



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Challenges in clinical genomics

The Harvard community has made this article openly available.
[Please share](#) how this access benefits you. Your story matters.

Citation	MacArthur, Daniel G. 2012. Challenges in clinical genomics. <i>Genome Medicine</i> 4(5): 43.
Published Version	doi:10.1186/gm342
Accessed	February 19, 2015 12:07:59 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:11177922
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

MEETING REPORT

Challenges in clinical genomics

Daniel G MacArthur^{1,2*}

Abstract

A report on the Genomic Disorders 2012: Genomics of Rare Diseases meeting, Hinxton, UK, 21-23 March 2012.

Genomics enters the clinic

The central theme of this year's meeting was very clear: genomic technologies are now entering clinical laboratories across the world on a scale that would have seemed unimaginable even a year ago.

The technology that has made the greatest impact in the past 12 months has been exome sequencing, which captures only the protein-coding regions of the genome. Despite only exploring about 1% of the sequence in a genome, it is a cost-effective approach for patients with severe diseases, in whom the majority of causal mutations are expected to disrupt protein sequence.

Exome sequencing is now being applied to patients on a massive scale - multiple presenters, including Han Brunner (Nijmegen Centre for Molecular Life Sciences) presented exome data from dozens or even hundreds of individuals - and that scale is set to increase exponentially. Margriet von Kogelenberg (Wellcome Trust Sanger Institute) described the UK-based Deciphering Developmental Disorders (DDD) project, which is currently recruiting 12,000 patients with developmental delay for array comparative genomic hybridization analysis and exome sequencing.

Genomics has unquestionably accelerated the discovery of the mutations underlying severe Mendelian diseases. Dozens of novel disease-associated genes were presented at the meeting, and there is a widespread (and, I suspect, justified) view that a substantial majority of Mendelian disease genes will be identified within the next 18 to 24 months.

In addition, the technology is already moving beyond the research domain and into the world of clinical diagnostics. Most of the research projects presented here

generate results that flow (via a process of validation) back to clinicians and patients. The task of assembling systems for translating genomic data into clinical interpretation was thus a key topic of discussion.

This meeting provided an eloquent counterpoint to media reports on the 'failure' of the Human Genome Project: medicine is clearly being transformed by genomics, with rare diseases at the vanguard of that transformation. There has never been a more exciting time to be working in the field.

The interpretation challenge

Although this meeting illustrated the acceleration of mutation discovery by next-generation sequencing, it is worth noting two key challenges to translating large-scale sequence data into clinical utility raised in presentations and informal discussions.

Firstly, there is the difficulty of answering a straightforward question: has a mutation seen in a patient with disease been previously reported in another individual with a similar phenotype? Unfortunately, as those working in the field can testify, obtaining this information is frustrating: there is currently no single, comprehensive, open-access database of known disease-causing mutations. Instead, knowledge is split between a variety of resources of varying comprehensiveness, usability and access policies. The lack of a reliable global database of known disease mutations remains one of the most embarrassing failures of modern human genetics.

Two presenters discussed independent approaches to addressing this failure: Donna Maglott (National Center for Biotechnology Information) discussed the relatively new ClinVar database, and Joanna Amberger (Johns Hopkins University) discussed the modernization of the venerable Online Mendelian Inheritance in Man (OMIM) resource. Both efforts are admirable, as are other independent ventures in the same field, such as mutaDATABASE and the Human Variome Project. Yet I and other attendees were left wondering which of these resources, if any, will ultimately provide the one-stop shop for high-confidence human disease mutations that is so sorely needed in the genomic era; certainly none do so currently.

Another challenge was pervasive in informal conversations at the meeting, but only rarely addressed in

*Correspondence: macarthur@atgu.mgh.harvard.edu

²Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

Full list of author information is available at the end of the article

presentations: exactly what are the standards that we should be applying when confirming that a novel mutation is in fact disease-causing? Several speakers, including James Lupski (Baylor College of Medicine), noted that many published disease-causing mutations are demonstrably erroneous; it is clear that historical standards for proving causation have been inadequate. The Mendelian genomics community needs stringent guidelines for establishing significant findings, similar to those developed by researchers working on genome-wide association studies of complex traits.

