Deriving Drug Discovery Value from Large-Scale Genetic Bioresources: Proceedings of a Workshop

DETAILS

114 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-44778-2 | DOI: 10.17226/23601

AUTHORS

Siobhan Addie, Amanda Wagner Gee, Steve Olson, and Sarah H. Beachy, Rapporteurs; Roundtable on Genomics and Precision Health; Forum on Drug Discovery, Development, and Translation; Board on Health Sciences Policy; Health and Medicine Division; National Academies of Sciences, Engineering, and Medicine

DERIVING DRUG DISCOVERY VALUE FROM LARGE-SCALE GENETIC BIORESOURCES

Proceedings of a Workshop

Siobhan Addie, Amanda Wagner Gee, Steve Olson,
and Sarah H. Beachy, *Rapporteurs*

Roundtable on Genomics and Precision Health

Forum on Drug Discovery, Development, and Translation

Board on Health Sciences Policy

Health and Medicine Division

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

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

This project was supported by contracts between the National Academy of Sciences and 23andMe (unnumbered contract); AbbVie Inc. (unnumbered contract); American Academy of Nursing (unnumbered contract); American College of Medical Genetics and Genomics (unnumbered contract); American Diabetes Association (unnumbered contract); American Heart Association (unnumbered contract); American Medical Association (unnumbered contract); American Society of Human Genetics (unnumbered contract); American Society of Microbiology (unnumbered contract); Amgen Inc. (Contract No. GHCCOPS-CSARF-85167); Association for Molecular Pathology (unnumbered contract); Association of American Medical Colleges (unnumbered contract); AstraZeneca (unnumbered contract); Baxter BioScience (unnumbered contract); Biogen (unnumbered contract); Blue Cross and Blue Shield Association (unnumbered contract); Burroughs Wellcome Fund (Contract No. 1016598); Centers for Disease Control and Prevention (Contract No. 200-2011-38807, Order No. 0039); College of American Pathologists (unnumbered contract); Critical Path Institute (unnumbered contract); Department of Veterans Affairs (Contract No. VA240-14-C-0037); Doris Duke Charitable Foundation (Contract No. 2015103); Eli Lilly and Company (unnumbered contract); FasterCures (unnumbered contract); Friends of Cancer Research (unnumbered contract); GeneDx (unnumbered contract); GlaxoSmithKline (Contract No. 016084); Health Resources and Services Administration (Contract No. HHSH250200976014I, Order No. HHSH25034017T and HHSH250201500001I, Order No. HHSH25034003T); International Society for Cardiovascular Translational Research (unnumbered contract); Janssen Research & Development, LLC (unnumbered contract); Johnson & Johnson (unnumbered contract); Kaiser Permanente Program Offices Community Benefit II at the East Bay Community Foundation (unnumbered contract); Merck & Co., Inc. (Contract No. CMO-140505-000393, Contract No. CMO-150107-000659, and Contract No. CMO-160208-001245); National Institutes of Health (Contract No. HHSN263201200074I; Task Order Nos. HHSN26300005/0002, HHSN26300066, and HHSN26300093: National Cancer Institute; National Center for Advancing Translational Sciences; National Human Genome Research Institute; National Institute of Allergy and Infectious Diseases; National Institute on Aging; National Institute of Mental Health; National Institute of Neurological Disorders and Stroke; and National Institute of Nursing Research); National Society of Genetic Counselors (unnumbered contract); *New England Journal of Medicine* (unnumbered contract); Northrop Grumman Health IT (unnumbered contract); Pfizer Inc. (unnumbered contract); P*h*RMA (unnumbered contract); Sanofi (unnumbered contract); and Takeda Pharmaceuticals (Contract No. 53108); and U.S. Food and Drug Administration (Contract No. 1R13FD005496-01). Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-44778-2
International Standard Book Number-10: 0-309-44778-X
Digital Object Identifier: 10.17226/23601

Additional copies of this publication are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2016. *Deriving drug discovery value from large-scale genetic bioresources: Proceedings of a Workshop.* Washington, DC: The National Academies Press. doi: 10.17226/23601.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.national-academies.org**.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Reports document the evidence-based consensus of an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and committee deliberations. Reports are peer reviewed and are approved by the National Academies of Sciences, Engineering, and Medicine.

Proceedings chronicle the presentations and discussions at a workshop, symposium, or other convening event. The statements and opinions contained in proceedings are those of the participants and have not been endorsed by other participants, the planning committee, or the National Academies of Sciences, Engineering, and Medicine.

For information about other products and activities of the Academies, please visit nationalacademies.org/whatwedo.

PLANNING COMMITTEE[1]

NADEEM SARWAR (*Chair*), President, Andover Product Creation Innovation System, Eisai Inc.
REBECCA BLANCHARD, Executive Director, Genetics and Pharmacogenomics, Head of Clinical Pharmacogenomics, Merck & Co., Inc.
JOHN CARULLI, Director, Precision Medicine, Biogen
GEOFFREY GINSBURG, Director, Center for Applied Genomics and Precision Medicine, Duke University
DAVID GOLDSTEIN, Director, Institute for Genomic Medicine, Columbia University
WILLIAM HOOS, Consultant, Pancreatic Cancer Action Network
GABRIELA LAVEZZARI, former Assistant Vice President, Science and Regulatory Advocacy, P*h*RMA
LAURA NISENBAUM, Advisor, Chorus Clinical Development, Eli Lilly and Company
DINA PALTOO, Senior Science Policy Analyst, Director of the Genetics, Health, and Society Program, Office of Science Policy, Office of the Director, National Institutes of Health
NALINI RAGHAVACHARI, Division of Geriatrics and Clinical Gerontology, National Institute on Aging

Genomics Roundtable Staff

SARAH H. BEACHY, Project Director
SIOBHAN ADDIE, Associate Program Officer
MEREDITH HACKMANN, Senior Program Assistant

Drug Forum Staff

ANNE B. CLAIBORNE, Project Director
REBECCA A. ENGLISH, Program Officer
AMANDA WAGNER GEE, Program Officer
MORGAN L. BONAME, Associate Program Officer
OLIVIA YOST, Senior Program Assistant

[1]The National Academies of Sciences, Engineering, and Medicine's planning committees are solely responsible for organizing the workshop, identifying topics, and choosing speakers. The responsibility for the published Proceedings of a Workshop rests with the workshop rapporteurs and the institution.

ROUNDTABLE ON GENOMICS AND PRECISION HEALTH[1]

[1]The National Academies of Sciences, Engineering, and Medicine's forums and roundtables do not issue, review, or approve individual documents. The responsibility for the published Proceedings of a Workshop rests with the workshop rapporteurs and the institution.

JILL M. HAGENKORD, Chief Medical Officer, 23andMe, Mountain View, CA

JENNIFER L. HALL, Representative of the International Society for Cardiovascular Translational Research; Associate Professor of Medicine, University of Minnesota, Minneapolis

RICHARD J. HODES, Director, National Institute on Aging, Bethesda, MD

MUIN KHOURY, Director, National Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA

DAVID LANFEAR, Representative of the American Heart Association; Head, Advanced Heart Failure and Cardiac Transplant Section, Henry Ford Hospital, Detroit, MI

GABRIELA LAVEZZARI (*until March 2016*), Assistant Vice President, Scientific Affairs, P*h*RMA, Washington, DC

THOMAS LEHNER, Director, Office of Genomics Research Coordination, National Institute of Mental Health, Bethesda, MD

DEBRA LEONARD, Representative of the College of American Pathologists; Professor and Chair of Pathology at the University of Vermont College of Medicine; Physician Leader of Pathology and Laboratory Medicine at Fletcher Allen Health Care, University of Vermont College of Medicine, University of Vermont, Burlington

ELIZABETH MANSFIELD, Deputy Office Director for Personalized Medicine, Office of In Vitro Diagnostics and Radiological Health, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD

JENNIFER MOSER, Health Science Specialist, Genomic Medicine Program, Office of Research and Development, Department of Veterans Affairs, Washington, DC

LAURA K. NISENBAUM, Advisor, Chorus Clinical Development, Eli Lilly and Company, Indianapolis, IN

ROBERT M. PLENGE (*until February 2016*), Vice President, Merck Research Labs; Head, Genetics and Pharmacogenomics, Merck Research Laboratories, Boston, MA

VICTORIA M. PRATT, Representative of the Association for Molecular Pathology; Associate Professor of Clinical Medical and Molecular Genetics and Director, Pharmacogenomics Diagnostic Laboratory, Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis

MARY V. RELLING, Member and Chair, Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN

NADEEM SARWAR, President, Andover Product Creation Innovation System, Eisai Inc., Cambridge, MA
JOAN A. SCOTT, Chief, Genetic Services Branch, Division of Services for Children with Special Health Needs, Maternal and Child Health Bureau, Rockville, MD
SAM SHEKAR, Chief Medical Officer, Health Information Technology Program, Northrop Grumman Information Systems, McLean, VA
KATHERINE JOHANSEN TABER, Director, Personalized Medicine, American Medical Association, Chicago, IL
DAVID VEENSTRA, Professor, Pharmaceutical Outcomes Research and Policy Program, Department of Pharmacy, University of Washington, Seattle
MICHAEL S. WATSON, Executive Director, American College of Medical Genetics and Genomics, Bethesda, MD
DANIEL WATTENDORF (*until May 2016*), Deputy Chief, Medical Innovations, Department of the Air Force; Program Manager, Defense Advanced Research Projects Agency/Defense Sciences Office, Arlington, VA
CATHERINE A. WICKLUND, Past President, National Society of Genetic Counselors; Director, Graduate Program in Genetic Counseling; Associate Professor, Department of Obstetrics and Gynecology, Northwestern University, Chicago, IL
ROBERT WILDIN, Chief, Genomic Healthcare Branch, National Human Genome Research Institute, Bethesda, MD
JANET K. WILLIAMS, Representative of the American Academy of Nursing; Professor of Nursing, University of Iowa, College of Nursing, Chair of Behavioral and Social Science, Iowa City

Genomics Roundtable Staff

SARAH H. BEACHY, Project Director
SIOBHAN ADDIE, Associate Program Officer
MEREDITH HACKMANN, Senior Program Assistant

FORUM ON DRUG DISCOVERY, DEVELOPMENT, AND TRANSLATION[1]

RUSS ALTMAN (*Co-Chair*), Stanford University, CA
STEVEN GALSON (*Co-Chair*), Amgen Inc., Thousand Oaks, CA
MARGARET ANDERSON, *FasterCures*, Washington, DC
HUGH AUCHINCLOSS, National Institute of Allergy and Infectious Diseases, Bethesda, MD
CHRISTOPHER AUSTIN, National Center for Advancing Translational Sciences, Bethesda, MD
ANN BONHAM, Association of American Medical Colleges, Washington, DC
LINDA BRADY, National Institute of Mental Health, Bethesda, MD
GAIL CASSELL, Harvard Medical School Department of Social and Global Medicine (Visiting), Carmel, IN
ANDREW DAHLEM, Eli Lilly and Company, Indianapolis, IN
JAMES DOROSHOW, National Cancer Institute, Bethesda, MD
JEFFREY DRAZEN, *New England Journal of Medicine*, Boston, MA
JODIE GILLON, AstraZeneca, Summit, NJ
KATHY HUDSON, National Institutes of Health, Bethesda, MD
LYNN HUDSON, Critical Path Institute, Tucson, AZ
S. CLAIBORNE JOHNSTON, Dell Medical School, University of Texas, Austin
RUSTY KELLEY, Burroughs Wellcome Fund, Research Triangle Park, NC
KATE KNOBIL, GlaxoSmithKline, Collegeville, PA
FREDA LEWIS-HALL, Pfizer Inc., New York, NY
BERNARD MUNOS, InnoThink Center for Research in Biomedical Innovation, Indianapolis, IN
JOHN ORLOFF, Baxalta, Deerfield, IL
ROBERT RATNER, American Diabetes Association, Alexandria, VA
MICHAEL ROSENBLATT, Merck & Co., Inc., Kenilworth, NJ
MICHAEL SEVERINO, AbbVie Inc., North Chicago, IL
LANA SKIRBOLL, Sanofi, Washington, DC
BRIAN STROM, Rutgers, the State University of New Jersey, Newark
JOHN WAGNER, Takeda Pharmaceuticals, Cambridge, MA

[1]The National Academies of Sciences, Engineering, and Medicine's forums and roundtables do not issue, review, or approve individual documents. The responsibility for the published Proceedings of a Workshop rests with the workshop rapporteurs and the institution.

JOANNE WALDSTREICHER, Johnson & Johnson, New Brunswick, NJ
CARRIE WOLINETZ, National Institutes of Health, Bethesda, MD
JANET WOODCOCK, U.S. Food and Drug Administration, Silver Spring, MD

HMD Staff

ANNE B. CLAIBORNE, Project Director
REBECCA A. ENGLISH, Program Officer
AMANDA WAGNER GEE, Program Officer
MORGAN L. BONAME, Associate Program Officer
OLIVIA YOST, Senior Program Assistant

Board on Health Sciences Policy Staff

ANDREW M. POPE, Director
HILARY BRAGG, Program Coordinator

Reviewers

This Proceedings of a Workshop has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published Proceedings of a Workshop as sound as possible and to ensure that the Proceedings of a Workshop 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 process. We wish to thank the following individuals for their review of this Proceedings of a Workshop:

LINDA BRADY, National Institute of Mental Health
COLLEEN A. CAMPBELL, University of Iowa
KATHERINE LAMBERTSON, Genetic Alliance
AIDAN C. POWER, Alzheon, Inc.

Although the reviewers listed above have provided many constructive comments and suggestions, they did not see the final draft of the Proceedings of a Workshop before its release. The review of this Proceedings of a Workshop was overseen by **MELVIN WORTH.** He was responsible for making certain that an independent examination of this Proceedings of a Workshop was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this Proceedings of a Workshop rests entirely with the rapporteurs and the institution.

Acknowledgments

The support of the sponsors of the Roundtable on Genomics and Precision Health was crucial to the planning and conduct of the workshop Deriving Drug Discovery Value from Large-Scale Genetic Bioresources and for the development of the workshop proceedings. Federal sponsors are the Centers for Disease Control and Prevention; Department of Veterans Affairs; Health Resources and Services Administration; and National Institutes of Health (National Cancer Institute; National Human Genome Research Institute; National Institute of Mental Health; National Institute of Nursing Research; and National Institute on Aging). Nonfederal sponsorship was provided by 23andMe; American Academy of Nursing; American College of Medical Genetics and Genomics; American Heart Association; American Medical Association; American Society of Human Genetics; Association for Molecular Pathology; Biogen; Blue Cross and Blue Shield Association; College of American Pathologists; Eli Lilly and Company; GeneDx; International Society for Cardiovascular Translational Research; Janssen Research & Development, LLC; Kaiser Permanente Program Offices Community Benefit II at the East Bay Community Foundation; Merck & Co., Inc.; National Society of Genetic Counselors; Northrop Grumman Health IT; and P*h*RMA.

The support of the Forum on Drug Discovery, Development, and Translation was also very important for the workshop. Federal sponsors are the U.S. Food and Drug Administration and National Institutes of Health (National Cancer Institute; National Center for Advancing Translational Sciences; National Institute of Allergy and Infectious Diseases; National Institute of Mental Health; and National Institute of Neurological Disorders and Stroke). Nonfederal sponsorship was provided by AbbVie Inc.; American Diabetes Association; American Society of Mi-

crobiology; Amgen Inc.; Association of American Medical Colleges; AstraZeneca; Baxter BioScience; Burroughs Wellcome Fund; Critical Path Institute; Doris Duke Charitable Foundation; Eli Lilly and Company; *FasterCures*; Friends of Cancer Research; GlaxoSmithKline; Johnson & Johnson; Merck & Co., Inc.; *New England Journal of Medicine*; Pfizer Inc.; Sanofi; and Takeda Pharmaceuticals.

The Roundtable on Genomics and Precision Health and the Forum on Drug Discovery, Development, and Translation wish to express gratitude to the expert speakers who explored how progress could be made in discovering and validating promising targets and medicines for those targets by using the data collected from large-scale genetic studies. The Roundtable and Forum also wish to thank the members of the planning committee for their work in developing an excellent workshop agenda. The project directors would like to thank project staff who worked diligently to develop both the workshop and the resulting proceedings.

Contents

Boxes and Figures

BOXES

FIGURES

Abbreviations and Acronyms

AMP	Accelerating Medicines Partnership
CTTV	Centre for Therapeutic Target Validation
EHR	electronic health record
FNIH	Foundation for the National Institutes of Health
FOP	fibrodysplasia ossificans progressiva
GSK	GlaxoSmithKline
GWAS	genome-wide association study
ITMI	Inova Translational Medicine Institute
LDL	low-density lipoprotein
NCATS	National Center for Advancing Translational Sciences
NIH	National Institutes of Health
PANCAN	Pancreatic Cancer Action Network
PEER	Platform for Engaging Everyone Responsibly
PMI	Precision Medicine Initiative
PPMI	Parkinson's Progression Markers Initiative
SGC	Structural Genomics Consortium
SIK	salt-inducible kinase
SNP	single-nucleotide polymorphism

1

Introduction and Themes of the Workshop[1]

The ability to accurately and rapidly sequence the human genome at a relatively low cost is likely to bring about widespread changes in the way medicine is practiced, including the introduction of novel genomics-based therapies for patients (Pasche and Absher, 2011). However, the process of discovering and developing a new drug or therapy is extremely costly and time consuming. Recently, it has been estimated that the creation of a new medicine costs on average more than $2 billion and takes 10 years to reach patients (DiMasi et al., 2016). The challenges associated with bringing new medicines to market have led many pharmaceutical companies to seek out innovative methods for streamlining their drug discovery research.

One way to increase the odds of success for compounds in the drug development pipeline is to adopt genetically guided strategies for drug discovery. A recent analysis of approved medicines indicated that using human genetic data to support the selection of drug targets and indications can roughly double the chances that a given drug will be clinically successful compared with drugs lacking genetic support (Nelson et al., 2015). The use of genetic information for drug discovery and development has led to a variety of innovations and new approaches in both the private and public sectors.

Recognizing the potential drug discovery benefits of collecting genetic and phenotypic information across specific populations, pharma-

[1]The planning committee's role was limited to planning the workshop. The Proceedings of a Workshop has been prepared by the rapporteurs as a factual account of what occurred at the workshop. Statements, recommendations, and opinions expressed are those of individual presenters and participants and have not been endorsed or verified by the National Academies of Sciences, Engineering, and Medicine. They should not be construed as reflecting any group consensus.

ceutical companies have started collaborating with healthcare systems and private companies that have curated genetic bioresources, or large databases of genomic information. Large-scale cohort studies offer an effective way to collect and store information that can be used to assess gene–environment interactions, identify new potential drug targets, understand the role of certain genetic variants in the drug response, and further elucidate the underlying mechanisms of disease onset and progression. The goal of many of these institutional partnerships is to identify and validate potential therapeutic targets, develop biomarker assays, and produce new and better targeted medicines. One example of this type of arrangement is the partnership between Regeneron Pharmaceuticals and the Geisinger Health System. Patient samples collected by Geisinger undergo DNA sequencing, providing researchers with information they can use to better understand the relationship between genes and human diseases.

At the same time, government agencies and academic researchers have new opportunities to work together to create new genetic bioresources and more sophisticated analytical tools. In these workshop proceedings, genetic bioresources are referred to as studies or initiatives in which biological specimens are collected from individuals with the intent to sequence DNA for research and/or discovery purposes. For example, the Precision Medicine Initiative (PMI), a program announced by President Obama in 2015, will assemble a cohort of 1 million U.S. volunteers for longitudinal research, including genetic studies (Collins and Varmus, 2015). Programs such as the PMI have the potential to spur new and innovative business models for integrating scientific and technological expertise across sectors and allowing for greater access to genomic data of clinical research participants. However, many questions remain about the design of large cohort studies, the types of data that should be collected, and which business models could engage stakeholders most effectively. To examine how genetic bioresources could be used to improve drug discovery and target validation, the Roundtable on Genomics and Precision Health (previously called the Roundtable on Translating Genomic-Based Research for Health) worked with the Forum on Drug Discovery, Development, and Translation to host a workshop on March 22, 2016, in Washington, DC, titled Deriving Drug Discovery Value from Large-Scale Genetic Bioresources.[2]

[2]The workshop agenda, speaker biographical sketches, statement of task, and registered attendees can be found in Appendixes A, B, C, and D, respectively.

BOX 1-1
Objectives of the Workshop

- To address how progress can be made in discovering and validating promising targets and medicines for those targets by using data collected from large-scale genetic studies.
- To highlight current genomics-enabled drug discovery activities in industry, academia, and government and to share best practices for study design and data collection.
- To examine enabling partnerships and business models that can facilitate the use of genetic data for drug discovery.

