cellular rejection |
| AHRQ | Agency for Healthcare Research and Quality |
| ASCO | American Society of Clinical Oncology |
| BCBS | Blue Cross Blue Shield |
| CAD | coronary artery disease |
| CAP | College of American Pathologists |
| CARGO | Cardiac Allograft Rejection Gene Expression Observational |
| CDC | Centers for Disease Control and Prevention |
| CLIA | *Clinical Laboratory Improvement Amendments of 1988* |
| CMF | cyclophosphamide, methotrexate, and fluorouracil |
| CMS | Centers for Medicare & Medicaid Services |
| CMV | cytomegalovirus |
| COI | conflict of interest |
| CONSORT | Consolidated Standards of Reporting Trials |
| CPRC | Cancer Center Protocol Review Committee |
| CTAF | California Technology Assessment Forum |
| CTMS | clinical trial management systems |
| CTSA | Clinical Translational Science Award |
| dbGaP | database of Genotypes and Phenotypes |
| DEAL | DNA-encoded antibody libraries |
| DNA | deoxyribonucleic acid |
| DoD | Department of Defense |
| DSMB | data safety monitoring board |
| EGAPP | Evaluation of Genomic Applications in Practice and Prevention |
| EMB | endomyocardial biopsy |
| EQUATOR | Enhancing the QUAlity and Transparency Of health Research |
| FDA | Food and Drug Administration |
| FISH | fluorescence in situ hybridization |
| FFATA | *Federal Funding Accountability and Transparency Act* |
| FFPE | formalin-fixed, paraffin-embedded |
| HER2 | human epidermal growth factor receptor 2 |
| HHS | U.S. Department of Health and Human Services |

| | |
|---|---|
| HPV | human papillomavirus |
| ICMJE | International Committee of Medical Journal Editors |
| IDE | investigational device exemption |
| IDMC | independent data monitoring committee |
| IHC | immunohistochemistry |
| IMAGE | Invasive Monitoring Attenuation through Gene Expression |
| IND | investigational new drug |
| IOM | Institute of Medicine |
| IOTF | Interagency Oncology Task Force |
| IRB | Institutional Review Board |
| ISHLT | International Society for Heart and Lung Transplantation |
| ISO | International Organization for Standardization |
| IVD | in vitro diagnostic |
| IVDMIA | in vitro diagnostic multivariate index assay |
| LDA | linear discriminate analysis |
| LDT | laboratory-developed test |
| LMS | Lung Metagene Score |
| MALDI | matrix-assisted laser desorption/ionization |
| MALDI-TOF | MALDI time-of-flight mass spectrometer |
| MF | methotrexate and fluorouracil |
| MIAME | Minimum Information About a Microarray Experiment |
| MIBBI | Minimum Information for Biological and Biomedical Investigations |
| MINDACT | Microarray In Node negative and 1-3 positive lymph node Disease may Avoid ChemoTherapy |
| miRNA | micro RNA |
| MoS | *Manual of Style: A Guide for Authors and Editors* |
| MRI | magnetic resonance imaging |
| mRNA | messenger RNA |
| MS | mass spectrometry |
| NAS | National Academies of Science |
| NCCN | National Comprehensive Cancer Network |
| NCI | National Cancer Institute |
| ncRNA | non-coding RNA |
| NIH | National Institutes of Health |
| NIST | National Institute of Standards and Technology |
| NMR | nuclear magnetic resonance |
| NOCEDP | National Ovarian Cancer Early Detection Program |
| NRC | National Research Council |
| NSABP | National Surgical Adjuvant Breast and Bowel Project |
| OHRP | Office of Human Research Protections |