Yet at this meeting there was no clear consensus on what such guidelines should look like, and there were worrying signs that false mutations from exome studies may already be creeping into the literature: for instance, many researchers seem to regard a *de novo* protein-altering mutation as having a very high probability of disease causation, when in fact such variants are carried by about 50% of healthy individuals. It is to be hoped that eminent researchers and journal editors will work together to develop strong and transparent standards for the field; otherwise, in an era of cheap exome sequencing, even a low false positive rate will rapidly swamp the literature with spurious findings.

Moving beyond the exome

Despite the current focus on exome sequencing, geneticists appreciate the importance of making sense of genetic variants that fall outside protein-coding regions: a non-trivial fraction of the mutations underlying rare diseases, and variants underlying the majority of the risk of common complex disorders, fall in the 'dark matter' outside the exome. This meeting provided reasons for both optimism and pessimism about the future of interpreting genetic variation within that dark matter.

The optimism comes from the unraveling of the functional map of human non-coding DNA through the application of new genome-scale approaches. Many of these approaches have been pioneered in the context of the massive ENCODE project, which is coming to fruition this year. Lucas Ward (Massachusetts Institute of Technology) presented data on behalf of ENCODE showing that many of the variants implicated in complex diseases either disrupt or create DNA sequences that are bound by transcription factors. Non-coding mutations can also underlie Mendelian disorders, of course: Cornelis Albers (Wellcome Trust Sanger Institute) described a wonderful genetic detective story in which a thrombocytopenia syndrome turned out to be caused by compound heterozygosity for a rare null allele and a more common non-coding polymorphism in the *RBM8A* gene, which encodes a vital component of the exon-junction complex involved in RNA processing.

The pessimism comes from data suggesting that global prediction of the functional impact of non-coding variants will be challenging even with genome-scale data. Joseph Hiatt (University of Washington) described an elegant experimental approach to explore the impact of mutations in enhancer elements, which showed that the vast majority of changes to enhancer elements have small or undetectable impacts on gene expression because of functional redundancy. Although Hiatt's approach provides an experimental method for exploring the effects of variation in enhancer elements, it also indicates that *de novo* prediction of these effects on the basis of sequence data alone will be extremely challenging.

The rate at which functional annotation of non-coding DNA is progressing is astonishing, and it is clear that many profound biological insights are emerging from the ENCODE project. However, it is equally clear that we remain far from a general model accurately predicting the functional impact of genetic variation in non-coding DNA.

Data: to return or not to return?

The final session of the meeting focused on ethical and policy issues. A strong theme in this session, and indeed throughout the meeting, was a familiar dilemma: should research participants be given access to 'incidental findings' related to disease risk but irrelevant to the primary research question - for instance, a mutation in the *BRCA1* breast cancer gene identified in a study on intellectual disability - and if so, how should this be done?

Opinions on this issue varied wildly among meeting participants; although this session by no means resolved the debate, it provided useful food for thought. Jonathon Berg (University of North Carolina) presented a strategy for binning genetic findings into categories related to their risk and clinical utility, and argued that the majority of genetic variants in a patient's genome lack clinical utility and should be excluded from medical records. Caroline Wright (Wellcome Trust Sanger Institute) discussed the data return strategy for the DDD project, under which 'clinically pertinent' findings related to the patient's primary diagnosis will be returned, but incidental findings will not.

Hopefully data return policies will ultimately be driven by the desires of participants rather than the views of ethicists. Unfortunately we currently know surprisingly little about participant expectations, but this is set to change: the DDD is currently coordinating an open survey at the GenomEthics website, of the views of research participants and other stakeholders, in which I would encourage readers to participate.

The informed uncertainty in this session was a fitting end to the meeting, which showed both that genomics is already transforming the diagnosis of rare diseases, and

also that confusion about important scientific and ethical issues remains to be resolved by the community. I look forward to seeing how the situation has changed at this meeting in 2013.

Abbreviations

DDD, Deciphering Developmental Disorders Project

Competing interests

The author declares that he has no competing interests.

Author details

¹Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA. ²Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA.

Published: 22 May 2012

doi:10.1186/gm342

Cite this article as: MacArthur DG: Challenges in clinical genomics. *Genome Medicine* 2012, **4**:43.