Unique Scientific Opportunities for the Precision Medicine Initiative

A Workshop of the Precision Medicine Initiative Working Group
of the Advisory Committee to the NIH Director
April 28-29, 2015
Natcher Conference Center, Bethesda, MD

EXECUTIVE SUMMARY

Background

On April 28-29, the National Institutes of Health (NIH) hosted a public workshop on the NIH campus in Bethesda, Maryland, to consider visionary scientific questions that could be addressed by the national research cohort proposed under the President's Precision Medicine Initiative (PMI). The workshop is part of a series of events and outreach that will inform a report on the design of the national research cohort that will be drafted.

The meeting featured speakers with a variety of expertise, including genomics, metabolomics, environmental and behavioral health, and informatics. Registration was open to the public and the workshop was also videocast. In addition, the workshop started a lively conversation on twitter, and some tweets from members of the public were shared for discussion by panelists and members of the working group.

Following the workshop, the ACD PMI working Group convened in a closed session to discuss the information shared during the workshop.

Workshop Summary

Sharon Terry (Genetic Alliance) kicked off the workshop by posing the question "Why PMI?" She reminded the audience that the time is right for the Precision Medicine Initiative with the advent of technologies like whole genome sequencing, and new tools for big data sets and biomedical analyses. But PMI is also an urgency: precision medicine may hold the keys to understanding causalities and discovering effective ways to prevent and treat conditions for that reduce the quality and length of the lives of ourselves and our loved ones. She challenged the audience to keep sight of the Initiative's long-term goals – to provide better understanding of health and disease, and better treatment and preventions strategies – and to break down boundaries, forge broad alliances, work selflessly for what really matters, and to give every person the opportunity to be a contributor to the success of PMI.

Following this introduction, the first three sets of panelists were asked to describe cutting-edge technologies that can inform, and be informed by, a national cohort.

Genomic and other 'Omic' technologies: The first session focused on the opportunities and need for the inclusion of genomics and other 'omics' technologies in the cohort. Euan Ashley (Stanford University School of Medicine) discussed the importance of large cohorts to the identification, reclassification, and interpretation of both common and rare variants, as well as the strength of familial analyses in genomic cohort research. Vamsi Mootha (Harvard Medical School) discussed the importance of the incorporating metabolic profiling into cohort data collection, noting that this data integrates the effects of both genes and environment. Vikram Bajaj (Google[x]) described the importance of combining multidimensional

data collection with longitudinal observation in order to maximize the value of cohort research to precision medicine outcomes.

Key Discussion Points:

- The power of the study to produce reliable results will depend on the reproducibility and stability of data derived from the samples, which depends on both the quality of data standards and the ability to accurately combine data across different EHRs.
- Designing the cohort to allow for identification of families and sibling groups could open up greater analytic possibilities for "omics." However, the inclusion of children (or children's samples) in the cohort has special considerations.
- Opinions varied on the most important "omic" data collection elements. Beyond genomics, including whole genome or whole exome sequencing, "omic" data elements important for consideration include the microbiome, metabolome, proteome, methylome, transcriptome, immune repertoire and cell-free DNA.
- The capture of these types of data sets will provide new opportunities for researchers to learn how to link them on a large scale, creating broad "omics" analysis.

Environmental and behavioral factors on health: The second session highlighted environmental and behavioral factors that could be collected and studied in the cohort to advance precision medicine. Marie Lynn Miranda (University of Michigan) stressed the importance of collecting geographical and time/date stamps for the cohort data , noting that Geographic Health Information Systems can link health system and social and environmental data to better understand health (and health disparities) at the individual and community level. Both Stephen Intille (Northeastern University) and Saul Shiffman (University of Pittsburgh) highlighted the potential for mobile applications to advance PMI research, which would allow more participants to be studied over time, increase the collection of self-reported data, and allow for the possibility of technology-based behavioral interventions.

Key Discussion Points:

- "Triangulating" different data types and research models will allow researchers to better dissociate correlative from causative factors for diseases.
- Partnering with the tech industry and requiring open source data will allow for sophisticated PMI technologies.
- Priorities for data collection include measuring chemical exposures, physical activities, and putting a geographic and time/date stamp on the data collected.

"Big Data": The third session focused on the potential uses and outcomes of "big data" that could be collected in the cohort. Russ Altman (Stanford University) outlined the potential for machine learning systems to identify diseases or disease relevant features without prior knowledge of the domain of interest, which may be valuable to identifying new factors that are relevant to human traits. Atul Butte (University of California, San Francisco) discussed the possibility of leveraging existing cohorts and currently-available storage systems (e.g., cloud computing) to maximize the potential of cohort data. Andrey Rzhetsky (University of Chicago) spoke to the importance of identifying both genetic and environmental factors, which will require developing deep clinical histories over time, and to consider more than one disease at a time in cohort research, as most databases are composed of large collections of heterogeneous phenotypes. During the discussion period, panelists raised the merits and challenges of distributed and centralized data models. While a distributed model may be socially, politically, and economically advantageous, a centralized model may prevent the loss of access to data sets should their sites lose funding. The model built (distributed, centralized, or a hybrid) will need to

carefully identify and guard against such challenges. Centralized and distributed data models may also pose challenges in balancing participants' information protections against the ability of qualified researchers to have access to the information needed to conduct meaningful research. A well-balanced approach will provide security for participants and freedom for researchers, while recognizing that there will be a range of tolerance and desire for health data privacy among individual participants. Some individuals and populations (such as rare disease populations) may favor wide sharing of their health data, especially if their "data altruism" is recognized and encouraged through return of information and appreciation.