| | |
|---|---|
| ORI | Office of Research Integrity |
| PBMC | peripheral blood mononuclear cells |
| PI | principal investigator |
| PHS | Public Health Service |
| PMA | premarket approval |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| PSA | prostate-specific antigen |
| QMS | quality management systems |
| qRT-PCR | quantitative reverse-transcriptase polymerase chain reaction |
| RCT | randomized controlled trial |
| REMARK | Reporting Recommendations for Tumor Marker Prognostic Studies |
| RNA | ribonucleic acid |
| RNAseq | RNA sequencing |
| rRNA | ribosomal RNA |
| RT-PCR | reverse-transcriptase polymerase chain reaction |
| RxPONDER | Rx for POsitive Node, Endocrine Responsive breast cancer |
| SAM | significance analysis of microarrays |
| SELDI | surface-enhanced laser desorption/ionization |
| SELDI-TOF | SELDI time-of-flight mass spectrometer |
| SERMs | selective estrogen receptor modulators |
| SGO | Society of Gynecologic Oncology |
| SNP | single nucleotide polymorphism |
| SOP | standard operating procedure |
| SPCI | Simone Protective Cancer Institute |
| SQUIRE | Standards for QUality Improvement Reporting Excellence |
| SRM | selected reaction monitoring |
| SS | similarity score |
| STARD | STAndards for the Reporting of Diagnostic accuracy studies |
| STROBE | STrengthening the Reporting of OBservational studies in Epidemiology |
| TAILORx | Trial Assigning IndividuaLized Options for Treatment |
| TEC | Blue Cross Blue Shield Technology Evaluation Center |
| TMQF | Translational Medicine Quality Framework |
| tRNA | transfer RNA |
| USPSTF | U.S. Preventive Services Task Force |

# Glossary

**Accrual (in clinical trials)—**the enrollment of qualified patients into clinical trials.

**Acute cellular rejection (ACR)—**when a transplanted organ is not accepted by the body of the organ recipient.

**Adjuvant therapy—**additional cancer treatment given after the primary treatment to lower the risk that the cancer will return. May include chemotherapy, radiation therapy, hormone therapy, targeted therapy, or biological therapy.

**Adnexal mass—**a lump in tissue near the uterus.

**Analyte—**a substance that is the subject of analysis.

**Analytical validation—**traditionally, "assessing [an] assay and its measurement performance characteristics, determining the range of conditions under which the assay will give reproducible and accurate data." With respect to omics, assessing a test's "ability to accurately and reliably measure the … analyte[s] … of interest in the clinical laboratory, and in specimens representative of the population of interest."

**Anastrozole (Arimidex)—**a drug that inhibits estrogen synthesis. This drug, an aromatase inhibitor, may inhibit tumor growth in some breast cancers.

**Angiography—**X-ray visualization of blood vessels that have been injected with a radiographic contrast dye.

**Apoptosis—**a type of cell death in which a series of molecular steps in a cell leads to its death. This is the body's normal way of getting rid of unneeded or abnormal cells.

**Archival tissue—**biological specimens collected from patients and stored for possible future use in medical care or research

**Area under the receiver operating curve (AUC)—**a measure of the ability of a test to accurately discriminate a result indicating a particular disease state from a result not indicating that disease state.

**Aromatase inhibitor—**a drug that prevents the formation of the hormone estradiol. This drug is a type of hormone therapy for postmenopausal women with certain types of breast cancer.

**Assay—**a biochemical or other measurement developed to quantify a biomarker.

**Baseline corrected—**allows for the removal of background "noise" or unnecessary peaks by running a blank set of samples that are subtracted from the data.

**Batch effects—**groups of measurements with different testing results because of variability in conditions such as testing on different days or by different equipment operators, rather than scientific or biological differences between samples.

**Bias—**the systematic but unintentional erroneous association of some characteristic with a group in a way that distorts a comparison with another group.

**Bioinformatics—**a field of study focused on developing fast, efficient algorithms for data reduction, data mining, and literature search techniques and developing biologically informative annotations related to DNA/RNA sequence, gene/protein expression, or the interaction of pathways, networks, phenotypes, and druggable targets.

**Biological plausibility—**data elucidating how the biological pathways leading from exposure to effect are useful.

**Biological products (biologics)—**a category of products regulated by the Food and Drug Administration (FDA), including vaccines, blood and blood components, allergenic compounds, somatic cells, gene therapy, tissues, and recombinant therapeutic proteins.