Participants at the workshop explored the current landscape of genomics-enabled drug discovery activities in industry, academia, and government; examined enabling partnerships and business models; and considered gaps and best practices for collecting population data for the purpose of improving the drug discovery process. (Box 1-1 lists the objectives of the workshop.) A wide variety of stakeholders presented their perspectives and participated in workshop discussions, including representatives of pharmaceutical companies, academic research institutes, health care systems, direct-to-consumer genetics companies, and patient advocacy groups.

Over the course of the workshop, several participants made suggestions for steps that could be taken within the public and private sectors to use large-scale genetic bioresources for the purpose of accelerating drug discovery. These suggestions are synthesized in Box 5-1 in the final chapter of these proceedings.

OVERVIEW OF THE WORKSHOP

Recently, the Forum on Drug Discovery, Development, and Translation has been mapping out the landscape of drug discovery and development and developing a tool to identify points in the ecosystem where bottlenecks and challenges arise, said Russ Altman, co-chair of the Forum on Drug Discovery, Development, and Translation and the Kenneth Fong Professor and a professor of bioengineering, genetics, medicine, and (by courtesy) computer science at Stanford University. Several of the activities of the Roundtable on Genomics and Precision Health in the past 2 years have been focused on clinical genomics and translation, so

the opportunity to jointly present this workshop comes at an excellent time, said Geoffrey Ginsburg, co-chair of the Roundtable on Genomics and Precision Health and the director of the Center for Applied Genomics and Precision Medicine at Duke University Medical Center. New large-scale genetic bioresources, activities, tools, and institutional arrangements are creating tipping points in science, culture, and business, said Nadeem Sarwar, president of Andover Product Creation Innovation Systems at Eisai Inc., and chair of the workshop. These tipping points are providing unprecedented opportunities to accelerate the delivery of innovative, targeted new medicines, Sarwar said.

Workshop speakers focused on new research and ideas in three primary areas: large cohort studies, genome-enabled discovery activities, and novel business models that support the development and use of data from bioresources. Throughout the workshop there was robust discussion of short- and long-term options for collaboration, translational research, and accelerated progress in the area of genomics-enabled drug discovery.

Designing Large Cohorts to Maximize Discovery Capabilities

Genetic and phenotypic data gathered from large cohorts make it possible to detect common genetic variants that increase or decrease the risk of disease, rare variants that can be used to identify drug targets, and genetic subpopulations in which particular treatments may be most effective, said Kári Stefánsson, the chairman and chief executive officer of deCODE genetics.

Several large studies that involve bioresources are already under way, including the Electronic Medical Records and Genomics (eMERGE) Network in the United States, and the Oxford Biobank in the United Kingdom, among others. Cohort studies are especially valuable when they include data collected over long periods of time, have extensive population participation, and allow for genotype- or phenotype-based recall studies, according to Mark Daly, the co-director of the Program in Medical and Population Genetics at the Broad Institute of the Massachusetts Institute of Technology and Harvard University. In addition, integrated approaches that include general population studies, family-based studies, research on founder populations, and studies of phenotypic-specific cohorts could potentially maximize opportunities for drug discovery, said Aris Baras, the vice president and co-head of the Regeneron Genetics Center. However, important issues can arise during cohort design, including

challenges with data sharing, informed consent, privacy, and analytical and computational issues, all of which are discussed in Chapter 2.

Genomics-Enabled Discovery Activities Related to Bioresources

Large genetic bioresources can facilitate the use of sophisticated analytical tools, new approaches to target validation and biomarker development, and studies of allelic variation in drug response (The Michael J. Fox Foundation, 2015; Whirl-Carrillo et al., 2012). Individual workshop speakers discussed discovery activities, many of which are in the precompetitive realm, that are enabling a greater understanding of human disease biology and of drug–target interactions across a wide variety of diseases (see Chapter 3). One example of this type of discovery activity is the Parkinson's Progression Markers Initiative (PPMI) of The Michael J. Fox Foundation. The PPMI is collecting and sharing data from subjects with Parkinson's disease in order to carry out research on—and develop biomarkers of—the mechanism and progression of Parkinson's disease.

Although genome-wide association studies (GWASs) have identified several thousand genetic variants that confer genetic risk for common diseases, there are several reasons that these results have not yet been translated into the predicted plethora of new medicines, including possible population stratification in GWASs, the lack of functional links between risk variants that lie in non-coding regions of the genome and the disorders they are connected to, and high rates of attrition for drug candidates that enter phase 2, "proof-of-concept" studies (Bunnage, 2011; McClellan and King, 2010). Target validation is an important step in the drug discovery process because it can help bridge the gap between basic science research and the development of new medicines (Smith, 2003). As a way to collaborate on target validation and accelerate translation to new medicines, GlaxoSmithKline (GSK), the European Bioinformatics Institute, and the Wellcome Trust Sanger Institute established the Centre for Therapeutic Target Validation (CTTV). The CTTV brings together expertise in drug discovery, chemistry, functional genomics, and electronic health records to study potential drug targets in the precompetitive realm, said Lon Cardon, the senior vice president of alternative discovery and development and the head of target sciences at GSK.

In addition to target validation, genomics-enabled drug discovery

requires a thorough understanding of the molecular, cellular, and organismal consequences of diseases and their treatments, said workshop speaker Tim Rolph, the vice president of program value enhancement at Pfizer Inc. Drug discovery programs will be successful if they can modulate and assay the activity of drug targets within cell and model systems and combine that with a better understanding of the biology of the potential target, according to Sally John, the vice president of computational biology and genomics at Biogen.

Business Models That Support Drug Discovery Across Stakeholder Groups

Progress of drug discovery and development efforts that begin within a single stakeholder organization can be greatly accelerated when such efforts are taken advantage of by larger collaborations involving multiple stakeholder groups. For example, patient and disease advocacy groups can gather genotypic and phenotypic data for investigators at academic institutions or pharmaceutical companies, help recruit individuals into clinical trials, provide treatment recommendations, and educate patient populations (see Chapter 4).

Multi-stakeholder collaborations supporting target identification, clinical development, patient stratification, and pharmacogenetics can similarly accelerate drug discovery and development (Hodes and Buckholtz, 2016). One such activity is the Accelerating Medicines Partnership (AMP), an effort that combines resources and expertise from different sectors such as government, industry, and the nonprofit sector, while increasing the scope, coordination, and efficiency of ongoing and future studies.

Collaborative partnerships also can result in the development of drug discovery tools and resources that can further increase participation in and the effectiveness of the collaboration. For example, knowledge portals such as AMP can allow integrated interrogation across multiple datasets while maintaining individual-level data privacy, said David Wholley, the director of research partnerships for the Foundation for the National Institutes of Health (FNIH). A major consideration in such collaborations is making sure that the initiatives are sustainable from scientific, financial, and administrative perspectives, so that promising opportunities can be realized.

Another example of a collaborative model discussed by workshop participants was the risk-sharing approach provided by the Structural

Genomics Consortium (SGC). The SGC brings together government agencies, academic scientists, philanthropic organizations, pharmaceutical companies, and others to provide resources for discovering and validating drug targets. As an open access research model, the SGC focuses on advancing the science, and therefore projects are less affected by commercial, institutional, or other financial interests.

Looking Toward the Future

The workshop participants discussed short-term and long-term options that could support genomics-based drug discovery. Individual speakers supported achieving a broader understanding of the underlying biology of human disease as a possible way to accelerate the drug discovery process. For example, there is much more to be learned about the role that genetic variants play in disease onset and progression. New technologies and computational tools, such as the development of "tissues on a chip" to understand biological mechanisms and test hypotheses, may constitute a disruptive change which hastens progress.

Another issue emphasized by some workshop participants was the importance of incentives for sharing data and resources among investigators in both the private and public sectors. Greater collaboration across disciplines could increase the sharing of data and expertise and the subsequent understanding of disease mechanisms. Collaborative efforts could reduce duplication, enhance progress, and break down siloes that create barriers within and among sectors. Other workshop speakers noted that many collaborative partnerships in genomics have already been formed, and that the sharing of research data is commonplace. However, it was also pointed out by individual speakers that the use of data from cohort studies around the world may be limited by restrictions on who can access the data and how that access is provided. One way to expand data sharing from these cohorts is to share summary-level data, said Stefánsson. Summary data are available for download through AMP, said David Wholley, but the ability to interrogate individual level data and develop reliable tools is a key aspect of AMP.

ORGANIZATION OF THE PROCEEDINGS

Following this introductory chapter, Chapter 2 reviews the features of and challenges associated with large cohorts set up for genetics re-

search and drug discovery. Examples of the useful features of large cohort studies suggested by some participants include consistent data collection methods; a large, stable, and engaged pool of participants; and the ability to perform genotype- or phenotype-based recall studies. Potential challenges associated with drug discovery using data collected from cohort studies include incomplete annotation of intergenic sequences and a lack of knowledge about the effects of rare variants. These benefits and challenges are discussed within the context of specific genomic cohort efforts in Finland, Iceland, an integrated translational medicine institute, a pharmaceutical company, and a direct-to-consumer genetics company. Perspectives on data quality, data sharing, and privacy are also shared in this chapter.

Chapter 3 focuses on drug discovery activities that can maximize the usefulness of genetic bioresources. One such activity is an informatics knowledge portal that curates and details information about the relationship between drugs and genes. This tool allows researchers to test their hypotheses in silico, saving precious time and funds. Another activity is public–private collaboration, where the goal is to bridge the gap from basic research discoveries to new medicines by expanding the precompetitive space and accelerating target validation. One program developed by a patient advocacy organization features open sharing of all data and is addressing the critical lack of biomarkers for Parkinson's disease. Understanding the biology of potential drug targets and pathways is an important step in maximizing findings from genetic cohort studies.

Chapter 4 features a discussion of multiple business models across government, academia, patient advocacy groups, and industry that are designed to hasten genomics-enabled drug discovery. The goals, scope, design, and available results from each model are presented for consideration and comparison.

The proceedings culminate in Chapter 5, which describes potential actionable ideas proposed by individual speakers for ways to improve genomics-based drug discovery. The chapter begins with an example of a new technology, an in vitro "tissue chip" platform for growing human tissues that has the potential to cause disruptive change in drug discovery and development. Individual speakers proposed the idea that cross-sector collaboration and increased data sharing could create a new ecosystem that could help to fulfill the potential for drug discovery that genomic resources offer.

2

Maximizing Discovery Capabilities Through Cohort Design

Important Points Highlighted by Individual Speakers

- Acceleration of drug discovery requires better target validation, indication discovery, and biomarker development. Integrated approaches to genomics-driven drug discovery, including general population studies, family-based studies, research on founder populations, and studies of phenotype-specific cohorts, can maximize the opportunities for drug discovery. (Baras)
- Large-scale genetic cohort studies point to the value of widespread national and international collaborations because no single cohort or country can generate all the needed insights on a specific disease or condition. (Daly)
- Useful features of cohort studies include consistent data collection in large populations over many years, extensive population participation, and the capability to do recall studies. (Daly)
- Large cohorts associated with genetic data are rich sources of information that make it possible to look for common genetic variants that increase or decrease the risk of disease, rare variants within genes that are valuable as drug targets, and genetic subpopulations in which particular treatments may be useful. (Stefánsson)
- Due to the large volume of patients seen at community health centers in the United States, these centers are an ideal venue for collecting and curating genetic data linked to longitudinal health information in electronic health records. (Vockley)

- Direct-to-consumer genetic testing companies can produce valuable genotypic and phenotypic data for large numbers of people. Both genome-wide and phenome-wide association studies are possible with these data, even with self-reported phenotypic information. (Scheller)
- To achieve maximum levels of participation, researchers need to assure patients that their data will not be misused or mishandled, even if data are widely shared among investigators. (Stefánsson)

Cohort studies, designed to search for links between risk factors and health outcomes, are increasingly collecting genomic data. Thus, it is important to consider designing those studies in ways that maximize the usefulness of the data. Speakers in the first session were asked to explore current gaps and opportunities regarding data collected in large cohort studies and what elements could enable a robust drug discovery toolbox. Speakers noted that genetics studies at the national level in countries such as Iceland and Finland can reveal valuable lessons about the potential of national registry data for research, along with constraints associated with gathering and analyzing large quantities of genotypic and phenotypic data. Many cohort studies are designed as longitudinal studies that collect data from the same group of participants over an extended period of time. Another approach to cohort studies is used by companies that offer direct-to-consumer genetic testing, meaning genetic tests that are marketed directly to the public. Each of these types of cohort studies offers important lessons in cohort study design and implementation. Issues that arise when designing large cohort studies include challenges concerning sharing data, protecting individual privacy, and overcoming analytical and computational barriers.

BIOBANK STUDIES IN FINLAND

Several large cohort studies in northern European countries have demonstrated the potential of genetic bioresources to enhance drug discovery and development. In Finland, a combination of linguistic and geographic isolation has resulted in a population that can trace much of its DNA back to bottleneck events in the 16th century, during which time internal migrations gave rise to isolated regional subpopulations. Over the years, these small subpopulations expanded, but without the genetic influence of outsiders. As Mark Daly, the chief of the Analytic and

Translational Genetics Unit at Massachusetts General Hospital and the co-director of the Program in Medical and Population Genetics at the Broad Institute, pointed out, members of the Finnish population have approximately one-third as many rare variants in their genomes as do the members of most other European populations. In addition, some rare variants in the Finnish population have increased to unusually high frequencies. For example, this population is enriched in loss-of-function variants that lower lipoprotein(a) levels and are strongly cardioprotective, providing, as Daly said, "precise clues to interventions that could be safe and effective."

Health data are most useful when they have been collected consistently in a large population over many years so as to track outcomes over long periods of time, Daly said. Finland, like other Nordic countries, has an enormous amount of national registry data, all coordinated with individual identification. This makes it possible to track all hospitalizations, outpatient visits to specialty clinics, and prescription medications that an individual has taken, Daly said. For example, the THL Biobank in Finland[1] now has samples from about 130,000 individuals from population cohorts and 70,000 individuals from disease collections (see Figure 2-1). The data in this biobank enable studies on a wide range of human diseases. "If you are interested in a specific phenotype, or a response to a specific medication, that information can not only be fleshed out over the life history of two individuals but over hundreds of thousands," Daly said. This enables new study designs that are difficult to do in a traditional academic environment, such as searching for adverse outcomes using prescription registry data, including adverse outcomes in people with specific genetic backgrounds, he said.

Another key element to effective cohort studies is extensive population participation and genotype-based recall, Daly said. Once a gene has been determined to play a role in a disease, understanding that gene's function can require additional studies in genotype-based or phenotype-based populations. Finland's Biobank Act of 2013[2] has made genotype-based recall very easy, Daly said, because it enables recall access to all individuals in the biobank, even those whose samples were collected years earlier. For example, Daly and his colleagues have identified loss-of-function

[1]For more information on Finland's THL Biobank, visit https://www.thl.fi/fi/web/thlfi-en/topics/information-packages/thl-biobank (accessed May 23, 2016).

[2]The Biobank Act of Finland can be found at http://www.findex.fi/fi/laki/.kaannokset/laki/kaannokset/2012/en20120688.pdf (accessed May 23, 2016).

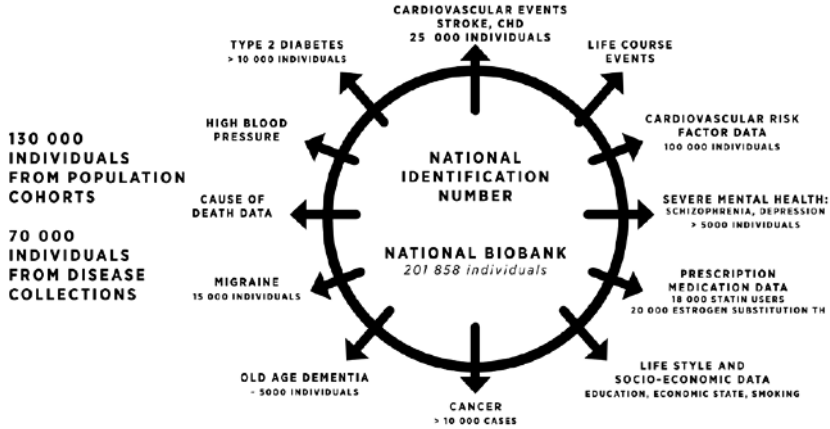

FIGURE 2-1 The THL Biobank in Finland includes more than 200,000 people, making possible large cohort studies in a wide range of disease areas.
NOTE: CHD = coronary heart disease.
SOURCE: Mark Daly, National Academies of Sciences, Engineering, and Medicine workshop presentation, March 22, 2016.

mutations in the *SLC30A8* gene as being strongly protective against type II diabetes (Flannick et al., 2014). These mutations in *SLC30A8* are extremely rare, but their existence in Finland and Iceland permitted focused and highly targeted genotype-based recall studies.

Even though the studies required subjects to have a meal and then undergo numerous blood draws over 3 hours, more than 75 percent of the eligible individuals in the biobanks made themselves available for the recall study. The Finnish population tends to participate in recall studies because they have high levels of education, access to their own medical data, and a great deal of trust in their doctors and researchers, Daly said.

Biobank studies should not be viewed in isolation or as competitors with other studies, Daly said. On the contrary, genetic cohort studies point to the value of widespread national and international collaborations because no one cohort or country can generate all the needed insights on a disease. Also, patient communities focused on specific diseases can often provide the sequenced genomes or exomes of more individuals than will be diagnosed with those diseases in many biobank studies. Most primary disease discoveries will continue to emerge from collaborations among investigators, and biobanks will then allow deeper exploration of these findings. “Insights into human genetics generally have

come and will continue to come from the extremes of the population—individuals with severe disease. We need to embrace data from the extremes of the population, and we need to then use biobanks . . . to allow much deeper exploration of those findings," Daly said.

GENETIC BIORESOURCES AND DISCOVERY EFFORTS IN ICELAND

As in Finland, a great deal is known about the genetic diversity of the population in Iceland. deCODE genetics,[3] a company based in Iceland, has been working to collect and curate genotypic and medical data on the Icelandic population, reported Kári Stefánsson, the chief executive officer and founder of deCODE genetics. Approximately half of the Icelandic population, 160,000 individuals, volunteered to have their DNA genotyped using an Illumina chip, he said. The whole genome has been sequenced in 20,000 Icelanders to a median depth of 30× coverage, and imputed variants are known down to a frequency of 0.01 percent in a cohort composed of 390,000 living or deceased Icelanders, Stefánsson said. A list is available of 16,000 Icelanders who are homozygous for a loss of function mutation in at least 1 of 1,800 genes, he added.

In addition, much is known about the phenotypes of the Icelandic cohort participants, including information on diseases, surgical procedures, prescriptions, mental disorders, and weight and height, along with educational attainment, socioeconomic status, and even memberships in associations—all of which can be used in studies in combination with data on genetic variants.

Researchers have been using the data that are available in Iceland in a few different ways. First, they are looking at common genetic variants that can either increase or decrease the risk of specific diseases. Stefánsson and his colleagues at deCODE have associated a large number of complex illnesses, including diabetes, osteoporosis, dementia, and cancer, with common variants in sequences (Stacey et al., 2015; Steinberg et al., 2015; Steinthorsdottir et al., 2014; Styrkarsdottir et al., 2016). These common variants usually have a subtle effect on the risk of disease, but an individual can carry a collection of them whose confluence creates a relatively high genetic risk.

Researchers at deCODE have also been looking at rare variants to

[3]For more information on deCODE genetics, see http://www.decode.com/research (accessed July 8, 2016).

identify genes that can be used as drug targets in a wide variety of conditions and diseases. This research can readily link diversity in sequence to diversity in phenotype, Stefánsson said. In addition, the genetic data from the Icelandic population have facilitated research on the potential side effects of therapeutic manipulation of the protein encoded by a particular genetic target. For example, antibodies that block the activity of PCSK9 cause an increase in the recycling of the low-density lipoprotein receptor and thus a reduction of low-density lipoprotein cholesterol levels in the blood (Blom et al., 2016). However, questions have arisen about whether reducing the activity of PCSK9 through antibody treatment might increase the risk of dementia or diabetes. By examining a loss of function PCSK9 variant (p.R46L), which is present in the Icelandic population, researchers found no evidence of such risks and a slightly longer lifespan, Stefánsson said. Additional work using the Icelandic biobank samples has uncovered another variant that has a greater protective effect for coronary artery disease, he added (Nioi et al., 2016).

Using Biobank Data for Alzheimer's Disease Research

Another example of how genetic biobank data can aid in the drug discovery process involves research on Alzheimer's disease, Stefánsson said. In August 2012, deCODE researchers and their colleagues announced the discovery of a coding mutation (p.A673T) in the amyloid protein precursor (*APP*) gene that confers protection against Alzheimer's disease and cognitive decline (Jonsson et al., 2012). Furthermore, when the *APP* mutation was examined in a cohort of nursing home residents in Iceland who had not been diagnosed with Alzheimer's disease, the variant was associated with a slowing of normal cognitive decline. Using a similar process, the researchers discovered a variant in another gene, *TREM2*, that results in a predisposition to Alzheimer's disease. TREM2 could be a particularly useful target in Alzheimer's disease (Jonsson et al., 2013).

