Genome **Medicine**

MEETING REPORT

Systems and genome-wide approaches unite to provide a route to personalized medicine

Barbara E Stranger[1,2,3]*, Johan Björkegren[4,5], M Eileen Dolan[6] and Marylyn D Ritchie[7]

Abstract

A report on the Keystone Symposium 'Complex Traits: Genomics and Computational approaches', Breckenridge, Colorado, USA, 20-25 February 2012.

Translating biological findings into healthcare

The study of complex traits in humans and model organisms has made considerable progress in recent years. Technological innovation facilitated an era of genome-wide association studies (GWASs) to investigate complex traits. Multi-dimensional high-resolution genomics data capture dynamics of cellular state and function, enabling the elucidation of complex biological networks. As echoed throughout the recent Keystone Symposium, vast amounts of genomic data are being generated, with implications for health and disease at both the population and individual levels. A key challenge for the community is how to best use this deluge of data. We need significant efforts in informatics, analytic methods development, meta-dimensional data integration, data sharing, data visualization, and strategies for bringing actionable biology into healthcare. The recognition of biology as a complex and informational science was seen as being first and foremost, and there was palpable excitement that we are truly on the cusp of discoveries that will revolutionize human health.

A view from the GWAS community

Current efforts of the GWAS community can be divided into two broad categories: discovery of novel risk loci, and extraction of biologically meaningful information from identified loci. Mark McCarthy (Oxford University) illustrated several successful approaches under way in type 2 diabetes (T2D) research. Meta-analysis of multiple

*Correspondence: bstranger@rics.bwh.harvard.edu
[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, 77 Avenue Louis Pasteur, Boston, MA 02115, USA
Full list of author information is available at the end of the article

case-control cohorts has resulted in new associations and identified a large number of variants with diminishing effect sizes. Fine mapping in non-Caucasian populations, functional genomics, and network-based approaches in disease-relevant tissues are identifying causal genes and elucidating functional mechanisms. Elizabeth Speliotes (University of Michigan) echoed these themes, with approaches implicating novel genes and pathways in obesity and non-alcoholic fatty liver disease. However, how to translate potentially causal genes into therapeutics can be less transparent. Sekar Kathiresan (Massachusetts General Hospital) described a Mendelian randomization approach to test whether the association of higher plasma high-density lipoprotein cholesterol (HDL-C) with reduced myocardial infarction (MI) risk is causal. The results of the study challenge the idea that raising plasma HDL-C will reduce MI risk; it is an important cautionary tale demonstrating that robust disease biomarkers may not always be feasible as therapeutic targets. A shift towards evaluation of rare and low-frequency variant effects on complex traits through whole-genome and exome sequencing is currently under way, as are epigenome-wide association studies, as exemplified by a genome-wide study of brain methylation in Alzheimer's disease (described by Manolis Kellis, Massachusetts Institute of Technology).

Many of these studies are moving from identified associations to an understanding of function. Kellis' whirlwind tour of data resources and analytic tools illustrated how the ENCODE project's data are being used to annotate dynamic regulatory elements in multiple human cell types, and can be mined to develop models of genetic effects.

Focus on health disparities

A workshop was held with the aim of better understanding how genomics research informs and impacts issues related to health disparities. Joshua Akey (University of Washington) provided a population genetics perspective by describing the National Heart, Lung, and Blood Institute Exome Resequencing project, comprising high-coverage exome sequencing of over 2,000 African-American and European-American individuals. A high

number of predicted deleterious variants were identified per individual, with the overall frequency spectrum dominated by very rare (mostly singleton) variants, consistent with human population demography models.

Population-level differences with respect to disease risk, drug efficacy, and side effects are areas in which the interplay of population genetics and functional genomics can inform mechanism. One of us (MED) described pharmacogenomics of anticancer agents in different populations. Cell-based models using HapMap lymphoblastoid cell lines are being used to elucidate functional effects and mechanisms of genetic variants influencing chemotherapeutic susceptibility. In contrast to trait mapping in ancestry-homogeneous populations, Elad Ziv (University of California, San Francisco) illustrated how populations of mixed ancestry can be used to map risk variants contributing to differences in disease incidence or age of onset, specifically focusing on benign neutropenia and breast cancer.

Personalized genomics

Personalized cancer therapeutics was a recurrent theme of the meeting. Joseph Lehár (Novartis Institutes for BioMedical Research) described large-scale efforts to test 45,000 drug combinations for synergy in 1,000 well-characterized cancer cell lines. These data, which are available as part of the Cancer Cell Line Encyclopedia, could facilitate methods development for linking pharmacological susceptibilities with genetic variation. Andrea Califano (Columbia University) described efforts to reconstruct and interrogate the regulatory logic of the cancer cell and develop a novel framework for cancer target discovery in a patient-specific manner. Dana Pe'er's (Columbia University) efforts to characterize patient-specific tumor network models are another step towards providing individualized treatment.

The realization that we are in the era of genomic medicine was emphasized by Atul Butte and Euan Ashley (both from Stanford University), who individually presented different aspects of the analysis of Stanford University investigator Stephen Quake's personal genome sequence. Together, they have created the largest curated database of human-disease associated single nucleotide polymorphisms (SNPs), and developed a pipeline for the analysis of clinically actionable findings from personal genomes. Butte discussed the importance of controlled vocabulary and methods for translating risk and effect size to clinicians, who will soon be faced with billions of patient data points to interpret.