**Biomarker—**"a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n] . . . intervention." (IOM, 2010)

**Biopsy—**the removal of tissues or cells so they can be examined by a pathologist.

**Biostatistics—**a field of study focused on applying experimental design and data analysis to a wide range of topics in biology.

**Blinding** [in a **controlled trial**]—the process of preventing those involved in a trial from knowing the comparison group to which a particular participant belongs. The risk of **bias** is minimized when as few people as possible know who is receiving the **experimental intervention** and who the **control** intervention. Participants, caregivers, outcome assessors, and analysts are all candidates for being blinded. Blinding of certain groups is not always possible; for example, if treatment involves active patient participation, such as attending a therapy session, the participant cannot be blinded to the type of treatment provided.

**CA125 (also called cancer antigen 125)—**a molecule that may be found in high levels in the blood of patients with certain types of cancer, such as ovarian cancer. CA125 levels may be an indicator of how well cancer treatments are working and whether a cancer will return.

**Candidate omics-based test—**an omics-based test in the test discovery and development phase.

**Chemotherapy—**treatment with drugs that kill cancer cells.

***Clinical Laboratory Improvement Amendments*** **(CLIA)—**amendments passed by Congress in 1988 that established quality standards for all non-research laboratory testing performed on specimens derived from humans for the purpose of providing information for the diagnosis, prevention, and/or treatment of disease, or impairment of or assessment of health. CLIA established quality standards for laboratories to ensure the accuracy, reliability, and timeliness of patient test results regardless of where the test is performed.

**Clinical/biological validation—**validation assessing a test's "ability to accurately and reliably predict the clinically defined disorder or phenotype of interest."

**Clinical endpoint—**a characteristic or variable that reflects how a patient feels, functions, or survives in response to an intervention.

**Clinical trial—**a formal study carried out according to a prospectively defined protocol that is intended to discover or verify the safety and effectiveness of procedures or interventions in humans.

**Clinical use stage—**the last phase in the test development process in which a validated omics-based test is assessed for use in patient management decisions.

**Clinical utility—**"evidence of improved measurable clinical outcomes, and [a test's] usefulness and added value to patient management decision making compared with current management without [omics] testing."

**Completely randomized trial design—**in this report, a test result is not used in the randomization of patients to different therapies, nor is patient accrual stratified according to the test results.

**Confidence interval—**a measure of the uncertainty around the main finding of a statistical analysis. Estimates of unknown quantities, such as the odds ratio comparing an **experimental intervention** with a **control,** are usually presented as a point estimate and a 95% confidence interval. This means that if someone were to keep repeating a study in other samples from the same **population**, 95% of the confidence intervals from those studies would contain the true value of the unknown quantity. Alternatives to 95%, such as 90% and 99% confidence intervals, are sometimes used. Wider intervals indicate lower precision; narrow intervals, greater precision.

**Confirmation—**in this report, verifying a candidate omics-based test on an independent sample set before the test validation phase.

**Conflict of interest—**"a set of circumstances that creates a risk that professional judgment or actions regarding a primary interest will be unduly influenced by a secondary interest."

**Confounding effects—**a situation in which an intervention effect is biased because of some difference between the comparison groups apart from the planned interventions, such as baseline characteristics, prognostic factors, or concomitant interventions.

**Coronary artery disease (CAD)—**damage to the heart caused by atherosclerotic constriction of arteries that supply blood to the heart.

**Cross-validation—**a statistical method for preliminary confirmation of a model's performance using a single data set, by dividing the data into multiple segments, and iteratively fitting the model to all but one segment and then evaluating its performance on the remaining segment.

**Cyclophosphamide—**a synthetic, chemotherapeutic agent with antineoplastic and immunosuppressive activities.

**Diagnosis—**a conclusion as to the presence of a disease.

**Diagnostic test—**the investigative tools and techniques used in biological studies to identify or determine the presence of a disease or other condition. Any laboratory-based test that can be used in drug discovery and development as well as in patient care and clinical decision making.