Data from genetic bioresources also can help answer the question of when to begin treating a chronic disease such as Alzheimer's, Stefánsson said. One approach is to group individuals based on carrier status in a particular gene, such as apolipoprotein E-4 (*APOE-4*) and look for signs of early cognitive decline as an indicator for when to begin treating. The use of genetic markers for patient stratification in clinical trials could increase the likelihood of successful drug development, Stefánsson said.

IMPROVING GENOMICS AND COHORT STUDIES

As part of their presentations and during the panel discussion following Session I, Baras and Stefánsson identified several potential improvements that could strengthen genomics as a research tool for drug discovery:

- Better annotation of intergenic sequences could provide a greater understanding of the mechanisms of disease caused by common variants. (Stefánsson)
- Improvements in technology could allow for longer reads on whole genomes. Curating longer-read data from large populations would help detect rare variants and replicate novel gene discoveries. (Baras, Stefánsson)
- Understand the basic underlying biology of human diseases. (Stefánsson)
- Better capabilities for biological and clinical target validation. (Baras)
- Novel computational methods for handling challenges with large datasets and rare variants. (Baras)

LONGITUDINAL STUDIES AT THE COMMUNITY HEALTH LEVEL

Longitudinal health data from patient populations, in combination with genetic data, are another source of valuable information. For example, the Inova Translational Medicine Institute (ITMI), in Falls Church, Virginia, has generated roughly 8,000 whole-genome sequences with matched expression, methylation, microRNA, clinical, and survey data from families that come to the Inova health center for care, said Joe Vockley, the chief operating officer, chief scientific officer, and senior vice president at the institute. Many of ITMI's studies focus on issues in neonatal and pediatric care, and anyone who delivers a child within the Inova health system has the opportunity to enroll in a longitudinal study, he said.

ITMI's study designs integrate physicians, genomics researchers, and bioinformatics scientists, and ITMI has standardized collection practices to maximize the usefulness of the data. The data gathered include clinical, family history, and survey information (including nutritional, behav-

ioral, and environmental data). Every sample at ITMI that is collected for the biobank is done so using standard operating procedures to ensure quality and consistency. For example, warm and cold ischemic time is minimized, meaning that biospecimens are cooled quickly after collection for long-term storage. Once samples are collected and cooled, they undergo barcoding and are stored at a constant temperature. The institute also has built a robust infrastructure for data analysis, using both an on-site supercomputer and cloud-based storage, with computing and other resources provided to researchers for accurate and integrated analysis of datasets.

In the past, a challenge for longitudinal studies has been the failure to adhere to high standards for data generation, Vockley said. For example, to save time, laboratories often generate data from collected samples using commercially available kits, but companies frequently update or modify their kits, and researchers do not always properly document or disclose these modifications. These modifications to research tools can significantly change the data, making them difficult to integrate and interpret, Vockley said. Another related challenge is the high number of potential biomarkers of human disease discovered in longitudinal studies and cited in peer-reviewed literature; Vockley noted only a small proportion of those are validated biomarkers. Therefore, researchers need to be aware that findings related to biomarkers in published studies can sometimes be incorrect, he said.

The community health care system in the United States is an excellent place to launch longitudinal cohorts, Vockley said, because it sees a great deal of patients. In 2014, approximately 23 million people were seen at community health centers supported by the Health Resources and Services Administration (HRSA, 2014). Tissues from large numbers of patients can be easily collected, biobanked, and made available for drug discovery research, he observed.

GENETICS-GUIDED DRUG DEVELOPMENT AT REGENERON

Studying human genetic data collected from longitudinal studies is one approach that Regeneron Pharmaceuticals is taking to accelerate the discovery and development of new drugs, said Aris Baras, the vice president and co-head of the Regeneron Genetics Center. Regeneron is incorporating human genetics into its research and development paradigm using a three-pronged approach:

- Target discovery: Identify new drug targets and pathways
- Indication discovery: Identify new indications for drug targets and programs
- Biomarkers: Use pharmacogenetics to predict an individual's response to drugs

A major bottleneck that is slowing the development of new drugs is identifying well-validated drug targets, Baras said. The pharmaceutical industry as a whole is looking at approximately 500 to 1,000 genes (accounting for less than 5 percent of all genes) as targets for medicines, he said, and Regeneron is currently pursuing 200 different targets, with support from a variety of sectors, including human genetics, mouse genetics, molecular biology, biochemistry, and clinical pharmacology. The company also uses principles from human genetics to identify links that can help expand drug indications, and it uses pharmacogenetics to identify patients who may be more responsive or who may have a predictable adverse effect to a particular medicine.

Regeneron's Collaboration with Geisinger

To support its research goals, Regeneron has taken an integrated approach to drug discovery, relying on general population studies, family-based studies, research on founder populations, and studies of phenotype-specific cohorts. One example of this approach is the partnership that Regeneron has developed with the Geisinger Health System in Pennsylvania. The goal of the partnership with Geisinger is to link genetic mutations with real-world outcomes in order to make actionable discoveries. Geisinger is a community-based health system that serves more than 2.5 million patients and that has used the same electronic health record (EHR) system for more than 20 years. Therefore, it has the world's largest longitudinal population in which iterative call-back phenotyping and sample collection have been operationalized, Baras said.

The DiscovEHR Study, a collaboration between Regeneron and Geisinger, is a large population-based project whose goal is to sequence the DNA of more than 250,000 participants. To date, Regeneron and Geisinger have developed a large number of high-quality, EHR-derived phenotypes for genetic analyses, Baras said, as well as a growing library of more than 2,000 quantitative and qualitative traits that are available

for high-throughput and in-depth analyses. Examples of traits for which in-depth datasets are available include coronary artery disease and lipid metabolism, chronic obstructive pulmonary disease and asthma, and bariatric traits and liver histology. Within the subjects sequenced to date, carriers of heterozygous loss-of-function variants have been found in more than 92 percent of genes, and the study has already found more than 1,300 genes with homozygous loss-of-function indications, including many Regeneron drug targets, Baras said. Increasing the number of sequenced individuals and expanding to studies in founder populations will boost the ability to understand the consequences of these mutations across many diseases, he added.

Using Genetics to Guide Drug Discovery

Through general population cohort studies, including the DiscovEHR Study, it was found that loss-of-function mutations in the *ANGPTL3* and *ANGPTL4* genes result in lower levels of triglycerides and a lower risk of coronary artery disease than is found in noncarriers (Dewey et al., 2016; Helgadottir et al., 2016). Both ANGPTL3 and ANGPTL4 are endogenous inhibitors of lipoprotein lipase, which mediates triglyceride metabolism and triglyceride-rich lipoproteins (Ono et al., 2003; Sukonina et al., 2006). Based on these findings, Regeneron developed an antibody to ANGPTL3 (evinacumab) which is now in late-stage development for dyslipidemias. The *ANGPTL3* and *ANGPTL4* variants also appear to have a protective effect for type II diabetes, which Regeneron researchers are now pursuing.

Researchers at Regeneron who are interested in pulmonary arterial hypertension, a cardiometabolic disorder, rely on family-based, Mendelian approaches that involve genetic data gathered from large pedigrees. Pulmonary arterial hypertension is characterized by high blood pressure in the pulmonary artery, which carries blood from the heart to the lungs; to date only 10 genes are known to be associated with pulmonary arterial hypertension (NLM, 2016). In looking at about 70 families and 190 singletons with familial pulmonary arterial hypertension enriched for pediatric onset, researchers from Columbia University Medical Center discovered rare deleterious variants in the *TBX4* gene, Baras said. The hope is that Regeneron researchers can take genetic insights like those and convert them to drug discovery opportunities and therapeutics for patients, he said.

As a final example, Baras shared information about Regeneron's efforts to study fibrodysplasia ossificans progressiva (FOP), a rare disorder that results in heterotopic ossification leading to severe disability and

early mortality. Most FOP patients have a mutation in a bone morphogenic protein type I receptor gene, $ACVR1^{R206H}$, which promotes heterotopic ossification. While examining all of the ligands for this receptor, researchers at Regeneron discovered that Activin A induces ossification in mice that carry the $ACVR1^{R206H}$ mutation. Based on this finding, a neutralizing antibody to Activin A is now in clinical development.

In summarizing Regeneron's work, Baras laid out several guiding principles for genetics-guided drug discovery:

- Leverage multiple approaches to maximize opportunities for gene discovery, including research in large general populations, Mendelian genetics and families, founder populations, and disease cohorts with deep phenotyping.
- Gather deep phenotypic data, including longitudinal data and quantitative traits.
- Emphasize large-effect coding variations.
- Take advantage of multi-ethnic cohorts.
- Use analytical strategies to interrogate wide-ranging biological questions.

Human genetics has proven itself to be a valuable way of optimizing and improving drug development, Baras concluded. "You cannot pursue modern drug discovery and development without incorporating human genetics into your R&D approaches," he said.

DRUG DISCOVERY THROUGH DIRECT-TO-CONSUMER GENETICS

Another source of genomic and phenotypic data is personal genetics companies that offer direct-to-consumer genetic tests. For example, when consumers send a sample of their saliva to the genetic testing company 23andMe, they have the option of giving the company permission to use their data for research purposes. More than 80 percent of people do agree to have their data used for these purposes, said Richard Scheller, the chief scientific officer and head of therapeutics at 23andMe. "We are receiving over 2 million data points per week, and people have answered over 345 million survey questions," he said.

23andMe currently genotypes about 600,000 SNPs (single-nucleotide polymorphisms) per sample and from that they are able to use Minimac, an

imputation software program, to expand the data to 15 million SNPs. The company has genotyped more than 1.2 million people, 77 percent of whom have been Euro-American, 10 percent Latino, 5 percent African-American, 4 percent East Asian, 2 percent South Asian, and 2 percent other. About 52 percent of customers are female, with a fairly even distribution of ages. The company uses a series of question to gather a wide variety of phenotypic data (see Box 2-1).

The company then provides a report to its customers with 35 carrier status reports. (This number is down from more than 200, Scheller said, as the company responded to concerns expressed by the Food and Drug Administration about the accuracy and comprehension of the health information being provided to consumers.[4]) Other reports cover ancestry, wellness, and other traits.

BOX 2-1
Examples of Survey Questions Posed by 23andMe to Participants (as presented by Richard Scheller)

- Diagnoses: "Have you ever been diagnosed with neuroblastoma?" (Other diagnoses include psoriasis, inflammatory bowel disease, high blood pressure, and high cholesterol.)
- Medication usage: "Which statin medications have you taken?"
- Response to medication: "Does Claritin (loratadine) cause you to have a fast or irregular heartbeat?"
- Family history of disease: "Have any of your grandparents, parents, brothers, sisters, aunts, or uncles ever been diagnosed with mild cognitive impairment?"
- Health behaviors: "Do you currently smoke cigarettes?"
- Geographic location: City, state, country
- Personality traits: "I am someone who is sometimes shy, inhibited." [Scale of 1–5]
- Environmental exposures: "Have you ever had a job in which you worked in petroleum products, such as benzene, toluene, or gasoline?"

[4]The warning letter from the U.S. Food and Drug Administration to 23andMe, dated November 22, 2013, is available at http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm (accessed October 14, 2016), and the close-out letter, dated March 25, 2014, is available at http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/ucm391016.htm (accessed October 14, 2016).

23andMe is trying to encourage people to become very engaged in their own genetics, Scheller said, and the goal is for non-experts to learn about their traits and how their genetics affects their health.

Self-Reported Health Information as a Discovery Tool

23andMe also uses data from consenting customers to do genome-wide and phenome-wide association studies, with the former focused on diseases of interest and the latter focused on SNPs or genes of interest. Because customers report their own health information through a series of questions, Scheller said, 23andMe wanted to carefully examine and validate the data and ensure that this approach was scientifically sound. One way that the company did this was by looking at a variant in the *CYP1A2* gene, whose product is responsible for more than 95 percent of the primary metabolism of caffeine (Kalow and Tang, 1993). When 23andMe researchers examined the participants who indicated that drinking a caffeinated beverage in the evening makes sleeping difficult, they found an enrichment of *CYP1A2* variants, indicating that in that instance the self-reported phenotype information corresponded very well to genetic data. They went on to find other genetic variants that were protective against getting jittery from drinking caffeine. "By asking [our participants] simple questions, we are able to see pathways that control how caffeine is metabolized," Scheller said.

Recently, researchers at 23andMe published their analysis of genetic associations of self-reported "morningness," or the preference to rise and to rest early (Hu et al., 2016). In analyzing data from a cohort of nearly 90,000 23andMe participants, the authors identified 15 loci that were associated with morningness, seven of which were located near genes involved in circadian rhythms. As further confirmation that self-reported information is useful, a survey question about psoriasis validated most of the well-known genetic associations with the condition and found several new associations that were not significant in other genome-wide association studies, Scheller said.

Self-reported medical data have enabled the efficient replication of more than 180 genetic associations (Tung et al., 2011), which, Scheller said, "is reassuring because I came to 23andMe to try to use this data to find drug targets." The database now contains large cohorts of people with specific conditions, including Parkinson's disease, cancer, depression, asthma, cardiovascular disease, and colorectal cancer, which are now the subject of genome-wide and phenome-wide association studies.

23andMe also has the ability to contact subpopulations to ask if they will participate in a study. "We have a target that we believe might be interesting in non-allergic asthma, but we needed to understand the natural history of the disease better before we could design a clinical trial," Scheller said. "So we sent out an email to 80,000 of our customers [with] asthma, and in 2 weeks we set up a survey where 8,000 people will answer questions once a month for 6 months."

PARTICIPATION, DATA SHARING, AND PRIVACY

A major consideration in collecting information from large cohorts is assuring the participants that their data will remain private. In single studies and collaborative studies where data are shared, patients need to know that their information will not be mishandled, Stefánsson said.

Once data privacy is assured, an argument can be made that data sharing among cohort members and among researchers is important. The sharing of information has led to many discoveries and new treatments, and patients should be willing to continue to share data so that the next generation can have access to improved health care, Stefánsson explained. Reflecting this perspective, Daly speculated about shifting to an opt-out rather than an opt-in model for patient participation in biobanks and research, where the default would be sharing. He also wondered if the academic model of publishing is slowing the dissemination and use of results.

A question was raised about participant compensation for participation in genetics-based cohort studies. "It is considered to be a violation of elementary fundamental principles of bioethics to pay people large sums of money for participation in research, and I think it is really inadvisable to propose to do that," Stefánsson said. Baras observed that some people are very willing to participate in follow-on research, though not everyone is.

Researchers, too, can be seen as having an obligation to share data among themselves, in part because it increases the rate at which new drugs can be delivered to patients, said Lon Cardon, the senior vice president of alternative discovery and development and head of target sciences at GlaxoSmithKline. "Those groups that learned to share earliest got their findings first," he said. There has been extensive data sharing in psychiatric genetics, though the topic has turned out to be extraordinarily difficult, Stefánsson said.

Several workshop participants commented that sharing is common in

many areas of genomic research. Regeneron, like all companies, must work within the confines of local laws and regulations, but sharing data with collaborators happens all the time, Baras said. Stefánsson pointed out that companies like Regeneron and Amgen are collaborating on genetics research, while successfully competing in drug development. "It's not just a willingness to collaborate," he said. "People are doing it."

Nordic countries do have some restrictions on the use of genetic data, Stefánsson emphasized. For example, laws in Iceland prohibit exporting genetic data or giving people free access to the data that have been gathered. But researchers nevertheless collaborate extensively and productively using the Icelandic data. "Anyone who comes to Iceland can have access to this data and work with [investigators] there," he said. Performing genomics research studies in the United States can be easier than doing so elsewhere, Baras said, because privacy laws and informed consent procedures are well established and constantly being refined. In certain circumstances, health systems in the United States share their data and health records, including with private companies, Baras said. By contrast, other countries may have laws about data leaving the country that require special procedures.

However, data sharing among investigators can heighten concerns among cohort participants about privacy, Stefánsson said. "It is a formidable challenge to convince people to participate at a high rate at the same time as you are making the primary data available to a very large number of people," he said. "You have to find some sort of a middle ground, [such as] sharing summary-level data." To some extent, people have a tendency to overestimate the dangers posed by sharing genetic data. Sharing stories widely about the success of genomics-driven drug discovery may take some of the fear out of it, Baras said.

Technological solutions may reduce concerns about the risk of data sharing. People are not going to accept putting all of their genomic and medical data on the Internet, Daly said, but novel informatics and computational solutions could generate results while addressing privacy concerns. It would be challenging to enact a law like Finland's Biobank Law[5] here in the United States, Daly said. "It would take quite a shift in understanding," he said. Educating the public on the potential benefits of genomics-driven drug discovery is critical for combating the distrust, Scheller said. The 23andMe experience has demonstrated that many people are willing to have their data used for research and to become en-

[5]The Biobank Act of Finland can be found at http://www.finlex.fi/laki/kaannokset/2012/en20120688.pdf (accessed July 12, 2016).

gaged with genetic information. "If they feel as though they are part of the process, they will be engaged and will let you use their data," Scheller said.

Another important issue with sharing of genotypic and phenotypic data is the quality of the data. Several international disease-focused consortia have addressed data quality challenges by agreeing on effective diagnostic and general criteria, Daly said, and those criteria have led to harmonious research tools that can be used across populations in the same disease state. Data quality issues are not insurmountable, Stefánsson said; even self-reported data from 23andMe is of high quality and has led to a lot of discoveries.

3

Discovery Activities Related to Genetic Bioresources

Important Points Highlighted by Individual Speakers

- New informatics tools based on genomic data can provide a systems approach to drug discovery and development, enabling in silico experiments that complement traditional approaches to drug discovery. Informatics approaches can reveal the complexities of pleiotropic effects of genes and drugs that come into play in medical interventions. (Altman)
- Enhancing target validation is one way to bridge the gap between basic research and medical applications. Much of this research can be in the precompetitive realm, which enables collaborations that can accelerate progress. (Cardon)
- If research on potential drug targets becomes a precompetitive activity, it could reduce competition and thus slow down the discovery process. (Stefánsson)
- Genetic approaches in combination with data collected from patients at the epidemiological and clinical levels could help to bring about new drugs and a more complete biological characterization of a disease. (Chowdhury)
- Understanding the role of allelic variation in target genes is critical to improving success in early drug discovery. (John)
- The development of effective and safe new drugs requires a more complete understanding of the molecular, cellular, and organismal mechanisms underlying potential drug targets and pathways. (Rolph)

The availability of large genetic bioresources creates the possibility of pursuing new approaches to drug discovery and development. Speakers in the second session were asked to describe selected current discovery activities enabled by new genetic cohort studies and to explore opportunities for cross-sector engagement.

The use of bioresources enables many novel approaches to research, including sophisticated analyses, new approaches to target validation and biomarker development, and studies of allelic variation in drug response. More broadly, large cohort studies are enabling a greater understanding of the biology of disease and of drug–target interactions, which facilitates the discovery of drugs to treat a wide array of diseases. This chapter also introduces ideas for incorporating patient perspectives and data in genomics-enabled drug development collaborations.

INFORMATICS APPROACHES TO UNDERSTANDING PHARMACOGENOMICS

The Pharmacogenomics Knowledge Base (PharmGKB)[1] developed at Stanford University is a comprehensive resource in which information about the impact of genetic variation on drug response has been curated for researchers and clinicians. The most popular feature of PharmGKB is the collection of pathways that it details, each of which has been curated from the literature and is supported by PubMed references, said Russ Altman, the Kenneth Fong Professor and a professor of bioengineering, genetics, medicine, and (by courtesy) computer science at Stanford University. The molecular pathways detailed in PharmGKB make possible a systems approach to drug discovery and understanding pharmacogenomics. PharmGKB provides a "platform to do new research projects on how drugs work, how they cause side effects, how they interact, and how we can discover new drugs," Altman said. Informatics approaches allow researchers to integrate data on genes and drugs across many different scales, including the molecular, cellular, physiological, and population levels, he continued.

One example of the molecular information that PharmGKB provides is details about promiscuous drug binding, which can result in adverse side effects. Nonspecific drug binding can result from the interaction of a small molecule with proteins that have similar molecular structures to the

[1]For more information on PharmGKB, see https://www.pharmgkb.org (accessed June 1, 2016).

target protein, a phenomenon that can be detected through three-dimensional modeling (Liu and Altman, 2011). It is easier to understand side effects when the potential interactions of a small molecule with a wide array of proteins—and not just with the intended targets—are examined, Altman said. Another useful informatics approach featured on PharmGKB involves the text processing of millions of research abstracts to enable a high-fidelity extraction of information about the relationship between genes and drugs. As an example, Altman explained that variations in the *ABCB1* gene can affect a patient's response to verapamil, a calcium channel blocker; PharmGKB's text processing feature distinguishes, identifies, and describes these variations in response.