Technology innovation

Pacific Biosciences' Stephen Turner presented the company's revolutionary technology that follows real-time enzyme activity, demonstrating that in addition to long-read DNA sequencing, the technology also characterizes nucleotide base modifications. The impression is that this technology is a novel frontier, but as detailed by Eric Schadt (Pacific Biosciences and Mount Sinai School of Medicine), it has already been deployed in important problems, including the 2011 *Escherichia coli* outbreak in Germany. It is expected that contributions from real-time understanding of living systems will be made in the near future. Similarly fascinating is Garry Nolan's (Stanford University) application of mass flow cytometry to sort sub-classes of leukocytes and then organize them into cellular networks to be used for more precise diagnosis. This application of flow cytometry evolved from an urgent clinical need - to individualize treatments of life-threatening lymphomas - and amazing results have been reported. In this capacity, Nolan's (and Pe'er's) work stands out as one of the few examples of the application of a computational systems approach currently in use in clinical care. Another exciting development is Leroy Hood (Institute for Systems Biology) and colleagues' efforts to make blood a 'window' into health and disease through the monitoring of organ-specific proteins in the blood.

On the informatics infrastructure side, Jeff Hammerbacher (Cloudera, Inc.) gave the 'Facebook' view of medical informatics. A completely information-driven schema based on a petabyte-scale platform for computational applications will allow routine data gathering and access. Another important data source is high-throughput genomic data readily available on the Internet, which have been minimally analyzed and from limited perspectives. Joel Dudley (NuMedii) presented a strategy for drug repositioning, in which publicly available gene expression data are used to predict new and often unexpected indications for established drugs.

Data sharing

Given that vast volumes of high-throughput genetic and genomic data are being gathered at an increasingly faster pace, Stephen H Friend (Sage Bionetworks) emphasized the need for more efficient data sharing and storage to enable discovery. Using Sage Bionetworks as a raw model, a 'federation' for efficient data sharing, storage and access has been formed in which members can collaboratively build disease models. Vicki L Seyfert-Margolis (US Food and Drug Administration) provided the administration's perspective on ways to enable drug trial data to be reused by the scientific community.

Informed consent is an important aspect of genomics data sharing. Jason Bobe (PersonalGenomics.org) detailed the problems inherent in making assurances to research volunteers that 'de-identified' or 'anonymized' data will remain confidential, even if data are shared widely. Bobe presented an 'open consent' solution stipulating that researchers: (1) do not promise anonymity and

confidentiality of data and (2) acknowledge risks of being re-identified from public data. Several speakers suggested that the public's apprehension about genomics data sharing will likely be tempered by actionable discoveries.

Network modeling

Another predominant theme at the meeting was systems-based approaches, including network or pathway modeling. The idea that GWAS has not uncovered the majority of the heritability for complex traits has been widely discussed, and here, the point was made that the simple, additive genetic components have been explored but the remaining 'missing heritability' lies elsewhere in the universe of molecular and cellular biology. For example, although common variation at the DNA level has been densely and routinely explored for single SNP associations, interactions have been largely ignored. Alexis Battle (Stanford University) described elegant approaches to look at epistasis, which has been considered primarily in model organisms (and was discussed by Leonid Krugylak, Princeton University, and Andy Clark, Cornell University). Methodologies and study designs that improve statistical power in humans are necessary and are clearly in development. For example, Trey Ideker (University of California, San Diego) described a framework integrating physical and genetic interaction maps to model regulatory and signaling networks, with implications for network-based patient stratification and drug target discovery.

In addition to complex interactions at the DNA level, a central focus is the integration of multiple data types. Meta-dimensional analysis, as described by one of us (MR), allows the consideration of variability that occurs through the genome, including gene expression patterns and proteomics. MR and colleagues have developed a data integration approach using evolutionary computing techniques along with data mining algorithms, such as neural networks. This type of analytical approach was also implemented by Iya Khalil (GNS Healthcare), who has used the methodology to predict disease phenotypes for complex traits. Pe'er presented novel approaches to integrate heterogeneous genomic data types into patient-specific tumor network models to identify key cancer drivers and their associated phenotypic effects, as well as to interrogate functionality of drug perturbations.

Considering data analysis in this comprehensive manner is supported by the evidence observed in several applications of expression quantitative trait loci (eQTLs), including in inflammatory disease (BES), T2D (Judy Zhong, New York University Medical School), and coronary artery disease (JB). Collectively, these studies emphasize that success depends on the collection of study populations, generation of high-throughput, well-defined cell- and tissue-specific genomic and phenotypic data, and development of powerful analytic strategies for meta-dimensional analysis. To truly elucidate this architecture, parallel non-human strategies are also needed, as highlighted by the efforts of Allan Attie (University of Wisconsin-Madison) to define eQTLs in mouse strains with a wide variety of disease susceptibilities.

The future

Leroy Hood's keynote address provided a big-picture view of the future of medicine. He predicts that we will transition from a clinically reactive to a proactive model, encompassing predictive, personalized, preventative, and participatory, or 'P4' medicine. This way of thinking relies on recognizing medicine as an informational science, both hypothesis-driven and hypothesis-generating, where systems approaches will allow one to understand wellness and disease in a more holistic way. Emerging technologies will allow us to explore new dimensions of patient data space, and new analytic tools will allow us to decipher the billions of data points for each individual. From the cutting-edge research discussed at this meeting, we can see that we are well on our way to that future.

Author details

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, 77 Avenue Louis Pasteur, Boston, MA 02115, USA. [2]Harvard Medical School, Boston, MA 02115, USA. [3]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA. [4]The Cardiovascular Genomics Group at the Vascular Biology Unit, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm 171 77, Sweden. [5]Department of Medical Pathology and Forensic Medicine, University of Tartu, 50090 Tartu, Estonia. [6]Department of Medicine and Committee on Clinical Pharmacology and Pharmacogenomics, University of Chicago, KCBD, 900 E. 57th Street, Chicago, IL 60637, USA. [7]Center for Systems Genomics, Pennsylvania State University, University Park, PA 16802, USA.