**Discovery phase—**the first phase in the omics-based test development process during which candidate omics-based tests are first identified and confirmed on an independent set of specimens, if available.

**Distant recurrence—**occurs when a cancer has metastasized to another location in the body following initial cancer treatment and remission.

**Effect modifier—**a measure that identifies patients who are most likely to be sensitive or resistant to a specific treatment regimen or agent. An effect modifier is particularly useful when that measure can be used to identify the subgroup of patients for whom treatment will have a clinically meaningful and favorable benefit-to-risk profile.

**Endocrine therapy—**treatment that aids, blocks, or removes hormones.

**Enrichment trial design—**the only patients entered into the clinical trial are those with positive test results at screening. These patients are randomized and/or treated.

**Epigenome—**the complete set of epigenetic modifications, which are heritable or transitory changes in phenotype or gene expression that result from mechanisms other than changes in the DNA sequence in a given individual, tissue, tumor, or population.

**Estrogen receptor (ER)—**a protein found in cells of reproductive tissue, some other types of tissue, and some cancer cells. The hormone estrogen binds to the receptor and may cause cells to grow.

**False negative—**the error of failing to observe a difference when in truth there is one.

**False positive—**occurs when a difference is observed even though in truth there is none.

**FDA approval—**the FDA can approve a device after reviewing a sponsor's premarket approval (PMA) application that has been submitted to the FDA. To acquire approval of a device through a PMA application, the applicant must provide reasonable assurance of the device's safety and effectiveness.

**FDA clearance—**the FDA can clear a device after reviewing a sponsor's premarket notification, otherwise known as a 510(k) (named for a section in the *Food, Drug, and Cosmetic Act*), that has been filed with the FDA. To acquire clearance to market a device using the 510(k) pathway, the 510(k) applicant must show that the medical device is "substantially equivalent" to a device that is already legally marketed for the same use.

**Fluorouracil (5-FU or F5U)—**an antimetabolite drug used in cancer treatment. The drug may kill cancer cells by stopping them from making DNA.

**Formalin-fixed, paraffin-embedded tissue—**a tissue sample that has been preserved to enable pathological or molecular analysis.

**Fully specified computational procedure—** all component steps of the procedure—namely, all data processing steps, normalization techniques, weights, parameters, and other aspects of the model, as well as the precisely defined mathematical formula or formulas used to convert the data into a prediction of the phenotype of interest—are completely formulated in writing.

**Genome—**the complete sequence of DNA in a cell or organism.

**Hazard ratio—**an expression of the risk of an event in one arm of a study as compared to the risk of the event happening in the other arm over time. This differs from the relative risk ratio, which is a proportion of the number of events that occur in one arm of the study as compared to the other arm.

**High-dimensional data—**large datasets characterized by the presence of many more variables than observations, such as datasets that result from measurements of hundreds to thousands of molecules in a relatively small number of biological samples. The analysis of such datasets requires appropriate computing power and statistical methods.

**Histopathology—**examination of tissue samples in order to understand disease processes in the organism from which the samples were obtained.

**Hormonal therapy—**see Endocrine therapy.

**Human epidermal growth factor receptor 2 (HER2)—**a growth factor receptor that is used as a breast cancer biomarker for prognosis and treatment with the drug trastuzumab (Herceptin), which targets the protein.

**In vitro device—**a test that can detect disease, infection, or other health conditions.

**Institutional Review Board (IRB)—**an institutional oversight body that protects human safety, privacy, and autonomy and ensures informed consent.

**Intended use—**a statement describing a device's intended application, taking into account whether such use could harm the patient or consumer. The manufacturer's intended use should be clearly marked on printed and graphic materials for proposed labels and promotional claims.

**Interpretation criteria—**a component of the validation process that describes the precise mathematical function used for interpretation of the assay results to ensure that it performs well on the assay results yielded by the test method in the clinical laboratory.