Another approach that PharmGKB researchers have taken is to process electronic health record (EHR) data, thereby detecting harmful and often unanticipated effects of drugs and drug combinations at the population level. For example, a statistical model was used to recognize two drugs that adversely altered blood glucose levels based on the adverse event signature reported through the U.S. Food and Drug Administration's Adverse Event Reporting System, a database that contains information on adverse events and medication errors. The statistical model revealed that pravastatin, a statin, and paroxetine, a selective serotonin reuptake inhibitor used to treat depression, can significantly increase blood glucose when taken together, even though neither is particularly associated with hyperglycemia (Tatonetti et al., 2011).

Together, these and other informatics approaches make it possible to test hypotheses in silico before scientists carry out experiments and prioritize them according to which appear most promising, Altman said. Discovering and developing effective and safe drugs requires an extensive amount of research in areas such as target structure and dynamics; drug recognition and binding; cellular response and molecular pathways; gene, drug, and phenotype associations; clinical responses; and population effect reporting, he said. This is in contrast to the approach that led to the discovery of *PCSK9* as a target for lowering low-density lipoprotein (LDL) levels, which involved defining a relevant phenotype (low LDL levels), finding phenotypic outliers, identifying causative variants, showing that a loss of function has positive effects, and developing and refining a targeted inhibitor. Altman noted that in comparison to the *PCSK9* example, drug discovery often involves a more comprehensive approach that includes defining relevant cell types and tissues; defining relevant molecular pathways; identifying genes and variants associated with pathway function and dysfunction; finding loss-of-function pathway

opportunities while seeking genetic evidence for modulation; conducting counter-screens for promiscuity, side effects, or efficacy issues; and developing and refining inhibitors of pathways.

In summary, Altman pointed out specific needs that could accelerate genomics-driven drug discovery (see Box 3-1). He also offered four primary observations about using informatics to advance drug discovery efforts. The first is that the majority of genetic contributors to disease consist of loss-of-function variants. Creating drugs to treat these diseases requires figuring out how to add back function, which is challenging. In contrast, when a loss-of-function mutation offers protection from a particular disease, it is relatively straightforward to develop a therapeutic strategy, he said.

Secondly, data integration across multiple scales—including molecular, cellular, physiological, medical (as represented in EHRs), and population levels—is essential for minimizing the risk of the drug discovery process. One reason why informatics is powerful, Altman said, is that it is not sensitive to the scale of the data, because "we can easily move between scales." For example, he said, informaticists can study and curate data across molecular, cellular, systematic, and population levels.

The third point Altman emphasized is that drugs, like genes, have vastly underestimated pleiotropic effects that need to be considered during drug development. Drugs can cause unintended adverse effects when they bind to "off-target" proteins (Karczewski et al., 2012).

Finally, considering biological pathways instead of single targets will be helpful in understanding how pleiotropic genes and drugs modulate biology and for predicting side effects and poor efficacy. Altman said that

BOX 3-1
Specific Needs That Could Accelerate Genomics-Driven Drug Discovery (as proposed by Russ Altman)

- A greater understanding of three-dimensional structures of human proteins as a way to understand drug promiscuity, network effects, and the influence of variation.
- Increased knowledge about the gene-expression response of tissues to drug exposures.
- Tissue-specific model systems to test pathway modulation.
- The use of large genetic cohorts to understand gene tolerance for mutations and the gain and loss of function responses of pathways.

he and his informatics colleagues think in terms of networks rather than targets. "Looking just at the target won't be the answer in 90 percent of the cases, but looking at its pathway may offer opportunities," he said.

USING HUMAN GENETIC DATA TO SUPPORT TARGET PRIORITIZATION

Drug discovery programs need to be set up so that targets in cell and model systems can be modulated and assayed in ways that recapitulate relevant parts of human physiology, said Sally John, the vice president of computational biology and genomics at Biogen. It is the translational piece, along with a better understanding of the biology of the targets, that is going to accelerate genomics-based drug discovery, she said.

John illustrated this point by describing work that she and her colleagues have done on salt-inducible kinase (SIK) inhibitors for the treatment of immune-mediated disease. Previous research demonstrated that inhibition of SIKs had the desirable effect of attenuating inflammatory cytokines while at the same time enhancing anti-inflammatory cytokines (Clark et al., 2012). Further research showed that these anti-inflammatory effects were mediated by one specific SIK isoform, *SIK2*. John and her colleagues went on to look for genetic data linking *SIK2* with inflammatory diseases, and through a literature search they discovered that variants in *SIK2* are associated with primary sclerosing cholangitis, an autoimmune disease of the liver (Liu et al., 2013). Meanwhile, data from the University of Cambridge Mendelian Randomisation Catalogue[2] indicated that the same *SIK2* variant had a protective effect for cardiovascular disease, indicating that drugs that inhibit SIK2 function may pose a potential safety concern, she said. Data from 23andMe also confirmed that the *SIK2* SNP carried an increased risk of immune phenotypes and a decreased risk of cardiovascular disease, John said. Examining data from multiple genetic bioresources was extremely important, she said, because they supported the idea that the inhibition of SIK2 may be beneficial for immune-mediated disease, but the data also pointed out potential safety concerns. If researchers wish to move forward and design a selective SIK2 inhibitor,

[2]The University of Cambridge's Mendelian Randomisation (MR) Catalogue is a curated database with publicly available results from large-scale genetic association studies. The MR Catalogue has been recently renamed to PhenoScanner, and the website is available at http://www.phenoscanner.medschl.cam.ac.uk (accessed June 17, 2016).

they may need to enlist a team of chemists to try and avoid off-target effects, she said.

To gain a better understanding of functional coding variants, Biogen started collaborating with several other companies and nonprofit organizations as part of the Industry Partnership for Human Genetics, John said. Focusing on about 50 target genes of interest, the partnership carried out a comprehensive assessment of loss-of-function and gain-of-function variants and enrichment in the Finnish registries across the genes. The phenotypic data were queried according to specific hypotheses based on each individual genetic target of interest, and association analyses of distinct alleles were done across the entire phenotypic spectrum.

The results revealed the cardioprotective effects of loss-of-function *PCSK9* variants. Individuals with these variants were much less likely to use lipid-modifying agents. However, they were more likely to use drugs prescribed for mental illness and were more likely to be hospitalized for psychiatric disorders. "Is this a potential safety indication?" John asked "I don't know whether these types of drugs and indications would overlap with the newer cognitive effects we are seeing occasionally on PCSK9 inhibition, but this is the sort of data that we would pay attention to."

Getting a complete picture of the role of allelic variation on target genes of interest is critical to improving success in early drug discovery, John concluded. In particular, she said, a better understanding of the biology leads to improved hypotheses.

STEPPING STONES TO UNDERSTANDING THE UNDERLYING BIOLOGY OF DRUG TARGETS

Understanding the biology of drug targets can increase the likelihood of successful drug discovery and development, said Tim Rolph, the vice president of program value enhancement at Pfizer Inc. It takes considerable time to understand molecular targets and disease mechanisms well enough to devise effective treatments, he said. Many view the development of PCSK9 inhibitors as an ideal model for genetics-driven drug discovery; however, the knowledge about cholesterol metabolism contributed by Nobel prize–winning researchers Michael Brown and Joseph Goldstein greatly aided the process, Rolph said.

There are several critical stepping stones that arise early on the path to creating novel medicines, Rolph said. The first step is to design a primary

screen using small molecules or antibodies to detect a phenocopy of a genetic variant. For example, the existing knowledge of PCSK9 biology enabled the design of an in vitro screen to detect monoclonal antibody phenocopying of the *PCSK9* loss-of-function variant. "We understood how PCSK9 interacted with the LDL receptor in terms of the extracellular domain," Rolph said, "and we constructed and immobilized the extracellular domain binding assay." Although the process seems very simple, it is an important step in the drug discovery and development process, he said.

The next step is to understand exposure–effect relationships by developing secondary screens (ex vivo or in vivo, or both) with a high confidence in the translation of an effect to the clinic. These secondary screens can be on either pathophysiological or physiological mechanisms, Rolph said, as long as they can detect the modulation that is being performed.

The third step is to demonstrate proof of pharmacology during first-in-human trials at a well-tolerated dose. A key aspect of this step, Rolph said, is to be able to translate preclinical pharmacokinetic-phamacodynamic (PK-PD) modeling to predict human PK-PD in the clinic so as not to rely on a "leap of faith" when trying a medicine in humans.

The fourth and final step is to demonstrate clinical proof of mechanism during short-duration studies in target patients across key safety biomarkers. "Positive proof of mechanism gives us great confidence to go and invest in clinical proof of concept," Rolph said. If there is not a clear understanding of the phenotype of a particular genetic variant at the molecular and cellular level, the chances of developing an effective drug are low, he concluded.

A PUBLIC–PRIVATE COLLABORATIVE APPROACH TO TARGET VALIDATION

The translation of 21st-century genetics into diagnostics, prognostics, and treatments has had many successes, especially in the areas of oncology, rare diseases, and drug safety, said Lon Cardon, the senior vice president of alternative discovery and development and head of target sciences at GlaxoSmithKline (GSK). However, in many areas the gap that exists between research results and medical applications is as wide as ever, he noted. "Where is genetics going to have its promise," he asked, "and what are the first areas of translation where the impact is going to be beyond oncology, rare diseases, and adverse events?"

One way to bridge the gap between discovery and clinical use is through the use of more robust target validation, Cardon said. As noted in Chapter 1, drugs with human genetic information are more than twice as likely to be successful clinically as those without such information (Nelson et al., 2015). Drugs that successfully reach patients are more likely to have genetic validation, Cardon said, and failures at each stage are more likely to occur with drugs without genetic validation. However, only about 10 to 15 percent of targets currently being pursued by pharmaceutical companies have genetic support for a specific indication, Cardon said. If the number of new targets with genetic support increased to 50 percent, there would be a 13 to 15 percent reduction in the overall cost of drug development, he noted. This would result in significant savings, as it was recently estimated that the average cost of developing and marketing a new drug is $2.6 billion dollars (Tufts Center for the Study of Drug Development, 2014).

However, it is important to remember that not all genes can serve as good drug targets, Cardon continued. A gene may provide a good predictor of a disease, but it is often the case that multifaceted research is needed to turn a predictor into a valid drug target. Among the many factors that need to be determined are the mechanism of action, pleiotropy, gene regulation, the gene's position within a pathway, tissue specificity, and its chemical tractability, Cardon said.

To address the need for research on a wide variety of factors to validate drug targets, the Centre for Therapeutic Target Validation (CTTV)[3] was established. CTTV is a public–private collaboration designed to harness the power of big data and genome sequencing information to improve the success rate for discovering new medicines. The three original founding organizations of the CTTV were GSK, the European Bioinformatics Institute, and the Wellcome Trust Sanger Institute. Recently, Biogen joined the effort, and other potential members have expressed interest in joining. The CTTV was founded on the premise that the study of drug targets themselves belongs in the precompetitive realm, Cardon said. Pharmaceutical companies compete in other areas, such as chemistry, clinical trials, and linking targets to relevant phenotypes. But in the precompetitive realm, companies can share research results without losing their competitive advantage, he said. Furthermore, validating targets can sometimes require a combination of skills that are not available in any one organization, public or private. To that end, the CTTV brings

[3]For more information on the Open Targets platform, part of the CTTV initiative, see https://www.opentargets.org (accessed June 2, 2016).

together people with different types of expertise to work collaboratively on projects. The concept of the CTTV is very exciting, Cardon said, because experts in the areas of drug discovery, functional genomics, and EHRs operate next to one another, all asking different questions of the data. Members of the center formally agree to pool their expertise and to share findings openly. The center is also a way to train a new generation of translational scientists, Cardon said. This collaborative concept appears to be taking off, he said, and although GSK provided the initial funding, it is reassuring that other organizations have now joined in as well.

Cardon also acts as an advisor to the Precision Medicine Initiative (PMI),[4] which has the goal of enabling a new era of medicine through research, technology, and policies that empower the development of individualized care. He said that the CTTV is focused on the earliest stages of drug discovery, which complements the efforts of initiatives like the PMI. "The PMI will follow from these drug discovery target validation activities just about perfectly," Cardon said, "and we need to find a way [to make] the PMI data equally accessible to all so that we can benefit all parties."

COLLECTING AND SHARING DATA TO TREAT A NEUROLOGICAL DISORDER

The Michael J. Fox Foundation, the world's largest nonprofit funder of research on Parkinson's disease, has the mission of accelerating the development of improved therapies and, ultimately, a cure for people living with Parkinson's disease. The foundation's research focuses mostly on translational to early-stage clinical research, said Sohini Chowdhury, the senior vice president for research partnerships at the foundation. The foundation supports not just the development of therapeutics, but also the tools that are required to transform a finding into a therapeutic, such as models, assays, reagents, and biomarkers.

As an example of the foundation's efforts in this area, Chowdhury cited the Parkinson's Progression Markers Initiative (PPMI)[5], which is an attempt to address the critical lack of biomarkers associated with dis-

[4]For more information on the Precision Medicine Initiative, see https://www.nih.gov/precision-medicine-initiative-cohort-program (accessed June 17, 2016).

[5]For more information on the Parkinson's Progression Markers Initiative, see https://michaeljfox.org/page.html?parkinsons-progression-markers-initiative-get-involved (accessed October 14, 2016).

ease mechanism, drug mechanism, dosage determination, study eligibility, stratification into disease subtypes, and clinical signals. It is a precompetitive initiative, and the foundation is the primary funder of the study and receives additional support from industry partners, nonprofit organizations, and private individuals. The study population includes idiopathic Parkinson's subjects, age- and sex-matched controls, individuals who clinically present with Parkinson's disease but do not show a dopamine deficit, and individuals with and without genetic risk factors. All of the subjects are followed for a minimum of 3 years and a maximum of 8 years and undergo the same array of assessments, including clinical data collection, imaging, and biosampling.

All of the data collected through the PPMI study are accessible. "We share everything," Chowdhury said. "You can go to the PPMI website and download all of the clinical data, all of the imaging data, all of the data that we collect. You can also request access to the samples. To date, we have had 500,000 data downloads, and 70 specimen requests."

This is an exciting and optimistic time for Parkinson's drug discovery and development, Chowdhury said. Several disease-modifying trials are in early-stage clinical testing, and genetic discoveries have provided tractable targets. But genetics can only take researchers so far in drug development, she said. Parkinson's is a complex disease whose effects go far beyond what is occurring in a patient's neurological system. As a result, Chowdhury said, all therapeutic development for Parkinson's disease needs to begin with thorough studies at the epidemiological and clinical levels that take into account the patient experience as part of the biological characterization of the disease. "It has to all come back to the patient and the experience of the patient," she said. Traditional disease research management holds that drug development unfolds in a linear fashion from basic research to the clinic. However, drug development should not be thought of as linear, Chowdhury said. Instead it should be viewed as a cycle, she said, and it is important to look for biomarkers in addition to drug targets because there is a critical need to measure the progression of the disease in actual patients.

The members of the foundation believe that achieving this perspective requires breaking down the siloes that exist between basic science, drug development, clinical research, and biomarker studies, Chowdury said. These fields should be communicating and working together, she said, and the research should all be grounded in the patient experience and seen through that holistic lens. Such a holistic view also

requires involving patients as partners, she added, elaborating as follows: "How are you getting the patient perspective? How are you getting access to patient samples? You have population-based studies, which are extremely important, but you also need studies that are focused on individuals with the disease that you are going after, because that's a very unique perspective."

Chowdhury noted that technology has made it possible for even small groups of patients with the disease to connect. The research landscape is changing, and integrating patients is important, she said. Patients who are living with a disease can be a valuable resource, she added, because they are often more willing to share their data than members of the general population, because they are in a position to benefit directly by doing so.

EXPLORING THE DEFINITION OF THE PRECOMPETITIVE RESEARCH SPACE

As discussed in previous sections in this chapter, one issue to be considered in a discussion of genomics-enabled drug discovery using data from large bioresources is the role of precompetitive research. (See Chapter 4 for more information on business models that support collaboration.) In the drug discovery process, there are research areas that certain groups view as precompetitive, but others see as competitive. In an effort to encourage collaboration and increase efficiency it may be useful to explore the early steps in drug discovery research and understand where all parties are willing to work together. Chowdhury advocated for "putting stakes in the ground" around areas that are precompetitive. "We need to start to think about what is competitive versus precompetitive," she said. For example, both the development of tools and clinical datasets can be precompetitive. Although following individuals for long time periods is expensive, she said, working collaboratively on tools, reagents, and assays, while perhaps legally complex, is not that cost intensive.

As an example of research that is largely precompetitive, John cited research on cholesterol metabolism. Another example, she said, is developing additional biomarkers. "If we can do that precompetitively and come up with more robust biomarkers that we can use in the clinic, that would accelerate our ability to translate basic genomics," she said. Genomics has been a field marked by collaboration, she noted, which facilitates precompetitive research.

One way to draw the line between precompetitive and competitive, Cardon suggested, would be to label the identification of targets as precompetitive while labeling experiments on those targets as competitive. However, he added, "where it gets complicated to me is when you start bringing molecules into it. Where do you draw that line, and how do you distinguish that, because molecules are also tools that can interrogate the targets themselves for validation?"

As an example of this distinction, Chowdhury noted that The Michael J. Fox Foundation spearheaded an initiative with three pharmaceutical companies that are all targeting the kinase LRRK2. All three companies shared their tool compounds with the foundation to determine whether a safety profile that was identified in one compound was actually relevant to all compounds that are targeting LRRK2. Precompetitiveness means identifying what is going to benefit the group as a whole, and in this case, all parties benefited from knowing the safety initiative data, Chowdhury said. This is an issue in academia as well as in industry, she continued. It is a field-wide challenge to understand that creative precompetitive arrangements can take place while still safeguarding proprietary interests, she said.

On the other hand, Stefánsson expressed some skepticism about the idea of making the target or research on the target precompetitive. "The target is not just a molecule," he said. "The target is a molecule in a context. . . . The science is always going to be focused on the discovery and the characterization of the target in the context that you want. So I think it would be devastating if we would make the target precompetitive. That's where we should be competing. That's where it's going to be exciting." Scientific information becomes public eventually, but between discovery and publication, industry can use that information to compete, Stefánsson emphasized.

Cardon responded, "If you don't want to share it, you don't share it. There is no obligation to share anything you work on. [But] we are tripping over each other making the same mistakes, doing the same experiments, reproducing failure, not knowing where to look in the pathways. It's more efficient from a societal and from a patient perspective to think of that part of the game as something we could share."

As part of the Industry Partnership for Human Genetics, John said, companies submitted a number of genes of interest to academic researchers. All of the participants saw the full list of genes, but they did not know which company had submitted which gene. Furthermore, participants were not aware of the indications that companies were interested in

pursuing. Biogen was very happy with this type of arrangement, John said, because it received in return all of the phenotype associations with its genes of interest.

Relatedly, there is a distinction between genotype–phenotype correlations that come from patients' DNA and downstream innovation, which is not necessarily performed by the same people uncovering the correlations, Daly said. "How to set that boundary in a way that maximizes the patient resources being used and available to everyone but retains the motivation that's needed is worth fleshing out further," he said.

For example, Cardon said, in the context of the PMI, all parties need access to information in the cohort study database. The initiative is patient-focused, but it also accommodates scientific and commercial interests. The generation of insights that lead to therapeutic hypotheses is going to remain precompetitive to some degree, Rolph observed. However, different models have been successful, such as those used in Iceland, in the U.S. commercial health care system, and with direct-to-consumer genetics companies, he said, and it is worth exploring the precompetitive space in each of these models.

INDUSTRY–UNIVERSITY RELATIONS

Another area related to the precompetitive space that can present challenges is the relationship between industry and university researchers, said John Carulli, director of precision medicine at Biogen. In terms of precompetitive investment by the pharmaceutical industry, he asked, which topics are the most important to fund? Because of the complexity of drug development, Cardon said, it is no longer an efficient approach for industry to blindly give money to university researchers and hope that something comes back. Both basic science and regulation are getting more sophisticated and complex, he said, and there needs to be novel funding mechanisms and new ways to bring together the expertise of relevant individuals.

University researchers can be fearful of translational research at times, Altman said, but that attitude can change quickly once a research project reaches its full potential. Furthermore, many young researchers have gotten interested in translational research, including those with a background in computer science and engineering. However, a missing ingredient is the expertise of clinician scientists, whose numbers are declining (NIH, 2014). "That is something we have to work on in terms of the educational pipeline," Altman said.