Key Discussion Points:

- There are many challenges presented by using EHRs as a major source of cohort data.
- The cohort design must balance near-term and long-term outcomes in order to keep the public invested in the PMI.
- There are merits and challenges to both the federated and centralized model of cohort platform development.

On the second day of the workshop, panelists were asked to discuss both near-term (3-5 years from now) and longer-term (10-20 years from now) research studies that would be uniquely suited to be conducted in the national research cohort.

Near-Term Cases: The fourth session highlighted potential uses for the cohorts within the next 3-5 years, focusing on studies that are on the cusp of feasibility and would be catalyzed by the cohort. Moderator Eric Dishman noted that we are on the cusp of exascale computing: during the near-term phase of the cohort, which can yield a quadrillion operations per second. Precision medicine may offer much-needed use cases for this magnitude of computing power. Sekar Kathiresan (Massachusetts General Hospital) provided two short-term examples of research where the cohort could provide valuable clinical information: the ability to distinguish correlative from causative factors in heart attacks and the ability for researchers to identify and study the mechanisms of naturally occurring genetic mutations that protect against the development of different diseases. Josh Stein (Adhere Tech) discussed potential benefits derived from using technologies, such as "smartpill" bottle, to both study and provide interventions for health. Howard Hu (University of Toronto) focused on the benefits of studying environmental exposures, stressed the need to collect environmental data at the outset of the cohort study, and highlighted the potential for global collaborations to advance PMI research. Anne Wojcicki (23andMe) discussed strategies to increase participant engagement in cohort research, which includes returning individual results and giving participants access to research results as a form of compensation (e.g., through publication in open-access journals).

Key Discussion Points:

- Participants may be more willing to engage in long-term research when they receive individual results and information about how and when their data are being used.
- The public worries about who is holding and using their data, so transparency will be key.
- Recruitment strategies should include both disease and healthy populations.

Long-Term Cases: The final session focused on examples of precision medicine insights that might emerge over 10-20 years. Isaac Kohane (Harvard Medical School) made a number of predictions regarding future uses of precision medicine, including: the creation of family biohistory banks, which utilize family histories, DNA, and analysis of disease-related cell types to predict disease risk; the creation of precision insurers, who would adjust rates according to the amount of information they have

about an individual; and the involvement of laypeople in the identification of disease risk using big data technologies. Jennifer Malin (Anthem) discussed long-term PMI outcomes that included tailoring cancer therapies, providing individual side-effect profiles of drugs, and preventative health care check-ups that include specific recommendations developed from interpreting an individual's genetic risk profile. Brenda Eskenazi (University of California, Berkley) discussed the importance of optimizing health across the lifespan by focusing on prenatal and multigenerational cohort data collection, including environmental exposures.

Key Discussion Points:

- Specific strategies (such as outreach to mothers) will need to be needed to successfully recruit families into the cohort.
- Including insurance companies as partners in cohort efforts may be highly beneficial, although the extent to which insurance companies will participate in widespread data sharing remains unknown.
- Strategies to successfully recruit underserved populations will also need to be considered during cohort design. Training members of the community to recruit and sustain relationships with participants has proven to be one successful strategy for engaging participants from underserved populations.

Recurring Points Made at the Workshop

Integrate Environmental Exposure and Genomic Data: The key to understanding common diseases will likely be through a better understanding of gene by environment interactions, which will require collecting a wide variety of environmental data, including environmental samples, geographical information and sociobehavioral data.

Collect Data Across Time: Workshop participants emphasized the power of longitudinal studies for the success of PMI, and recommended that the cohort collect data across as wide a timeframe as possible by including developmental data and multigenerational data. Several workshop participants emphasized the importance of recruiting families in the cohort.

Incorporate mHealth Technologies into Cohort: The incorporation of mobile heath technologies in the cohort may provide powerful new tools with which to study, and possibly intervene, in human health.

Integrate Data to the Greatest Extent Possible: The success of PMI will rely, in large part, on our ability to merge and analyze the range of biological and environmental data that may influence an individual's health. Current technologies are rapidly growing in their capacity to compute large amounts of data, although they might not yet be able to handle the volume of biomedical research being proposed. However, integration of real-time data into research will help to accelerate the research process and the production of results that are relevant to precision medicine.

Engage Participants: Engaging participants at the individual, family, and community levels are important ways to encourage cohort participation and the uptake of research results at each of these levels. Social media has advanced such that it might be a particularly fruitful way to engage individual participants.

Maximize Cohort Diversity: Maximizing cohort diversity will allow researchers to more easily identify associations between biological, environmental and sociobehavioral health factors and the development of disease. Cohort diversity will need to be achieved at the population (many different types of participants) and individual (many different types of data) level.

Train a New Workforce in Precision Medicine: There is a need for training to develop a workforce that is capable of analyzing, interpreting and applying research results to the field of precision medicine.