**Investigational device exemption (IDE)—**an FDA designation that allows an investigational device to be used in a clinical study to collect safety and effectiveness data supporting a premarket approval application or a premarket notification submission.

**Laboratory-developed tests (LDTs)—**laboratory tests used in patient care that have been developed and are performed in a CLIA-certified clinical laboratory, but have not been reviewed by the FDA.

**Lipidome—**the complete set of lipids in a biological sample.

**Locked-down—**in this report, various aspects of the test, such as the precisely defined series of computational steps performed in processing the raw data, the mathematical formula or formulas used to convert the data into a prediction of the phenotype of interest, and the clinical assay for measuring the selected features are recorded and no longer changed at specific points in the development process.

**Mass spectrometry—**a method for separating ionized molecular particles according to mass by applying a combination of electrical and magnetic fields to deflect ions passing in a beam through the instrument.

**Mechanism of action—**the biological pathway by which a drug affects its target in the body.

**Medical device—**any instrument, apparatus, appliance, material, or other article intended to be used to affect the structure or any function of a human or animal body.

**Metabolome—**the complete set of small molecular metabolites found with a biological sample (including metabolic intermediates in carbohydrate, lipid, amino acid, nucleic acid, and other biochemical pathways, along with hormones and other signaling molecules).

**Metadata—**information about a dataset and how it was generated.

**Methotrexate—**a drug with antineoplastic and immunosuppressant activities that have the effect of inhibiting synthesis of DNA and RNA.

**Microarray—**a high-throughput biological assay in which different probes are deposited on a chip surface (glass or silicon) in a miniature arrangement.

**Multivariate model—**measuring the impact of more than one variable at a time while analyzing a set of data, for example, looking at the impact of age, sex, and occupation on a particular outcome.

**Negative predictive value (NPV)—**the probability that an individual with a negative test result is truly unaffected and/or does not have the particular disease or characteristic that the test is designed to detect.

**Omics—**scientific disciplines comprising study of related sets of biological molecules. Examples of omics disciplines include genomics, transcriptomics, proteomics, metabolomics, and epigenomics.

**Omics-based test—**an assay composed of or derived from many molecular measurements and interpreted by a fully specified computational model to produce a clinically actionable result.

**Overfitting—**a false pattern that is found between large numbers of possible predictors and an outcome due to high complexity and "noise" in the data. Overfitting leads to erroneous conclusions about the data. This can be identified by checking the reproducibility in a separate, independent group of individuals.

**Overlap (in datasets)—**the use of the same samples in more than one phase of test development.

**Patient management—**decisions about the care and treatment of individual patients, based on information about their disease status and history.

**Performance characteristic—**the sensitivity, accuracy, and specificity of an omics-based test.

**Phase I clinical trial—**clinical trial in a small number of patients in which the toxicity and dosing of an intervention are assessed.

**Phase II clinical trial—**clinical trial in which the safety and preliminary efficacy of an intervention are assessed in patients.

**Phase III clinical trial—**large-scale clinical trial in which the safety and efficacy of an intervention are assessed in a large number of patients. The FDA generally requires new drugs to be tested in Phase III trials before they can be put on the market.

**Phenotype—**the physical traits of an individual.

**Positive predictive value (PPV)—**the probability that an individual with a positive test result has, or will develop, the particular disease or characteristic that the test is designed to detect. It is a measure of the ratio of true positives to (false + true positives).

**Preanalytical variables**—aspects of sample collection and handling that need to be standardized and documented prior to test development and use.

**Predictive factor**—an effect modifier.

**Premarket approval (PMA)**—an FDA approval for a new test or device that enables it to be marketed for clinical use. To receive this approval, the manufacturer of the product must submit the clinical data showing the product is safe and effective for its intended use.

**Premarket notification or 510(k)**—an FDA review process that enables a new test or device to be marketed for clinical use. This review process requires manufacturers to submit data showing the accuracy and precision of their product and, in some cases, its analytical sensitivity and specificity. Manufacturers also have to provide documentation supporting the claim that their product is substantially equivalent to one already on the market. This review does not typically consider the clinical safety and effectiveness of the product.