4

Business Models That Support Bioresource Discovery and Collaboration

Important Points Highlighted by Individual Speakers

- A patient-centered approach to using large-scale genomic analyses for the discovery of targeted therapies can foster the acceleration of research in this area. (Matrisian)
- A registry of information obtained directly from patients on their experiences and clinical outcomes is a valuable supplement to genetic and clinical data and may create new opportunities for drug discovery. (Matrisian)
- Pooling the expertise and resources of several companies in a precompetitive environment could support accelerated target identification, clinical development, patient stratification, and pharmacogenetics discoveries. (Ehm)
- Public–private partnerships with well-defined deliverables and timelines can facilitate the discovery and validation of new targets that companies can incorporate into their therapeutic development programs; provide new insights into known, existing targets; enable a significant increase in the knowledge of tractable disease biology and disease pathways; and create a rich, comprehensive, integrated knowledge base that is easy to use and available to the entire global research community. (Wholley)
- The Structural Genomics Consortium demonstrates how government agencies, philanthropies, private companies, and university researchers can come together to pool resources and share risks in an effort to discover and develop new medicines. (Bountra)

Multiple business models, including models in the precompetitive space, can support collaborations that use bioresources for discovery purposes. These models generally include multiple stakeholder groups across sectors, because each sector has strengths and resources that it can bring to a collaboration. Any given collaborative effort may be based in the private sector, the public sector, or the nonprofit sector, which includes patient advocacy groups. But in each case, as demonstrated by the examples discussed at the workshop, the goal of individuals in an effective collaborative model is to be part of and draw on broader systems of expertise and capabilities from a variety of stakeholder perspectives. Speakers in the third session were asked to examine potential precompetitive business models and investments that can support discovery efforts and opportunities across stakeholder groups.

ENGAGING PATIENTS FOR DRUG DISCOVERY

Pancreatic cancer is the only major type of cancer with a 5-year survival rate in the single digits and is a health threat in the United States and around the world, said Lynn Matrisian, the vice president of science and medical affairs at the Pancreatic Cancer Action Network (PanCan). In 2010, the network adopted a goal to double the survival rate for pancreatic cancer by 2020. PanCan has adopted a comprehensive approach to advancing research, which involves a central government affairs office in Washington, DC, affiliate networks in more than 60 U.S. cities, a research program to support translational and clinical research, and a patient services group. The patient services group, called Patient Central, operates a call center that receives more than 12,000 requests for information per year from patients and their families. Representatives at Patient Central can link callers to pancreatic cancer specialists in their area, give them information about educational events and webinars on pancreatic cancer, and help them find clinical trials in which they can enroll. Through this effort, the network aims to improve patient access to data and participation in clinical trial research.

Recognizing that genomic information can accelerate progress, PanCan has created a program called Know Your Tumor that over 2 years has enrolled more than 500 patients from across the United States. Patients are recruited not just from academic institutions but also from community settings, Matrisian said. Biopsies taken at the point of care are sent to a central location, where tissues are then allocated to different

molecular diagnostic companies. The results for the first group of patients have included information on a 343-gene panel designed for all solid tumors and an analysis of 23 separate proteins related to pancreatic cancer. This work has resulted in the identification of a diverse set of genes thought to be implicated in pancreatic cancer that can be used as a starting point for drug development programs (see Figure 4-1). The work can also be used to guide the selection of appropriate targeted treatments for an individual patient.

Different subtypes of cancer have distinct targets, treatment indications, biomarkers, therapies, and patient subpopulations, Matrisian said. However, she added, the approach taken by PanCan "leverages what we know in [other cancers] and takes advantages of the distinctions." This approach would not be possible without large-scale genomic analysis and targeted therapies, she said.

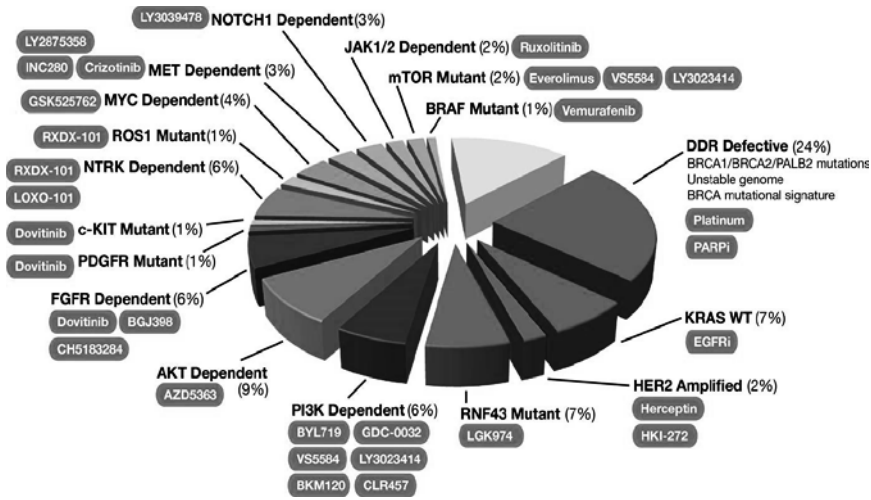

FIGURE 4-1 The actionable genome for pancreatic cancer contains a wide variety of potential drug targets.
NOTE: Figure was developed by David Chang, Peter Bailey, and Andrew Biankin, University of Glasgow, and is based on data from 457 patients in the International Cancer Genome Consortium (ICGC PACA-AU) cohort. Percentages correspond to the percent of total pancreatic cancers with mutations and/or copy number alterations in the indicated gene or pathway, and shaded ovals indicate possible therapies based on the molecular aberrations.
SOURCE: Lynn Matrisian, National Academies of Sciences, Engineering, and Medicine workshop presentation, March 22, 2016.

PanCan also maintains a comprehensive clinical trial database on all pancreatic cancer trials that are currently open in the United States, Matrisian said.[1] By using the Clinical Trial Finder, patients get personalized information on trials that they may be eligible for. It is estimated that in 2014 only 4.2 percent of newly diagnosed pancreatic cancer patients entered a clinical trial, Matrisian said. However, among the subset of patients who contacted PanCan for information, 15.5 percent entered clinical trials. In addition to its role in providing information for patients, the organization's database can be used by health care professionals to easily survey the landscape of clinical trials in order to identify gaps and plan new trials. Trial sponsors can also receive information on how many times information about their trial is given out.

PanCan has contact with more pancreatic cancer patients than any other organization and is now collaborating with Genetic Alliance on its Platform for Engaging Everyone Responsibly (PEER) registry system. On the PEER registry, patients can record their experiences and outcomes, and can manage via the PEER portal their preferred privacy settings and access to data they have provided, Matrisian said. Depending on the desires of the patients, certain data are made available to researchers. This system allows researchers to access another layer of information that supplements clinical records and thus creates opportunities for novel discoveries, Matrisian said.

COLLABORATIVE APPROACHES TO GENOMIC DRUG DISCOVERY

Utilizing EHR-Linked Biobanks for Drug Discovery

Previous efforts to assess the impact of genetic variation on clinical phenotypes have relied on population studies performed through genetic analysis consortiums, said Meg Ehm, the director of external strategic alliances for genetics at GlaxoSmithKline (GSK). However, the phenotypes used in these studies have often been reduced to a "common denominator," she said, and the studies rarely have produced longitudinal information. Only a few of these collections have included information on drug usage or drug response, and many conditions have not yet been studied, Ehm said.

[1]For more information on the Pancreatic Cancer Action Network's clinical trial database, see https://clinicaltrials.pancan.org (accessed June 8, 2016).

To address these issues, GSK is focusing its efforts on building a new entity, the Genomic Resources for Drug Discovery Consortium, which will utilize electronic health record (EHR)-linked biobanks more effectively for drug discovery purposes. The new consortium is designed to support target identification, clinical development, patient stratification, and pharmacogenetics. The overarching goal of the consortium is to realize the value in integrated medical and genomic resources in order to identify and prioritize targets and to better understand drug response, Ehm said. The consortium has five key features: comprehensive and diverse health data, comprehensive genetic data, access to biological samples, informatics, and the ability to recontact patients.

The consortium was designed with several key tenets in mind, according to Ehm. The first is the idea that it is important to develop a research-enabled environment that facilitates the precompetitive space, because accelerating drug discovery will require more resources than any one company can provide. The second is that engaging users early in the design and development of genomic and EHR-linked bioresources is critical if one is to leverage unique insights into drug discovery and development questions. Finally, consortium developers recognize that the harmonization of multiple resources, such as data collected from many types of health settings and from participants of all ages and ethnicities, will be needed to realize the full value of the partnership.

As an example of how the consortium could help research to progress more efficiently and effectively, Ehm cited work on the identification of a rare missense variant in the glucagon-like peptide 1 receptor (*GLP1R*) gene. GLP1R agonists, which are used in the treatment of type 2 diabetes, act as mimics of the incretin hormone GLP-1 and increase insulin levels in response to orally-consumed glucose (Scott et al., 2016). The rare missense variant that was discovered in *GLP1R* was found to mimic the effects of GLP1R agonists, so researchers combined genetic and phenotypic data using in silico approaches to test the association of the variant with disease outcomes. This analysis revealed that the variant was associated not only with a reduced risk of type 2 diabetes, but also with a reduced risk of coronary heart disease, thus providing supportive evidence that GLP1R agonists are not likely to be associated with an unacceptable increase in cardiovascular risk.

The Genomic Resources for Drug Discovery Consortium grew out of enthusiasm about associating genetic variants with clinical outcomes, but participating companies also intend to work toward developing evidence that links potential drug targets to disease progression, predicting out-

comes of target modulation, and using systems biology and systems pharmacology to facilitate these goals.

A Public–Private Partnership to Accelerate Genomic Medicines

The Foundation for the National Institutes of Health (FNIH) is a nongovernmental organization established by an act of Congress to develop public–private partnerships that support the mission of the National Institutes of Health (NIH). One project managed by the foundation is the Accelerating Medicines Partnership (AMP), which brings high-level government, industry, and nonprofit foundation partners together to identify and validate the most promising biological targets for therapeutics. The specific goals of AMP are to discover and validate new drug targets that companies can incorporate into their therapeutic development programs; to provide new insights into known, existing targets; to enable a significant increase in knowledge of tractable disease biology and disease pathways; and to create a rich, comprehensive, integrated knowledge base that is easy to use and available to the entire global research community, said David Wholley, the director of research partnerships for FNIH. Launched in early 2014, AMP brings together about 20 companies, NIH, and multiple nonprofit organizations to do precompetitive research and share data broadly and quickly, with the funding split between the public and private sectors. To ensure the broadest possible opportunity for commercialization, preemptive patenting is not allowed, and data become freely available as soon as quality control is finished, Wholley said.

AMP is focused on three main areas: Alzheimer's disease, type 2 diabetes, and immune-related disorders such as rheumatoid arthritis and systemic lupus erythematosis. Within each of these disease areas, the stakeholders involved in the partnership set timelines for deliverables. AMP is governed by steering committees that are co-chaired by a representative from the NIH and an industry partner. The steering committees "are really designed to make sure that the research is continuing on track and according to our milestones and goals," Wholley said. Steering committee members are personally committed to the goals of the program and make it a point to join each and every meeting, Wholley continued.

As an example of how AMP is using genetic and genomic data, Wholley described efforts within the program to find new medicines to treat type 2 diabetes. In order to identify potential drug targets for type 2

diabetes, AMP has mapped out a 5-year program that links human genetic data on disease risk to phenotypic data. Similar to the Genomic Resources for Drug Discovery Consortium, researchers in the AMP diabetes program realized that they could accelerate progress by aggregating their data, as opposed to working separately, Wholley said. Though funding procedures and workflows have been complicated at times, AMP has succeeded in building a knowledge portal that provides aggregated human genetic data from more than 150,000 individuals across a wide range of ages and ethnicities, and has created tools within the portal to allow easy, integrated interrogation across multiple datasets while maintaining individual-level data privacy, he said.

The challenges for this type of collaboration include both incentivizing investigators to share data and managing data restrictions, such as consent and regulations on exporting data out of countries, Wholley said. Another difficulty has been data integration, which is challenging because of the heterogeneity of the data, analytical needs, funding sources, legacy support, and data platforms. It has been critical to address concerns about publications, authorship, and acknowledgment, he said, although intellectual property issues have not been a concern because of the well-defined precompetitive agreement that all partners must agree to before joining the collaboration. Members also debate about how to prioritize spending. Finally, aligning interests and perspectives between government, academic, and industry scientists can present difficulties, Wholley said.

The Structural Genomics Consortium: Pooling Resources and Sharing Risks

Three major challenges confront researchers working on drug discovery and development both in the public and private sectors, said Chas Bountra, a professor of translational medicine at the University of Oxford. The first is target discovery—the identification of molecules (e.g., proteins) that can be modulated to treat diseases in a subset of patients. The second is the amount of duplication occurring in biomedicine across all sectors. Many groups are working on the same targets, and, unfortunately, many of those targets are destined for failure, he said. The third challenge is the development of high-quality drugs to alter disease pathways.

The Structural Genomics Consortium (SGC)[2] has tried to address these challenges in several ways, Bountra said. First, members of the consortium have pooled resources and shared risk by assembling a network of government agencies, philanthropies, and eight large pharmaceutical companies. Second, through collaborations with hospitals, clinicians, patient organizations, and individual patients, the SGC can easily access such resources as primary human tissues for target validation studies and can maintain focus on patient needs within the context of specific diseases. Third, the consortium is making all of its reagents and tools for drug discovery freely and immediately available. This has generated great enthusiasm for collaborations, so that the consortium is now working with more than 300 academic labs all over the world for free, Bountra said. Finally, the consortium has used the expertise, chemical compound collections, and other resources made available through commercial partnerships to work on proteins or protein families that have been previously viewed as intractable or unusable in a drug development program.

As an example of this approach, Bountra cited work on a family of proteins with a special motif called bromodomains, which were thought to be undruggable due to their complex biology and protein structures. However, in collaboration with researchers at GSK and Harvard University, Bountra and his colleagues identified a novel inhibitor of bromodomain-containing proteins that reduced proliferation in tumors (Filippakopoulos et al., 2010). Since then, Bountra said, the molecule has been given to more than 1,000 laboratories around the world and has been found to be effective not only for a range of cancers but for sepsis, cardiac hypertrophy, male contraception, fibrosis, and chronic obstructive pulmonary disease, as reported in more than 300 publications. Pharmaceutical companies that partnered with the SGC began proprietary programs with the molecule as a starting point. To date, six companies have developed six unique investigational drugs based on the original molecule and are currently testing them in clinical trials, Bountra said.

More recently, with funding from the Wellcome Trust, the SGC has been working with clinicians, geneticists, and disease experts to identify high-priority genes that are most likely to be therapeutic targets. The group then generates the purified proteins, related active molecules, inactive control molecules, and other tools that can facilitate the development

[2]For more information on the Structural Genomics Consortium, see http://www.thesgc.org (accessed July 7, 2016).

of drugs for patients. Twenty-five of these target-enabling packages are being developed in the areas of neuropsychiatry, metabolic disease, cancer, and inflammatory disease. The reagents are peer reviewed by external academic scientists to assess their quality before being released, Bountra said.

With the work of the SGC, "we are trying to generate de-risked targets which industry can then take to generate proprietary molecules, fast track them through large-scale clinical studies, and take them all the way to the marketplace," Bountra said. "We believe this is the right thing to do for patients. It's the right thing to do for society. It's the right thing to do for industry." There is hope that the creation of a new collaborative research ecosystem that involves the pharmaceutical industry, academic researchers, patients, government, and private funders, will generate many more novel medicines, Bountra continued.

THE SUSTAINABILITY OF BUSINESS MODELS

How can collaborative business models for drug discovery be sustainable over time? Successfully meeting short-term program goals is, of course, one key to sustainability, Matrisian said, because it keeps all the partners invested and moving forward. The SGC has been sustainable because of its ability to evolve and take on new research projects once the network was in place and results were being generated, Bountra said, and that evolution has allowed for additional funding. Clearly defining deliverables early on was a significant part of the respective business plans for AMP and the Genomic Resources for Drug Discovery Consortium, according to Wholley and Ehm. Data also need to remain relevant to important research questions for an initiative to be sustainable, Ehm said.

The availability of data generated from large-scale genetic bioresources tends to create opportunities for incentivized challenges, which can democratize the discovery process, Wholley said. Generating large datasets and making them broadly available can attract innovative researchers with an entrepreneurial spirit, he said. Data harmonization and the ability to associate genetic variants with longitudinal outcomes are key practical markers of efficient data accessibility and utility, Ehm said. Making the needs of patients a priority during the design of research studies is also important, Wholly said. One way to ensure that research is focused on the goals of patients is to issue direct challenges and competitions for specific funding opportunities, Matrisian said.

One shortcoming of public–private partnerships, Wholley said, is that few of them have clearly defined metrics for when enough work has been done on a topic or outline plans for how the partnership will prioritize and advance projects. Looking forward to future steps to ensure consortium relevance for the long term is also important.

5

Potential Next Steps in Using Genomics to Advance Drug Discovery

What is the one thing that would enhance our ability to translate insights from human genetics into new medicines within the next 3 to 5 years, asked Nadeem Sarwar, the workshop chair and president of Andover Product Creation Innovation Systems at Eisai Inc. One example of a disruptive technology that could accelerate genomics-driven drug discovery is being developed by the Tissue Chip for Drug Screening program at the National Center for Advancing Translational Sciences (NCATS); it is an in vitro platform that uses human tissues to assess the safety, efficacy, and toxicity of potential drugs. Following a brief presentation on the Tissue Chip, individual workshop participants revisited Sarwar's question and focused on two main topics: potential ways to foster a broader view of biology that fuses genetics with research on many other facets of health and disease; and ideas for sharing information, including data from genetic cohort studies, to enhance progress in developing new disease treatments. This final chapter of these proceedings synthesizes these discussions. Box 5-1 provides a list of possible next steps for genomics-enabled drug discovery efforts that were drawn from remarks made by individual speakers in the final workshop session.

DISRUPTIVE TECHNOLOGIES FOR DRIVING DRUG DISCOVERY

As an example of a technology that could produce disruptive changes in the drug discovery and development process, Danilo Tagle,

BOX 5-1
Possible Next Steps Proposed by Individual Workshop Participants

Broadening the Scope of Research

- Encourage collaboration between basic scientists and clinical researchers as a way to link findings from large genetic cohort studies to the underlying biological mechanisms and pathways. (Ginsburg)
- Engage the broader research and clinical communities to think about using genomic data as a way to break down silos and encourage collaboration. (Daly)
- Explore the potential of polypharmacology, the design or use of pharmaceutical agents that act on multiple targets, to alter disease outcomes. (Altman)

Fostering Collaboration

- Come up with a clearer definition of the realm of precompetitive research and develop standards to enhance collaboration. (Cardon)
- Develop and disseminate a list of best practices based on past successful collaborative frameworks. (Ginsburg)
- Develop new collaborative models that foster cross-disciplinary information sharing. (Daly)
- Establish an online resource that would contain information about the existence of genetic cohorts, types of available data, and ways to initiate collaborations. (Altman, Ginsburg)

Increasing Patient-Focused Research

- Identify the needs and perspectives of patients and incorporate them at the center of genomics-enabled drug discovery research. (Matrisian)
- Broadly communicate to patients and the public the value that is provided by the new tools and resources associated with genomics-driven drug discovery efforts. (Sarwar)

Rethinking Industry's Approach to Genomics-Driven Drug Discovery

- Identify the proportion of drugs in the preclinical pipeline that have been genetically validated and attempt to increase that number over time. (Sarwar)
- Set up mechanisms that encourage the cessation of failed projects at early stages in the drug development pipeline, thus allowing companies to be more nimble and flexible. (Cardon)

the associate director for special initiatives at NCATS, described the Tissue Chip for Drug Screening program.[1] The goal of the Tissue Chip program is to develop an in vitro platform in which human tissues are embedded on a chip to evaluate the safety, efficacy, and toxicity of promising drug candidates. Data from the Tissue Chip program could also assist regulatory agencies with making evidence-based decisions, Tagle said. As part of the program, 10 human physiological systems—circulatory, endocrine, gastrointestinal, immune, integumentary, musculoskeletal, nervous, reproductive, respiratory, and urinary—will be functionally represented by human tissue constructs. The various systems will be captured in a platform that is genetically diverse and has relevance to the physiology and the pathology of the relevant tissues. The chips will be modular and reconfigurable, tissues will be viable for at least 4 weeks, and the materials will be made widely available, Tagle said.

The development of tissues on chips requires input from cross-disciplinary teams consisting of engineers, material scientists, and cell biologists, Tagle said. These teams must take many factors into account when creating the platforms, including computational design, functional readouts, host response, innervation, bioreactors, perfusion, spatial and temporal patterning, structure, cell type, and scaffolding. To survive, the tissues require microvasculature, dedicated bioreactors optimized for the appropriate tissue response, and realistic immune responses, Tagle said. For example, researchers are currently working on a "gastrointestinal system on a chip" that can undergo peristalsis and secrete digestive enzymes. Another system seeks to re-create microvasculature with an embedded colon tumor, Tagle said. The "heart on a chip" has been used to model a rare disease called Barth syndrome and search for potential treatments (Wang et al., 2014).

Tissues on chips have the potential to incorporate many of the genetic resources discussed at the workshop and have numerous potential future applications, Tagle said. In 2016, NCATS plans to use the tissue chip technology to model rare diseases, and it will further expand the program to more common diseases in 2017. There are also plans to harness gene-editing technology to introduce various polymorphisms into induced pluripotent stem cells to look at individual drug responses. "My goal eventually would be to be able to conduct clinical trials on chips," Tagle said, which could make preclinical and clinical research more cost effective.