**Prognosis**—an assessment of the probable course of a disease given the risk factors present in an individual; this assessment may affect treatment decisions.

**Prognostic factor**—a measure correlated with a clinical outcome in the setting of natural history or a standard of care regimen; it is a variable used to estimate the risk of or time to clinical outcomes.

**Prospective clinical trial**—a clinical trial in which patients are identified and then followed forward in time.

**Prospective–retrospective clinical study**—an analysis using archived specimens from previously conducted prospective clinical trials that addressed the intended clinical use of the test.

**Proteome**—the complete set of proteins expressed by a cell, tissue, or organism.

**Qualification**—evidentiary process of linking a biomarker with biological processes and clinical endpoints.

**Randomized block trial design**—a test result needs to be available at the time of screening patients for accrual, and the result is used to stratify the randomization of patients to different therapies.

**Risk stratification**—the classification of patients into groups based on the likelihood of developing or suffering effects from a disease.

**Sample set**—in the report, a collection of specimens or other biological materials and the data derived from or associated with them that is used to discover or validate an omics-based test.

**Sensitivity (analytical)**—the lowest concentration that can be distinguished from background noise. This concentration is termed an assay's detection limit.

**Sensitivity (clinical)—**a measure of how often a test correctly identifies patients with a specific diagnosis. It is calculated as the number of true-positive results divided by the number of true-positive plus false-negative results.

**Serum—**the fluid portion of the blood obtained after removal of fibrinogen, other clotting factors, and cells; a clear watery fluid, especially the moistening surface of serous membranes.

**Single nucleotide polymorphism (SNP)—**a variant DNA sequence in which the purine or pyrimidine base (e.g., cytosine) of a single nucleotide has been replaced by another such base (e.g., thymine).

**Specificity (analytical)—**how well an assay detects only a specific substance and does not detect closely related substances.

**Specificity (clinical)—**a measure of how often a test correctly identifies the proportion of persons without a specific diagnosis. It is calculated as the number of true-negative results divided by the number of true-negative plus false-positive results.

**Standard of care—**in medicine, treatment that experts agree is appropriate, accepted, and widely used. Also called best practice and standard therapy.

**Standard operating procedures (SOPs)—**instructions detailing steps and activities of a process or procedure.

**Statistical significance—**a result that is unlikely to have happened by chance. The usual threshold for this judgment is that the results, or more extreme results, would occur by chance with a probability of less than 0.05 if the null hypothesis was true. Statistical tests produce a p-value used to assess this.

**Statistics and bioinformatics validation—**verifying that the omics-based test can perform its intended task. Ideally, this involves ensuring that the test can accurately predict the clinical outcome of interest in an independent set of samples that were not used in developing the test. Such validation is particularly important as omics tests typically involve algorithms or statistical models whose parameters can be "overfit" in any single dataset, leading to an overly optimistic sense of the test's accuracy.

**Systemic therapy—**treatment using substances that travel through the bloodstream and reach and affect cells throughout the body.

**Tamoxifen—**a drug that interferes with activity of the hormone estrogen and is used to reduce risk of breast cancer.

**Test validation phase—**the second phase in the omics-based test development process where the test method is defined, and analytical and clinical/biological validation are performed in a CLIA-certified laboratory. The goal of this process is a fully defined and validated test.

**Training set—**a group of data used to derive a computational model.

**Transcriptome—**the complete set of RNA transcripts from DNA in a cell.

**Trastuzumab (Herceptin)—**a monoclonal antibody that binds to HER2 (human epidermal growth factor receptor 2) and can kill HER2-positive cancer cells. Used to treat breast cancer that is HER2 positive.

**Undifferentiated tumor—**a cancer with cells that do not have specialized structures or functions. Undifferentiated tumor cells often grow and spread quickly.

**Utilization—**contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed.

**Validation—**the process of assessing the assay or measurement performance characteristics.