[1]To read more about the Tissue Chip for Drug Screening program, see https://ncats.nih.gov/tissuechip (accessed June 14, 2016).

LINKING GENETIC DATA TO BASIC RESEARCH ON BIOLOGICAL FUNCTION

Large-scale population cohorts with extensive genetic and phenotypic information are rich sources of information about potential drug targets, said Geoffrey Ginsburg of Duke University in summarizing the session on large genetic cohort design (see Chapter 2). However, he continued, findings from cohorts need to be linked to the underlying mechanistic biology in order to improve drug discovery and development. Russ Altman of Stanford University said that genetics research is an integral component of any data portfolio and can help drive drug discovery, but he cautioned that genetic data on their own are not sufficient to get the job done.

"When we understand the biology, we can make enormous inroads in drug discovery," said Tim Rolph of Pfizer. That has happened with certain types of cancer over the past few years—with childhood leukemia, for example, for which survival rates have substantially increased. The same thing could happen with other common diseases, such as cardiovascular disease, Rolph said. Researchers should focus their efforts on areas of health care where there are unmet needs with regard to the underlying biology and try to create targeted therapies, he said. Having a clear understanding of disease pathogenesis is extremely important, Rolph said, because without that information, drug developers have to take "leaps of faith" without good ways of managing risk.

The need for a broad view of biology was made apparent in the presentations on genome-enabled discovery activities, said John Carulli, the director of precision medicine at Biogen and moderator of the session on current genome-enabled discovery activities (see Chapter 3). Identifying genetic information to help facilitate the drug discovery process is only the first step, he said. The molecular, cellular, and organismal consequences of target modulation—including the effects of such modulation on the health trajectory of patients—are also critical to success in drug development. In-depth biological information is essential, Carulli said, adding, "We need rich phenotypic data, and we need it to be longitudinal."

Polypharmacology, or the design or use of pharmaceutical agents that act on multiple targets, has the potential to increase our understanding of molecular disease mechanisms. It is much easier to treat a disease when you can take multiple shots at a pathway, said Russ Altman of Stanford University. Polypharmacology is becoming increasingly common in medicine, he noted. "We treat HIV with three drugs, we treat tuberculo-

sis with four drugs, we treat depression with two to four. . . . The targetability of a pathway statistically is better than the targetability of a single element of the pathway."

CHALLENGES ASSOCIATED WITH COLLECTING PHENOTYPIC DATA

Gathering rich and accurate phenotypic data can be more complicated than gathering genetic data. For example, one workshop participant commented that while the concept of 23andMe is exciting, the population that the company is studying may not accurately represent the general population. And while disease registries tend to have relatively good phenotypic datasets, electronic health records generally do not contain research-grade phenotypic data, the participant noted, saying "phenotypic data has not yet reached a level that we've achieved for genotypic data."

While phenotype measurements can be subjective, some fields are beginning to rethink the boundaries of phenotypic descriptions so that they are more biologically meaningful, said Mark Daly of the Broad Institute of the Massachusetts Institute of Technology and Harvard University. Daly noted marked progress in the area of mental health: "There's now a push toward developing simple-to-apply [and] broad tools that can characterize mental state in a way that will be able to be consistently applied across populations around the world."

Phenotypic descriptions can be captured at various levels of detail, Altman noted, ranging from relatively simple measures such as the presence or absence of a condition, to more complex phenotypes such as the status of multiple biopsies. Informaticists are currently developing ways of using proxy variables, algorithms, and post-processing to extract valuable phenotypic data from the information that is available, he said, adding that there are ways to be very clever in extracting the data needed to answer a particular scientific question.

In many cases, callbacks to cohort participants for deeper phenotyping will be necessary, Ginsburg said. For example, there are many mutations known to exist in the population for which there is no knowledge about the phenotypes associated with them. One possible approach would be to use information gathered through unconventional means to inform genetic studies. For instance, Ginsburg suggested, perhaps it would be possible to use the 23andMe survey platform more ubiquitously throughout cohort studies. Another possibility would be to use mobile

health technology as a way to capture phenotypic information in a standardized way, he said.

CONSIDERING THE NEEDS OF PATIENTS

The broader view of biology that is developing could accelerate genomics-enabled drug discovery and should also incorporate the perspectives of patients, said one workshop participant. Genomics-enabled drug discovery is not moving as fast as it needs to move from the perspective of patients, the participant said, even though many patients are willing to share their genetic and medical information, which could result in the discovery of a wealth of potential drug targets.

A widespread cultural shift toward incorporating the needs and desires of patients during the design of research is something that could enhance outcomes, said Lynn Matrisian of the Pancreatic Cancer Action Network. Such a shift will create a sense of urgency and lead to the acknowledgment that one person cannot do this, that it is going to take working together as a team, she said. This approach is becoming more common, she added, but it is not yet widespread. New incentives, structures, and goals will be able to facilitate a cultural shift in this direction, she said.

On the subject of the involvement of patients, Lon Cardon of GSK expressed concern about the way that hope can sometimes turn into hype. It is difficult to predict what will happen in the next 10 years, he said, but overpromising and underdelivering remains a problem. However, hype tends to be cyclical in the field of genomics, he added, and the field is in a less hyperbolic phase now, because new drugs are just starting to emerge from research.

PROMOTING INFORMATION SHARING AND COLLABORATION

The potential for collaboration and data sharing to enhance drug discovery has been demonstrated in the past, Mark Daly said. Over the past 7 years, more than 100 genetic associations have been discovered for schizophrenia, he said, which happened because more than 50 independent groups around the world were willing to deposit their genetic data on a single server and engage in a collaborative analysis activity

(Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). It is likely that these advances would not have been achieved in such a short amount of time if the participating groups were not willing to work collaboratively, he said.

The opportunity for data-sharing success, as exhibited by collaborations in the mental health field, underscores the need for new collaborative models that would discourage investigators in academia and industry from keeping their data to themselves, Daly said. Genome-wide association studies provide an opportunity to share results in a way that addresses a more nuanced set of questions that go beyond simple genetic associations and attend to details about the tissues, cells, and developmental time periods in which genetic variants are active, Daly said.

Data Sharing as a Way to Reduce Research Duplication

The sharing of genetic data is a powerful opportunity because it engages an entire community of data analysts, each of whom might have been previously working in individual data silos, Daly said. Geneticists need to do a better job of reaching out to the broader molecular biology community with well-annotated genetic results in order to make these data more accessible to the entire research community, he continued. Today, a great deal of redundant research activity takes place as people do the same set of basic data processing and analysis activities. Currently, the discussion is too focused around raw data access, Daly said, arguing that there should instead be a focus on collaboratively completing the tedious elements so that the entire community can move on with the research.

Other speakers also pointed to the duplication that could be eliminated through expanded data sharing. For example, Chas Bountra of Oxford University observed that many companies work on the same drug target, and if a target fails, considerable time and resources can be lost. While replicating research results is essential, time and money could be saved by reducing the amount of unnecessary experimental duplication, particularly before a target is validated.

Factors That Could Contribute to Expanded Data Sharing

A clearer and more widely accepted definition of the boundaries of precompetitive research could enhance information sharing, as would standards to increase interoperability, Cardon said. “The competitive

part is making the drugs, but if we can share resources on unraveling the function of these targets, then maybe we can do a better job of developing drugs," he said. Even companies like 23andMe share data extensively. 23andMe has many academic collaborations, Scheller said, even though the data cannot leave the company because of its informed consent process.

Both the benefits and risks of information sharing should be openly communicated in order to optimize partnerships and drug discovery, said Rajesh Ranganathan, who at the time of the workshop was the vice president of science and regulatory advocacy at P*h*RMA. It is important to figure out what works for all parties involved, Ranganathan said. In several successful collaborations, the parties have determined what each can bring to the table and how they can potentially benefit. Leadership is also important in bringing disparate groups together in collaborations, Ranganathan said. As a result, collaborations will vary in their degree of success, depending on the type of leadership involved.

To increase data sharing, Altman suggested establishing a website, potentially named geneticcohorts.gov that would function like ClinicalTrials.gov, where all applicable clinical trials in the United States must be preregistered before beginning participant enrollment. Such a website could announce the existence of a genetics-based cohort study, its size, the types of metadata associated with it, and information on whom to contact to initiate a collaboration. The site would not contain the raw genetic data, but inclusion in the site could be a condition for publication, he said.

Trust between patients and researchers is essential in data sharing, Ginsburg said, which means that "our job is to educate." Most of the organizations conducting population studies are reaching out to communities and creating principles of engagement, but studies need to treat these individuals of the community as partners, not just as subjects, he said. "Shaking the hands of the participants and having them understand the transaction that's going to occur and how they are going to benefit from it is something that we have to take quite seriously."

Sharing appears to be common in some areas and organizations but not everywhere, Cardon said, adding, however, that the tide may be turning. From the business perspective, sectors have become increasingly entrepreneurial and collaborative, creating alternative ways to think about drug discovery and development. "We have an opportunity that is not just once in a generation but once in a lifetime," Cardon said. "The information sitting in front of us has the power and the potential to

transform the way we think about the next generation of medicines." The technologies and data exist or are imminent. But an acceleration of progress will require working together to translate research into medicines, which in turn will require innovative and not just incremental thinking. "It's really exciting, but it's incumbent upon us to make it happen," Cardon said.

CONCLUDING REMARKS

Three themes were touched on by individual speakers during the workshop: hope, incentivizing entrepreneurs, and patient-centricity. First, it is critical to broadly communicate the opportunities that genomics-driven drug discovery provides, Sarwar said. One example of an exciting genomics-enabled prospect is the ability to understand which subsets of patients are most likely to benefit from a particular medicine, he said, and this idea has changed how new medicines are developed. The realization that genetic support for drug targets increases the odds of success has underlined the importance of genomic research. This new perspective has contributed to substantial advances in the ability to treat and potentially cure diseases, and hope for continued advances is a strong motivator and enabler. Agreeing with Cardon, Sarwar said, "This may be a once-in-a-lifetime opportunity for us to truly change our odds of success for developing new medicines."

Second, scientific and commercial advances can also act to incentivize entrepreneurs. The challenge, Sarwar said, will be to put in place the structures to support entrepreneurship. Would better ways of integrating and applying data unlock new abilities to make medicines? Can the right combinations of expertise be brought together to begin realizing those opportunities? "Are there business opportunities that we've missed?" he asked.

Finally, patient-centricity needs to remain at the forefront of drug discovery and development, Sarwar observed. Patients are the primary drivers of research and drug development. What endpoints address patient needs? What are measures of progress versus measures of impact? What medicines do patients want, and how can these desires be met? "We don't make medicines because it's easy, or because it's a guaranteed return on investment," Sarwar said. "We make medicines because they are needed by patients. It's a privilege to be able to make new medicines, and it's a privilege to try to understand why diseases are caused."

Each sector involved in drug discovery and development is now in a position to help realize the potential behind that privilege.

References

Blom, D. J., R. Dent, R. C. Castro, and P. P. Toth. 2016. PCSK9 inhibition in the management of hyperlipidemia: Focus on evolocumab. *Vascular Health and Risk Management* 12:185–197.

Bunnage, M. E. 2011. Getting pharmaceutical R&D back on target. *Nature Chemical Biology* 7(6):335–339.

Clark, K., K. F. MacKenzie, K. Petkevicius, Y. Kristariyanto, J. Zhang, H. G. Choi, M. Peggie, L. Plater, P. G. Pedrioli, E. McIver, N. S. Gray, J. S. Arthur, and P. Cohen. 2012. Phosphorylation of CRTC3 by the salt-inducible kinases controls the interconversion of classically activated and regulatory macrophages. *Proceedings of the National Academy of Sciences of the United States of America* 109(42):16986–16991.

Collins, F. S., and H. Varmus. 2015. A new initiative on precision medicine. *New England Journal of Medicine* 372(9):793–795.

Dewey, F. E., V. Gusarova, C. O'Dushlaine, O. Gottesman, J. Trejos, C. Hunt, C. V. Van Hout, L. Habegger, D. Buckler, K. M. Lai, J. B. Leader, M. F. Murray, M. D. Ritchie, H. L. Kirchner, D. H. Ledbetter, J. Penn, A. Lopez, I. B. Borecki, J. D. Overton, J. G. Reid, D. J. Carey, A. J. Murphy, G. D. Yancopoulos, A. Baras, J. Gromada, and A. R. Shuldiner. 2016. Inactivating variants in ANGPTL4 and risk of coronary artery disease. *New England Journal of Medicine* 374(12):1123–1133.

DiMasi, J. A., H. G. Grabowski, and R. W. Hansen. 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics* 47:20–33.

Filippakopoulos, P., J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. La Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp, and J. E. Bradner. 2010. Selective inhibition of BET bromodomains. *Nature* 468(7327):1067–1073.

Flannick, J., et al. 2014. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nature Genetics* 46(4):357–363.

Helgadottir, A., S. Gretarsdottir, G. Thorleifsson, E. Hjartarson, A. Sigurdsson, A. Magnusdottir, A. Jonasdottir, H. Kristjansson, P. Sulem, A. Oddsson, G. Sveinbjornsson, V. Steinthorsdottir, T. Rafnar, G. Masson, I. Jonsdottir, I. Olafsson, G. I. Eyjolfsson, O. Sigurdardottir, M. S. Daneshpour, D. Khalili, F. Azizi, D. W. Swinkels, L. Kiemeney, A. A. Quyyumi, A. I. Levey, R. S. Patel, S. S. Hayek, I. J. Gudmundsdottir, G. Thorgeirsson, U. Thorsteinsdottir, D. F. Gudbjartsson, H. Holm, and K. Stefánsson. 2016. Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nature Genetics* 48:634–639.

Hodes, R. J., and N. Buckholtz. 2016. Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) knowledge portal aids Alzheimer's drug discovery through open data sharing. *Expert Opinion on Therapeutic Targets* 20(4):389–391.

HRSA (Health Resources and Services Administration). 2014. HRSA Health Center Program, Bureau of Primary Health Care. http://bphc.hrsa.gov/about/healthcenterfactsheet.pdf (accessed June 20, 2016).

Hu, Y., A. Shmygelska, D. Tran, N. Eriksson, J. Y. Tung, and D. A. Hinds. 2016. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nature Communications* 7:10448.

Jonsson, T., J. K. Atwal, S. Steinberg, J. Snaedal, P. V. Jonsson, S. Bjornsson, H. Stefansson, P. Sulem, D. Gudbjartsson, J. Maloney, K. Hoyte, A. Gustafson, Y. Liu, Y. Lu, T. Bhangale, R. R. Graham, J. Huttenlocher, G. Bjornsdottir, O. A. Andreassen, E. G. Jonsson, A. Palotie, T. W. Behrens, O. T. Magnusson, A. Kong, U. Thorsteinsdottir, R. J. Watts, and K. Stefánsson. 2012. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488(7409):96–99.

Jonsson, T., H. Stefansson, S. Steinberg, I. Jonsdottir, P. V. Jonsson, J. Snaedal, S. Bjornsson, J. Huttenlocher, A. I. Levey, J. J. Lah, D. Rujescu, H. Hampel, I. Giegling, O. A. Andreassen, K. Engedal, I. Ulstein, S. Djurovic, C. Ibrahim-Verbaas, A. Hofman, M. A. Ikram, C. M. van Duijn, U. Thorsteinsdottir, A. Kong, and K. Stefánsson. 2013. Variant of TREM2 associated with the risk of Alzheimer's disease. *New England Journal of Medicine* 368(2):107–116.

Kalow, W., and B. K. Tang. 1993. The use of caffeine for enzyme assays: A critical appraisal. *Clinical Pharmacology & Therapeutics* 53(5):503–514.

Karczewski, K. J., R. Daneshjou, and R. B. Altman. 2012. Chapter 7: Pharmacogenomics. *PLOS Computational Biology* 8(12):e1002817.

Liu, T., and R. B. Altman. 2011. Using multiple microenvironments to find similar ligand-binding sites: Application to kinase inhibitor binding. *PLOS Computational Biology* 7(12):e1002326.

Liu, J. Z., et al. 2013. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nature Genetics* 45(6):670–675.

McClellan, J., and M. C. King. 2010. Genetic Heterogeneity in Human Disease. *Cell* 141(2):210–217.

The Michael J. Fox Foundation. 2015. Parkinson's Biomarker Initiative expands genetic study. https://www.michaeljfox.org/foundation/publication-detail.html?id=578&category=7 (accessed June 6, 2016).

Nelson, M. R., H. Tipney, J. L. Painter, J. Shen, P. Nicoletti, Y. Shen, A. Floratos, P. C. Sham, M. J. Li, J. Wang, L. R. Cardon, J. C. Whittaker, and P. Sanseau. 2015. The support of human genetic evidence for approved drug indications. *Nature Genetics* 47(8):856–860.

NIH (National Institutes of Health). 2014. *Physician-scientist workforce (PSW) report 2014*. Rockville, MD: National Institutes of Health. https://report.nih.gov/workforce/psw/index.aspx (accessed July 3, 2016).

Nioi, P., et al. 2016. Variant ASGR1 associated with a reduced risk of coronary artery disease. *New England Journal of Medicine* 374(22):2131–2141.

NLM (National Library of Medicine). 2016. Pulmonary arterial hypertension. https://ghr.nlm.nih.gov/condition/pulmonary-arterial-hypertension#resources (accessed May 25, 2016).

Ono, M., T. Shimizugawa, M. Shimamura, K. Yoshida, C. Noji-Sakikawa, Y. Ando, R. Koishi, and H. Furukawa. 2003. Protein region important for regulation of lipid metabolism in angiopoietin-like 3 (ANGPTL3): ANGPTL3 is cleaved and activated in vivo. *Journal of Biological Chemistry* 278(43):41804–41809.

Pasche, B., and D. Absher. 2011. Whole-genome sequencing: A step closer to personalized medicine. *JAMA* 305(15):1596–1597.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511(7510):421–427.

Scott, R. A., et al. 2016. A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. *Science Translational Medicine* 8(341):341ra376.

Smith, C. 2003. Drug target validation: Hitting the target. *Nature* 422(6929):341, 343, 345.

Stacey, S. N., H. Helgason, S. A. Gudjonsson, G. Thorleifsson, F. Zink, A. Sigurdsson, B. Kehr, J. Gudmundsson, P. Sulem, B. Sigurgeirsson, K. R. Benediktsdottir, K. Thorisdottir, R. Ragnarsson, V. Fuentelsaz, C. Corredera, Y. Gilaberte, M. Grasa, D. Planelles, O. Sanmartin, P. Rudnai, E. Gurzau, K. Koppova, B. A. Nexo, A. Tjonneland, K. Overvad, J. G. Jonasson, L. Tryggvadottir, H. Johannsdottir, A. M. Kristinsdottir, H. Stefansson, G. Masson, O. T. Magnusson, B. V. Halldorsson, A. Kong, T. Rafnar, U. Thorsteinsdottir, U. Vogel, R. Kumar, E. Nagore, J. I. Mayordomo, D. F. Gudbjartsson, J. H. Olafsson, and K. Stefánsson. 2015. New basal cell carcinoma susceptibility loci. *Nature Communications* 6:6825.

Steinberg, S., H. Stefansson, T. Jonsson, H. Johannsdottir, A. Ingason, H. Helgason, P. Sulem, O. T. Magnusson, S. A. Gudjonsson, U. Unnsteinsdottir, A. Kong, S. Helisalmi, H. Soininen, J. J. Lah, DemGene, D. Aarsland, T. Fladby, I. D. Ulstein, S. Djurovic, S. B. Sando, L. R. White, G. P. Knudsen, L. T. Westlye, G. Selbaek, I. Giegling, H. Hampel, M. Hiltunen, A. I. Levey, O. A. Andreassen, D. Rujescu, P. V. Jonsson, S. Bjornsson, J. Snaedal, and K. Stefánsson. 2015. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nature Genetics* 47(5):445–447.

Steinthorsdottir, V., G. Thorleifsson, P. Sulem, H. Helgason, N. Grarup, A. Sigurdsson, H. T. Helgadottir, H. Johannsdottir, O. T. Magnusson, S. A. Gudjonsson, J. M. Justesen, M. N. Harder, M. E. Jorgensen, C. Christensen, I. Brandslund, A. Sandbaek, T. Lauritzen, H. Vestergaard, A. Linneberg, T. Jorgensen, T. Hansen, M. S. Daneshpour, M. S. Fallah, A. B. Hreidarsson, G. Sigurdsson, F. Azizi, R. Benediktsson, G. Masson, A. Helgason, A. Kong, D. F. Gudbjartsson, O. Pedersen, U. Thorsteinsdottir, and K. Stefánsson. 2014. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nature Genetics* 46(3):294–298.

Styrkarsdottir, U., G. Thorleifsson, S. A. Gudjonsson, A. Sigurdsson, J. R. Center, S. H. Lee, T. V. Nguyen, T. C. Kwok, J. S. Lee, S. C. Ho, J. Woo, P. C. Leung, B. J. Kim, T. Rafnar, L. A. Kiemeney, T. Ingvarsson, J. M. Koh, N. L. Tang, J. A. Eisman, C. Christiansen, G. Sigurdsson, U. Thorsteinsdottir, and K. Stefánsson. 2016. Sequence variants in the PTCH1 gene associate with spine bone mineral density and osteoporotic fractures. *Nature Communications* 7:10129.

Sukonina, V., A. Lookene, T. Olivecrona, and G. Olivecrona. 2006. Angiopoietin-like protein 4 converts lipoprotein lipase to inactive monomers and modulates lipase activity in adipose tissue. *Proceedings of the National Academy of Sciences of the United States of America* 103(46):17450–17455.

Tatonetti, N. P., J. C. Denny, S. N. Murphy, G. H. Fernald, G. Krishnan, V. Castro, P. Yue, P. S. Tsao, I. Kohane, D. M. Roden, and R. B. Altman. 2011. Detecting drug interactions from adverse-event reports: Interaction between paroxetine and pravastatin increases blood glucose levels. *Clinical Pharmacology and Therapeutics* 90(1):133–142.

Tufts Center for the Study of Drug Development. 2014. How the Tufts Center for the Study of Drug Development pegged the cost of a new drug at $2.6 billion. http://csdd.tufts.edu/files/uploads/cost_study_backgrounder.pdf (accessed July 3, 2016).

Tung, J. Y., C. B. Do, D. A. Hinds, A. K. Kiefer, J. M. Macpherson, A. B. Chowdry, U. Francke, B. T. Naughton, J. L. Mountain, A. Wojcicki, and N. Eriksson. 2011. Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS ONE* 6(8):e23473.

Wang, G., M. L. McCain, L. Yang, A. He, F. S. Pasqualini, A. Agarwal, H. Yuan, D. Jiang, D. Zhang, L. Zangi, J. Geva, A. E. Roberts, Q. Ma, J. Ding, J. Chen, D. Z. Wang, K. Li, J. Wang, R. J. Wanders, W. Kulik, F. M. Vaz, M. A. Laflamme, C. E. Murry, K. R. Chien, R. I. Kelley, G. M. Church, K. K. Parker, and W. T. Pu. 2014. Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nature Medicine* 20(6):616–623.

Whirl-Carrillo, M., E. M. McDonagh, J. M. Hebert, L. Gong, K. Sangkuhl, C. F. Thorn, R. B. Altman, and T. E. Klein. 2012. Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacololgy & Therapeutics* 92(4):414–417.

A

Workshop Agenda

Deriving Drug Discovery Value from Large-Scale Genetic Bioresources:
A Workshop
March 22, 2016

The Keck Center of the National Academies of Sciences, Engineering, and Medicine, Room 100
500 Fifth Street, NW
Washington, DC 20001

MEETING OBJECTIVES

- To address how progress can be made in discovering and validating promising targets and medicines for those targets by using data collected from large-scale genetic studies.
- To highlight current genomics-enabled drug discovery activities in industry, academia, and government and to share best practices for study design and data collection.
- To examine enabling partnerships and business models to facilitate the use of genetic data for drug discovery.

AGENDA

8:30–8:35 a.m. **Welcoming Remarks**

Geoffrey Ginsburg, *Co-Chair of the Roundtable on Genomics and Precision Health*
Director, Duke Center for Applied Genomics and Precision Medicine; Professor of Medicine and of Pathology and Biomedical Engineering, Duke University Medical Center

Russ B. Altman, *Co-Chair of the Forum on Drug Discovery, Development, and Translation*
Professor of Bioengineering, Genetics, Medicine, and (by courtesy) Computer Science, Stanford University

8:35–8:45 a.m. **Charge to Workshop Speakers and Participants**

Nadeem Sarwar, *Workshop Chair*
President, Andover Product Creation Innovation Systems, Eisai Inc.

SESSION I: DESIGNING COHORTS TO MAXIMIZE DISCOVERY CAPABILITIES

Objectives: To explore current gaps and opportunities in data collected in large cohort studies and what elements could enable a robust discovery toolbox.

Moderator: Geoff Ginsburg, *Co-Chair of the Roundtable on Genomics and Precision Health*; Director, Duke Center for Applied Genomics and Precision Medicine

8:45–9:30 a.m. Joe Vockley
Chief Operating Officer and Chief Scientific Officer, Senior Vice President, Inova Translational Medicine Institute

Mark Daly
Co-Director, Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard University

Aris Baras
Vice President, Co-Head, Regeneron Genetics Center, Regeneron Pharmaceuticals

9:30–9:45 a.m. **Break**

Kári Stefánsson
Chairman, Chief Executive Officer, and Founder, deCODE genetics

Richard Scheller
Chief Science Officer, 23andMe

10:15–10:45 a.m. **Discussion with Speakers and Attendees**

SESSION II: CURRENT GENOME-ENABLED DISCOVERY ACTIVITIES RELATED TO BIORESOURCES

Objectives: To discuss current activities among stakeholder groups and innovative technologies, to identify gaps, and to explore opportunities for cross-sector engagement.

Moderator: John Carulli, Director, Precision Medicine, Biogen

10:45 a.m.–12:00 p.m. Russ Altman
Professor of Bioengineering, Genetics, and Medicine, Stanford University

Lon Cardon
Senior Vice President of Alternative Discovery and Development, Head of Target Sciences, GlaxoSmithKline

Sohini Chowdhury
Senior Vice President, Research Partnerships, The Michael J. Fox Foundation

Sally John
Vice President, Computational Biology and Genomics, Biogen Idec

Tim Rolph
Vice-President, Program Value Enhancement, Pfizer Inc.

12:00–12:45 p.m. **Discussion with Speakers and Attendees**

12:45–1:45 p.m. **WORKING LUNCH**

SESSION III: BUSINESS MODELS THAT SUPPORT BIORESOURCE DISCOVERY COLLABORATION

Objectives: To examine potential precompetitive business models and investments that can support discovery efforts and opportunities across stakeholder groups.

Moderator: Rajesh Ranganathan, former Vice President of Science and Regulatory Advocacy, P*h*RMA

1:45–2:45 p.m.

Lynn Matrisian
Vice President, Scientific & Medical Affairs, Pancreatic Cancer Action Network

Meg Ehm
Director, External Strategic Alliances, Genetics, GlaxoSmithKline

David Wholley
Accelerating Medicines Partnership and Foundation for the National Institutes of Health; Director, Research Partnerships, Foundation for the National Institutes of Health

Chas Bountra
Professor of Translational Medicine, Head of Structural Genomics Consortium, University of Oxford

2:45–3:30 p.m. **Discussion with Speakers and Attendees**

3:30–3:45 p.m. **BREAK**

SESSION IV: WHERE DO WE GO FROM HERE?

Objectives: What are the short-term and long-term next steps for achieving effective collaboration among stakeholders? Are there incentives that should be explored? How can we maximize use of large-scale genetic bioresources to effectively translate research into drug discovery?

Moderator: Nadeem Sarwar, *Workshop Chair*; President, Andover Product Creation Innovation Systems, Eisai Inc.

3:45–3:55 p.m.	Dan Tagle Associate Director for Special Initiatives, Office of the Director, National Center for Advancing Translational Sciences, National Institutes of Health
3:55–4:25 p.m.	**Panel Discussion** Russ Altman Chas Bountra Lon Cardon Mark Daly Richard Scheller
4:25–5:10 p.m.	**Discussion with Speakers and Attendees**
5:10–5:25 p.m.	**SESSION HIGHLIGHTS** John Carulli Geoff Ginsburg Rajesh Ranganathan
5:25–5:30 p.m.	**CONCLUDING REMARKS** Nadeem Sarwar, *Workshop Chair* President, Andover Product Creation Innovation Systems, Eisai Inc.
5:30 p.m.	**ADJOURN**

B

Speaker Biographical Sketches[1]

Russ B. Altman, M.D., Ph.D., is the Kenneth Fong Professor and a professor of bioengineering, genetics, medicine, and (by courtesy) computer science, and past chairman of the Bioengineering Department at Stanford University. His primary research interests are in the application of computing technology to basic molecular biological problems of relevance to medicine. He is particularly interested in informatics methods for advancing pharmacogenomics, the study of how human genetic variation affects drug response. Other work focuses on the analysis of functional sites within macromolecules with an emphasis on understanding the action, interaction, and adverse events of drugs. Dr. Altman holds an M.D. from Stanford Medical School, a Ph.D. in medical information sciences from Stanford, and an A.B. from Harvard College. He has been the recipient of the U.S. Presidential Early Career Award for Scientists and Engineers and a National Science Foundation CAREER Award. He is a fellow of the American College of Physicians, the American College of Medical Informatics, and the American Institute of Medical and Biological Engineering. He is a past president, founding board member, and fellow of the International Society for Computational Biology. He is an organizer of the annual Pacific Symposium on Biocomputing. He leads one of seven National Institutes of Health–supported National Centers for Biomedical Computation, focusing on physics-based simulation of biological structures. He won the Stanford Medical School graduate teaching award in 2000. He is a member of the National Academy of Medicine. Dr. Altman is the chair of the Food and Drug Administration (FDA) science board, advising the FDA commissioner.

[1]This text was revised after the prepublication release.

He is the president-elect of the American Society for Clinical Pharmacology & Therapeutics. He is the founder of Personalis, Inc., which focuses on using next-generation sequencing for clinical diagnostics.

Aris Baras, M.D., M.B.A., serves as vice president and co-head of the Regeneron Genetics Center (RGC). Previously, he held various roles and responsibilities at Regeneron in research and development strategy and business development, technology development, and translational medicine, where he contributed to the development of new technologies and therapeutics. Prior to Regeneron, Dr. Baras worked with Liquidia Technologies, a nanotechnology company developing vaccines and respiratory therapeutics, where he contributed to the advancement of novel technologies for inhaled therapeutics and drug candidates in early stages of development. While at Duke University, Dr. Baras conducted research in immunology and oncology studying the mechanisms of action and resistance of B-cell depleting immunotherapies as well as ErbB2/HER2-targeting therapeutics in the setting of inflammatory breast cancer.

In his current role, Dr. Baras co-leads Regeneron's broad research programs in human genetics, with responsibility for overall scientific strategy, ensuring operational excellence, and establishing critical research collaborations, encompassing large-scale sequencing and genetic analysis, such as with the Geisinger Health System (GHS) and a network of 20 other academic, National Institutes of Health, and health system collaborators. Dr. Baras co-founded the RGC and established a collaboration with the GHS in 2013, with the goal of enrolling and sequencing as many as 250,000 participants from the MyCode Community Health Initiative at Geisinger, linking genomic data to health record data with the goal of guiding drug development activities and genomic medicine implementation. To date, more than 60,000 consented MyCode participants have been sequenced. Initial research efforts at the RGC have produced important new findings and gene discoveries. Examples include the identification, in collaboration with Columbia University Medical Center and Dr. Wendy Chung, of multiple families with disease-causing variants in a novel PAH gene, TBX4, and the discovery of marked reductions in the odds of coronary artery disease among carriers of known and novel inactivating mutations in ANGPTL4 in the first 40,000 sequenced MyCode participants at GHS. In establishing the RGC, Dr. Baras helped build a fully integrated genetics program spanning large-scale sequencing and operations, informatics and analytical expertise, and translational and functional genomics capabilities. Samples from nearly 100,000 par-

ticipants have been sequenced at the RGC across dozens of ongoing projects and collaborations, which he oversees. Dr. Baras received his bachelor's degree in biology and economics, his M.D., and his M.B.A., all from Duke University.

Chas Bountra, Ph.D., is a professor of translational medicine in the Nuffield Department of Clinical Medicine and an associate member of the Department of Pharmacology at the University of Oxford. He is also a visiting professor in neuroscience and mental health at Imperial College, London. Dr. Bountra is an invited expert on several government and charitable research funding bodies and an advisor for many academic, biotech, and pharma drug discovery programmes. Prior to coming back to Oxford, Dr. Bountra was the vice president and head of biology at GlaxoSmithKline. He was involved in the identification of more than 40 clinical candidates for many gastrointestinal, inflammatory, and neuro-psychiatric diseases. More than 20 of these molecules progressed into patient studies, and more than five of these delivered successful "proof of concept" data and hence progressed into late-stage development. He was involved in the launch and development of the first treatment for irritable bowel syndrome (alosetron) and was the first to show that neurokinin NK1 antagonists are anti-emetic in preclinical and clinical studies. His current interests are (1) using X-ray structures of novel human proteins to generate small molecule inhibitors, screening in human cells to identify novel targets for drug discovery, and then developing clinical candidates for evaluation in patients, pre-competitively; (2) focusing on epigenetic and genetically identified proteins, because these are likely to represent better targets for drug discovery, for many cancer, inflammatory, metabolic, and neuro-psychiatric diseases; (3) working with colleagues in Oxford to build major programs in rare diseases and in Alzheimer's disease and creating a "bioescalator" for the rapid translation of Structural Genomics Consortium (SGC) science; and (4) building stronger links with local hospitals, patient groups, regulatory agencies, private investors, contract research organizations, biotech companies, and large pharma companies to create a new, more efficient ecosystem for pioneer drug discovery. Dr. Bountra believes the SGC has become a leader in human protein structural biology and epigenetics chemical biology and that it is arguably one of the most successful open-innovation public–private partnerships in the world. Furthermore, with the many recent local developments (e.g., Target Discovery Institute, Kennedy Institute), he

believes Oxford is emerging as one of the major academic drug discovery centers in Europe. He has given more than 300 invited lectures. In 2012 he was voted one of the "top innovators in the industry."

Lon Cardon, Ph.D., joined GlaxoSmithKline (GSK) in 2008, initially as special vice president of genetics and shortly thereafter combining genetics with the departments of statistics, epidemiology, computational biology, and clinical pharmacokinetics to form the quantitative sciences division. Most recently he was the head of Alternative Discovery and Development, a pan-therapeutic division focused on novel disease areas and research paradigms in drug discovery and development. In 2015 he created and continues to lead GSK's Target Sciences, a new division aimed at harnessing the latest discoveries and technologies in genomics and other forms of "big data" to advance the next generation of drug targets. He is a member of the research and development executive team and chairs the governance body responsible for investment in all of GSK's discovery performance units. Prior to joining GSK, Dr. Cardon was a senior academic in the United Kingdom and the United States as a professor of bioinformatics at the University of Oxford until 2006 and then as a professor of biostatistics at the University of Washington and the Fred Hutchinson Cancer Research Center in Seattle. He undertook his Ph.D. training at the Institute for Behavioral Genetics at the University of Colorado and conducted his postdoctoral research in the Department of Mathematics at Stanford University. He has received a number of scientific awards, including election to the United Kingdom's Academy of Medical Sciences and the American Association for the Advancement of Science. He has authored more than 200 scientific publications and 15 books and chapters.

John Carulli, Ph.D., is the Director of Translational Genomics at Biogen. His research is focused on discovering genetic contributions to disease, validating genetically defined drug targets, and translating genomic discoveries into clinical applications. Dr. Carulli's research has spanned a number of therapeutic areas, including bone disease, autoimmune disease, and neurodegenerative disease. Human Genetics has revealed rare and common variants that can be directly targeted by drugs or that may identify pathways to target. Dr. Carulli's work in Mendelian genetics has identified rare variants in LRP5 that cause autosomal dominant high bone mass, directly identifying a target and pathway that are subjects of osteoporosis drug discovery. Together with academic and industrial

collaborators, his work has also identified common variants associated with rheumatoid arthritis and systemic lupus erythematosus that also highlight pathways that can be targeted. In addition to identifying drug targets, human genetics and genomics discoveries have clinical applications including genetic testing and biomarker development. Dr. Carulli's team at Biogen is working with collaborators to develop tests and treatments for spinal muscular atrophy and myotonic dystrophy. This work includes detailed transcriptomic analysis of spinal muscular atrophy (SMA) animal models and the impact of experimental drugs on the transcriptome. Additionally, the team has developed a rapid and inexpensive SMA genotyping assay suitable for newborn screening from dried blood spots. The goal of this work is to identify the right patients, to treat them at the right time, and to monitor the impact of treatment on their disease.

Sohini Chowdhury, M.A., joined The Michael J. Fox Foundation (MJFF) in 2005. In her current role at MJFF, Ms. Chowdhury oversees a team that focuses on three areas: (1) increasing engagement and developing partnerships with various stakeholders, including industry (pharma and biotech), insurance companies, and academic networks/medical groups; (2) developing and implementing strategies to improve recruitment for Parkinson's disease trials; and (3) managing pre-competitive partnerships, including the Parkinson's Progression Markers Initiative, an approximately $65 million clinical biomarker study. Prior to joining MJFF, Ms. Chowdhury worked at the World Economic Forum for 5 years. As the senior community manager of the forum's Technology Pioneers program, she was responsible for annually selecting and integrating innovative biotech, energy, and information technology companies into forum activities. Ms. Chowdhury also worked directly for the forum's chief executive officer, acting as his liaison with key forum stakeholders and overseeing several in-house projects. Ms. Chowdhury graduated with an M.A. from Georgetown University and holds a B.A. in international studies from Vassar College.

Mark Daly, Ph.D., is the founding chief of the Analytic and Translational Genetics Unit (ATGU) at Massachusetts General Hospital and an assistant professor in the Harvard Medical School. His research has historically focused on the development and application of statistical methods for the discovery and interpretation of genetic variation responsible for complex human disease, and with the recent creation of the ATGU, he and other core faculty are now focused on the interpretation of ge-

nome sequence and the use of genome information in clinical settings. Dr. Daly is also a senior associate member and a co-director of the Program in Medical and Population Genetics at the Broad Institute, where he leads many large-scale genome-sequencing studies in autism and inflammatory bowel disease. Dr. Daly's group has developed numerous methods and widely used software tools, including GENEHUNTER and HAPLOVIEW, genetic analysis tools used in thousands of laboratories worldwide, and GRAIL and DAPPLE, Web-based utilities for the interpretation of the results of genome-wide association studies (GWASs), and they have contributed to additional widely distributed tools developed in the Broad community such as PLINK and GATK. Dr. Daly's earlier work at the Whitehead Institute and Whitehead/Massachusetts Institute of Technology (MIT) Center for Genome Research (precursor to the Broad Institute) was instrumental in developing an understanding of patterns of variation in the human and mouse genomes and in the use of these patterns in disease gene mapping. While developing computational and statistical methods that can be broadly applied, his group has several primary medical genetics research foci. His lab serves as the analytic hub for the Psychiatric GWAS Consortium, an international consortium leading the largest collaborative GWAS studies in five major psychiatric disorders. He also has a long-standing effort in the mapping of genes for Crohn's disease and ulcerative colitis, where he helped found and lead the International Inflammatory Bowel Disease (IBD) Genetics Consortium, an international effort that has identified more than 150 genetic risk factors. More recently, his group has launched, with the support of the Helmsley Trust, a major exome sequencing effort to better elucidate functional alleles from the gene mapping to date, and in collaboration with Dr. Ramnik Xavier's group at the Center for the Study of Inflammatory Bowel Disease, the team pursues the functional interpretation and clinical ramifications of emerging findings from these continued gene discovery efforts. Dr. Daly also leads an extensive research program in neuropsychiatric genetics at ATGU—particularly in autism, schizophrenia, and attention deficit hyperactivity disorder (ADHD)—and has led large-scale GWAS and exome sequencing efforts in this area. The ATGU serves as the analytic hub for the Psychiatric Genomics Consortium, an international consortium leading the largest collaborative GWAS studies in five major psychiatric disorders. More recently, the group has facilitated numerous studies using exome sequencing to articulate the genetic origins of rare inherited diseases, early-onset and pediatric cancers, and severe adverse drug responses. Dr. Daly received his B.S. in physics

from MIT and his Ph.D. in human genetics from Leiden University, Netherlands. Dr. Daly was a recipient of the 2014 Curt Stern Award from the American Society of Human Genetics.

Margaret (Meg) G. Ehm, Ph.D., is a director of genetics at GlaxoSmithKline (GSK) in King of Prussia, Pennsylvania. She develops and manages external alliances that bring together GSK with academic and industry groups to build innovative capabilities capitalizing on genetic data that will drive the identification of high-quality drug targets. Her recent work has focused on the use of electronic health record data and genetic information to drive drug discovery and development. Prior to this role, she led the statistical genetics group at GSK through a series of progressively more challenging roles within the international pharmaceutical company. She received her B.S. degree from Vanderbilt University in mathematics and computer science and her M.A. and Ph.D. degrees from Rice University in statistics. She completed a brief postdoctoral post at North Carolina State University in 2001 where she remains an adjunct professor of statistics.

Geoffrey Ginsburg, M.D., Ph.D., is the founding director for the Center for Applied Genomics in the Duke University Medical Center and the founding executive director of the Center for Personalized and Precision Medicine in the Duke University Health System. He is a professor of Medicine, Pathology, and Biomedical Engineering at Duke University. He is an internationally recognized expert in genomics and personalized medicine with funding from the National Institutes of Health (NIH), the Department of Defense (DOD), Air Force, the Defense Advanced Research Projects Agency (DARPA), the Gates Foundation, and industry. Prior to Duke he was at Millennium Pharmaceuticals Inc. where he was vice president of Molecular and Personalized Medicine and responsible for developing pharmacogenomic and biomarker strategies for therapeutics. Dr. Ginsburg serves as an expert panel member for Genome Canada, as a member of the Board of External Experts for the National Heart, Lung, and Blood Institute (NHLBI), as co-chair of the National Academies of Sciences, Engineering, and Medicine's Roundtable on Genomics and Precision Health, as a member of the advisory council for the National Center for Advancing Translational Sciences, as co-chair of the Cures Acceleration Network, as an advisor to the Pharmacogenetics Research Network, and as a member of the World Economic Forum's Global Agenda Council on the Future of the Health Sector.

Sally John, Ph.D., is the vice president of genomics and computational biology at Biogen Idec, where her group focuses on applying human genetics, genomics and analytical methods to support drug discovery and development, from identification of new targets through to understanding genetic variability in drug response and patient stratification. She is a passionate advocate of the value of human genetics to support early drug discovery and has founded a number of groups within industry to strengthen the application of genetics, including a human genetics group at Pfizer and the statistical genetics group at AstraZeneca. Her most recent position prior to Biogen Idec was as the head of clinical genetics and bioinformatics at Pfizer. She gained a Ph.D. in Molecular Biology from the University of Manchester, United Kingdom, and held academic positions including senior lecturer in genetic epidemiology at the University of Manchester. Her academic focus of research has been in the area of inflammatory genetics, including rheumatoid arthritis, asthma and pain, and genetic epidemiology methods as applied to the analysis of complex traits. She is active in the external and pre-competitive community and has acted as the co-chair of the International Serious Adverse Events Consortium, an industry-led precompetitive consortium focusing on the genetic basis of drug-induced serious adverse events.

Lynn M. Matrisian, Ph.D., M.B.A., is the vice president of scientific and medical affairs at the Pancreatic Cancer Action Network, based in Manhattan Beach, California, and Washington, DC. She focuses on understanding the scientific and medical activities within the pancreatic cancer field and facilitating these activities through a grants program, a patient support program, and special research initiatives. Under her guidance, the grants program was expanded to include mechanisms focused on translational and clinical research activities, and the Know Your Tumor and patient-report outcomes registry initiatives were initiated. She continues to work with various stakeholders within the field to advance the organization's goal to double survival from pancreatic cancer by the year 2020. Dr. Matrisian is formerly a professor and the founding chair of the Department of Cancer Biology at Vanderbilt University. She received her Ph.D. in molecular biology from the University of Arizona and M.B.A. from Vanderbilt University. She is a past president of the American Association of Cancer Research (AACR), a fellow of the AACR Academy, and the recipient of the Paget-Ewing award from the Metastasis Research Society. She served as co-chair of the National Cancer Institute's (NCI's) Translational Research Working Group and as a

special assistant to the director of the NCI. Research in her laboratory revolved around the molecular mechanisms underlying tumor progression and metastasis, with emphasis on the biology of matrix-degrading proteinases.

Tim Rolph, D.Phil., is the vice president of program value enhancement at Pfizer Inc. Formerly the chief scientific officer of Pfizer's Cardiovascular and Metabolic Disease Research Unit, he established it in Cambridge, Massachusetts. During his leadership, ertuglifozin was discovered and progressed to Phase 3 in partnership with Merck as a fixed-dose combination with Januvia™. He received a B.Sc. in biochemistry from the University of London (UK), and a D.Phil. from University of Oxford (UK). His pre- and postdoctoral training was at the Nuffield Institute for Medical Research, studying metabolic adaptations of skeletal and cardiac muscle during development. Subsequently, he joined Glaxo's veterinary research and development, initially studying modulation of growth for food production, then carrying out research for antiparasitic vaccines against protozoan (anti-coccidial for poultry, Paracox™) and metazoan species (gastrointestinal helminths). Beginning in a similar role at Pfizer, he became the leader of human anti-infective research at Sandwich (UK), during which the prototypical CCR5 antagonist maraviroc (Celzentry™) was discovered and launched as a novel antiretroviral for HIV. He then became head of research at Pfizer's Sandwich laboratory and more recently at Groton, Connecticut. Through his career, he has led groups that have taken many different therapeutic mechanisms into Phase 2, covering HIV, diabetes, and inflammatory and renal diseases.

Nadeem Sarwar, M.Pharm., M.R.Pharm.S., M.Phil., Ph.D., has expertise in human genetics-guided drug discovery and precision medicine. He has executive-level experience in academia (Tenured Faculty, University of Cambridge, United Kingdom), big-pharma (Senior Director, Pfizer Inc.), mid-size pharma (Vice President, Genetics & Human Biology, Eisai Inc.) and biotech-like organizations (President, Eisai Andover innovative Medicines [AiM] Institute). His research interests stem from the intersection of innovation in human genetics and collaborative business models to accelerate delivery of novel, targeted therapeutics. He has experience of several therapeutic areas including dementia, immuno-oncology, auto-immunity and cardiometabolic diseases. His research has been published in leading medical journals (e.g., *New England Journal*

of Medicine, *Lancet*, *JAMA*), presented at international meetings (e.g., American Diabetes Association, Prix Galien Foundation, Hitachi Innovation), and covered by international media (e.g., BBC, Bloomberg, Forbes). He has been invited to provide expert insights on human genetics and drug discovery for The World Dementia Envoy, Genomics England, The UK Minster for Life Sciences, and Scottish Enterprise.

Dr. Sarwar is the founder and president of the newly launched Eisai AiM Institute, an industry unique discovery innovation unit within the greater Boston biopharma hub of 90 integrated scientists in Andover, Massachusetts. The AiM Institute's exclusive mission is to realize human genetics driven drug discovery, with a predominant focus on delivering precision medicines for the subset of patients with immune-driven pathology in dementia (immunodementia) and oncology (immunooncology). The AiM Institute currently has assets in early stage clinical trials, late stage pre-investigational new drug (IND) and early stage preclinical discovery. To realize the vision of human genetics driven drug discovery, Dr. Sarwar has established a number of entrepreneurial collaborations and scientific partnerships (including cross-sector, open innovation and precompetitive) with a range of collaborators. Dr. Sarwar joined Eisai in 2013 as vice president and global head of Genetics & Human Biology, in which capacity he established and served as director of the Integrated Human Genomics (IHGx) Research Unit. In October 2015 he was named to his current position, and launched the Eisai AiM Institute in June 2016.

Before joining Eisai, Dr. Sarwar served as senior director, head of Population Research and head of Cardiometabolic Genetics at Pfizer Inc., where he was also a member of the Atorvastatin Core Advisory Board. Key projects under his leadership at Pfizer included leading the development and initiation of a new Cardiovascular Therapeutic Area Strategy; establishing and serving as co-director of the University of Cambridge/Pfizer Center for CV Genomics; creating a "genes to targets" strategy delivering novel targets into the exploratory cardiometabolic research portfolio; and serving on the clinical design team for a 20,000 participant phase 3 outcome trial.

Prior to working in the pharmaceutical industry, Dr. Sarwar was a tenured faculty member at the School of Clinical Medicine, University of Cambridge, where he led large-scale research consortia and new bioresources to identify causal and predictive disease risk factors. He served as Principal Investigator of the IL6R Genetics Consortium and the Triglycerides & Coronary Disease Genetics Consortium; and Steering

Committee & Leadership Team member of the Emerging Risk Factors Collaboration, the Pakistan Risk of Myocardial Infarction Study, Generation Scotland, and the Reykjavik Study Predictive Biomarkers Project.

Dr. Sarwar served as chair of The World CNS Summit; The Drug Discovery Stream of the Festival of Genomics; and The External Advisory Board of the National Institute on Aging (NIA) Longevity Genomics Study. In addition, he serves as industry lead for the Alzheimer's Disease Neuroimaging Initiative (ADNI) Genetics Core; sits on the National Academies of Sciences, Engineering, and Medicine's Roundtable on Genomics and Precision Health; serves on the Genetics Scientific Advisory Group of the European Prevention of Alzheimer's Disease Consortium; and facilitates the Industry Partnership for Human Genetics.

Richard H. Scheller, Ph.D., joined 23andMe in January of 2015 as its chief scientific officer and head of therapeutic development. He is responsible for translating genetic information into new therapies. Hired in 2001 as the senior vice president of research and member of the Genentech executive committee, from 2009 through 2014 he was the executive vice president and head of Genentech research and early development and a member of the Roche executive committee. Dr. Scheller received his bachelor of science in biochemistry in 1975 from the University of Wisconsin–Madison and his doctorate in chemistry in 1980 from Caltech. After postdoctoral fellowships at Caltech and Columbia, Dr. Scheller joined the Stanford faculty from 1982 until 2001. An investigator with the Howard Hughes Medical Institute from 1994 to 2001, Scheller has been an adjunct professor at the University of California, San Francisco, since 2004. In 2014 he was named a trustee of Caltech.

Dr. Scheller's research elucidating molecular mechanisms governing neurotransmitter release earned him the 2013 Albert Lasker Basic Medical Research Award, the 2010 Kavli Prize in Neuroscience, and the 1997 U.S. National Academy of Sciences Award in Molecular Biology. He is a fellow of the American Academy of Arts and Sciences, and a member of the National Academy of Sciences and the National Academy of Medicine.

Kári Stefánsson, M.D., Dr.Med., has served as the president, the chief executive officer and a director of deCODE genetics since he founded the company in August 1996. Dr. Stefánsson was appointed the chairman of the board of directors of deCODE genetics in December 1999. From

1993 until April 1997, Dr. Stefánsson was a professor of neurology, neuropathology, and neuroscience at Harvard University. From 1983 to 1993, he held faculty positions in neurology, neuropathology, and neurosciences at the University of Chicago. Dr. Stefánsson received his M.D. and Dr.Med. from the University of Iceland and is board-certified in neurology and neuropathology in the United States. He has published numerous articles on the genetics of common/complex diseases and has been among the leaders of the world in the discovery of variants in the sequence of the human genome that are associated with the risk of common/complex traits. Dr. Stefánsson was chosen by *Time* magazine as 1 of the 100 most influential men of the year for 2007 and by *Newsweek* as 1 of the 10 most important biologists of the 21st century. He was the recipient of the Jakobus Award in 2007, the World Glaucoma Association Award for present scientific impact in 2007, the European Society of Human Genetics Award in 2009, and the Andre Jahre Award in 2009.

Danilo A. Tagle, Ph.D., is the associate director for special initiatives at the National Center for Advancing Translational Sciences (NCATS). He leads and provides scientific and programmatic oversight and coordination to the following trans-National Institutes of Health (NIH) programs: (1) NIH Microphysiological Systems (a.k.a. tissue chip) program, (2) Extracellular RNA Communication program, and (3) SPARC (Stimulating Peripheral Activity to Relieve Conditions) program. These activities involve coordination with other NIH institutes and centers as well as partnerships with other government agencies, such as the Food and Drug Administration, Defense Advanced Research Projects Agency, and Defense Threat Reduction Agency, and the private sector. Prior to joining NCATS, Dr. Tagle was a program director for neurogenetics at the National Institute of Neurological Disorders and Stroke (NINDS), where he was involved in developing programs in genomics-based approaches for basic and translational research in inherited brain disorders. Dr. Tagle obtained his Ph.D. in molecular biology and genetics from Wayne State University School of Medicine in 1990. He was an NIH National Research Service Awards postdoctoral fellow in human genetics at the laboratory of Dr. Francis S. Collins at the University of Michigan. Prior to joining NINDS in 2001, Dr. Tagle was an investigator and the section head of molecular neurogenetics at the National Human Genome Research Institute beginning in 1993, and he has been involved in the highly collaborative effort towards the positional cloning of genes for Huntington's disease, ataxia-telangiectasia, and Niemann-Pick type C

disease. In addition to being the associate director for special initiatives, Dr. Tagle recently served as the acting director for the NCATS Office of Grants Management and Scientific Review and also as the executive secretary to the NCATS Advisory Council and the Cures Acceleration Network Review Board. He has served in numerous committees and advisory boards, and was on the editorial board of the journal *Gene* as well as the *International Journal of Biotechnology*. He has more than 150 scientific publications and has garnered numerous awards and patents. Central to Dr. Tagle's accomplishments and goals is leveraging key resources and expertise through partnerships with various stakeholders in biomedical research, including various government agencies, nonprofit organizations, patient advocacy groups, industry, and pharmaceutical corporations.

Joe Vockley, Ph.D., is the chief operating officer and chief scientific officer of the Inova Translational Medicine Institute. Dr. Vockley brings 25 years of experience in academic, pharmaceutical, and biotechnology contract research organizations and government research. He has broad and deep expertise in the fields of genetics, genomics, molecular diagnostics, bioinformatics, and large program management.

Dr. Vockley is a results-oriented manager and scientist. He is an inventor on numerous U.S. and international genomic and bioinformatic technology patents in the areas of DNA diagnostics, laboratory methods for microarray analysis, gene discoveries, and bioinformatic tool development. His basic research interests are in the fields of cancer and inborn errors of metabolism.

Dr. Vockley has previously held positions of chief scientific officer, vice president of research, director of genomics, and director of bioinformatics. Most recently, he was the director of the National Cancer Institute's Cancer Genome Atlas Project and the Cancer Genome Atlas Program Office.

David Wholley, M.Phil., manages the Research Partnerships Division of the Foundation for the National Institutes of Health, which is responsible for major research collaborations, including the Accelerating Medicines Partnership, the Biomarkers Consortium, the LungMAP precision medicine trial in lung cancer, and the Alzheimer's Disease Neuroimaging Initiative. Mr. Wholley has also served as the director of the Genetic Association Information Network, a public–private partnership dedicated to helping discover the genetic basis of common disease, and he led the

development of a major public–private partnership in drug safety with the biopharmaceutical industry and the Food and Drug Administration. Prior to joining the foundation in 2006, Mr. Wholley's career spanned nearly 25 years in health care technology business management, including extensive experience in product development, sales, marketing, corporate strategy, and partnership and project development. Mr. Wholley has held senior management roles in several venture-funded technology startup companies, including as head of global marketing and development for First Genetic Trust, Inc., which developed software for large-scale collaborative genetic research and personalized medicine. During a 16-year career at IBM, he co-led the corporate strategy team that guided IBM's formation of its life sciences industry organization. Mr. Wholley holds an M.Phil. from Rutgers University and a certificate in business administration from the Stern School of Business at New York University.

C

Statement of Task

An ad hoc committee will plan and conduct a 1-day public workshop to examine and discuss how large-scale genetic data could be used to improve the likelihood of bringing effective and targeted therapies to patients. The goal of the workshop will be to address how progress could be made in discovering and validating promising targets and medicines for those targets by using the data collected from large-scale genetic studies. Discussions will be held with a broad array of stakeholders, which may include representatives from pharmaceutical and biotech companies, information technology and data science companies, research institutes, investors, providers, patients, payers, and regulators. The planning committee will develop the workshop agenda, select and invite speakers and discussants, and moderate the discussions. A summary of the workshop will be prepared by a designated rapporteur in accordance with institutional policy and procedures.

D

Registered Attendees

Olu Adeniyi
Food and Drug Administration

Russ Altman
Stanford University

Margaret Anderson
FasterCures

John Aquino
Bloomberg BNA

Hugh Auchincloss
National Institute of Allergy and Infectious Diseases

Aris Baras
Regeneron Pharmaceuticals

John Baras
University of Maryland

Miriam Bayes
Thomson Reuters

Robert Beckman
Georgetown University Medical Center

Adam Berger
Department of Health and Human Services

Gouri Shankar Bhattacharyya
Fortis Hospital

Rebecca Blanchard
Merck & Co., Inc.

Bruce Blumberg
Kaiser Permanente

Kimberly Boucher
Inova Health System

Chas Bountra
University of Oxford

Khaled Bouri
U.S. Food and Drug Administration

Linda Brady
National Institute of Mental Health

Joel Brill
Predictive Health LLC

PJ Brooks
National Center for Advancing Translational Sciences

Apryl Brown
American Public Health Association Genomics Forum

Tara Burke
Association for Molecular Pathology

Colleen Campbell
University of Iowa

Robert Campbell
Brown University

Lon Cardon
GlaxoSmithKline

John Carulli
Biogen

Ann Cashion
National Institute of Nursing Research

Sohini Chowdhury
The Michael J. Fox Foundation for Parkinson's Research

Charles Cywin
National Institute of Neurological Disorders and Stroke

Andrew Dahlem
Eli Lilly and Company

Mark Daly
Massachusetts General Hospital

Susan Delaney
Coriell Institute for Medical Research

Joe Donahue
GeneDx

Michael Dougherty
American Society of Human Genetics

Erin Durkin
Inside Health Policy

Meg Ehm
GlaxoSmithKline

Raith Erickson
Independent

Lynn Etheredge
Rapid Learning Project

Greg Feero
Journal of the American Medical Association

Caroline Fox
Merck Research Laboratories

Steven Galson
Amgen Inc.

Michael Garvin
AstraZeneca

Taylor Gilliland
National Center for Advancing Translational Sciences

Geoffrey Ginsburg
Duke University

Tina Grande
Healthcare Leadership Council

Christian Grimstein
U.S. Food and Drug Administration

Jill Hagenkord
23andMe

Jennifer Hall
University of Minnesota

Erin Hauenstein
Northrop Grumman

Heonia Hillock
Inova Health System

Carolyn Hoban
Hartford Hospital

Arthur Holden
International Serious Adverse Event Consortium

William Hoos
Pancreatic Cancer Action Network

Lynn Hudson
Critical Path Institute

Sally John
Biogen

Brett Johnson
StoneFace Ventures

Robert Karp
National Institute of Diabetes and Digestive and Kidney Diseases

Marina Kozak
Friends of Cancer Research

Audrey Kusiak
Department of Veterans Affairs

Joan Lakoski
American Association of Colleges of Pharmacy

Katherine Lambertson
Genetic Alliance

David Lanfear
Henry Ford Hospital

Debra Leonard
University of Vermont Medical Center

Sharon Liang
Food and Drug Administration

Klaus Lindpaintner
Pfizer Inc.

Bolan Linghu
Pfizer Inc.

Katrina Loomis
Pfizer Inc.

Lynn Lund
Cogstate

Lynn Matrisian
Pancreatic Cancer Action Network

Robert McCormack
Janssen Oncology

Shelly Menolascino
Washington Square Psychiatry and Transcranial Magnetic Stimulation

Melissa Miller
Pfizer Inc.

Ron Miller
IMS Health

Adele Mitchell
Merck Research Labs

Cliona Molony
Pfizer Inc.

Bernard Munos
InnoThink Center for Research in Biomedical Innovation

Laura Nisenbaum
Eli Lilly and Company

James O'Leary
Genetic Alliance

Steve Olson
Self-Employed

John Orloff
Baxalta

Mike Pacanowski
Food and Drug Administration

Heather Pierce
Association of American Medical Colleges

Liz Powell
Government to Growth Consulting

Vicky Pratt
Association for Molecular Pathology

Rajesh Ranganathan
Pharmaceutical Research and Manufacturers of America

Nagarajan Rangarajan
National Institute of Neurological Disorders and Stroke

Robert Ratner
American Diabetes Association

Turna Ray
GenomeWeb

Samantha Roberts
Friends of Cancer Research

Tim Rolph
Pfizer Inc.

Mary Rubino
Rubino & McGuire Associates

Nadeem Sarwar
Eisai Inc.

Richard Scheller
23andMe

Kathryn Schubert
Society for Maternal–Fetal Medicine

Marion Schwartz
The Cholangiocarcinoma Foundation

Joan Scott
Health Resources and Services Administration

Kelly Servick
Science Magazine

Sam Shekar
Northrop Grumman

Chun-Pyn Shen
EMD Serono

Lana Skirboll
Sanofi

Fabrice Smieliauskas
University of Chicago

Brad Smith
FasterCures

Gyan Srivastava
Merck Research Laboratories

Kári Stefánsson
deCODE genetics

Mark Stewart
Friends of Cancer Research

Katie Johansen Taber
American Medical Association

Danilo Tagle
National Center for Advancing Translational Sciences

Sharon Terry
Genetic Alliance

Douglas Throckmorton
U.S. Food and Drug Administration

Patty Vasalos
College of American Pathologists

Hetal Vig
Rutgers Cancer Institute of New Jersey

Joseph Vockley
Inova Translational Medicine Institute

Carrie Wager
Pfizer Inc.

John Wagner
Takeda

Wes Walker
Cerner

Michael Watson
American College of Medical Genetics and Genomics

Susan L. Weiner
Children's Cause for Cancer Advocacy

David Wholley
Foundation for the National Institutes of Health

Catherine Wicklund
National Society of Genetic Counselors

David Wierz
OCI Group

Bob Wildin
National Human Genome Research Institute

Jane Wilkinson
Broad Institute

Janet Williams
American Academy of Nursing; University of Iowa College of Nursing

Emre Yucel
University of Texas, Houston School of Public Health