

MEETING REPORT

Back to the family: a renewed approach to rare variant studies

Dina Zielinski¹, Melissa Gymrek^{1,2,3,4} and Yaniv Erlich^{1,*}

Abstract

A report on the 62nd Annual Meeting of the American Society of Human Genetics, San Francisco, California, USA, 6-10 November 2012.

Introduction

The annual meeting of the American Society of Human Genetics is a major - some would say overwhelming - conference that aims to present the state of the field as a whole, with presentations on nearly all aspects of human genetics. This year, the meeting had a record number of nearly 7,000 participants, with approximately 450 scientific presentations and over 3,200 posters. In addition, more than 200 vendors presented their products, including all major DNA sequencing companies, bioinformatic services for data analysis, and clinics specializing in genetic medicine. In order to navigate this scientific maze, the organizers developed a smartphone application to allow users to browse abstracts based on schedule, presenter, and topic, and to generate a snapshot of the current events at the conference. Intriguing buzzwords had participants running back and forth between rooms to hear speakers from parallel sessions.

Some of the most frequent terms used throughout the meeting, according to quantitative text analysis of #ASHG2012 tweets, were 'rare' and 'common' with both almost equally represented. Indeed, in the spirit of the US presidential election, which coincided with the first day of the meeting, there seemed to be a common and rare variant party division. Entire sessions were devoted exclusively to either class of variant, such as 'GWAS from head to toe' and 'Cancer genetics I: rare variants'. In an attempt to find a bipartisan resolution, one session was dedicated to 'Common variants, rare variants, and

everything in between', and presented the advantages of an integrated approach that examines association signals of both variant classes. Interestingly, several talks showed that such integrative studies based on whole exome sequencing can simultaneously replicate known genome-wide association study (GWAS) signals and uncover a distinct set of genes that harbor rare etiological variations.

The interpretation gap

A recurrent issue throughout the meeting was the gap between data generation and data interpretation, especially for rare variants. This challenge is exacerbated by the application of sequencing in the clinic, where evaluation of pathogenic variants and incidental findings may determine the course of treatment. A significant number of talks described potential techniques to overcome this interpretation gap. Some speakers, such as Heidi Rehm (Harvard Medical School, USA) suggested the development of specific databases for clinical interpretation. Others, including Marc Greenblatt (University of Vermont, USA), focused on the development of standards for variant interpretation. Another set of presentations described the recent successes of community-based interpretation contests, such as Boston Children's Hospital's CLARITY challenge and Berkeley's Critical Assessment of Genome Interpretation.

The main challenge when studying rare variants is that robust statistical inference of their effects in case-control studies requires a large amount of sequencing data. Daniel MacArthur (Massachusetts General Hospital, USA) suggested a brute force approach that includes sequencing a large number of individuals, postulating that 'in order to understand one genome, we need to sequence tens of thousands of genomes.'

Family studies: a shortcut to analyze rare variants

Other speakers suggested alternative, more efficient approaches to rare variant interpretation. In fact, one of the prevailing themes at this year's meeting was the renewed interest in large pedigrees and isolated populations to assess the effect of rare variants on common traits. Michael Province (Washington University, USA)

*Correspondence: yaniv@wi.mit.edu

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

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

presented one potential problem of the brute force approach in a session centered on family studies as a means to investigate complex traits. Citing recent studies about rapid population growth in humans, he noted that the number of extremely rare alleles in the population is much higher than thought. Even doubling the sample size does not help very much, as the new sample will simply present new rare alleles, rather than add statistical power. As an alternative, he suggested focusing on large pedigrees, where the allelic diversity is smaller: 'pedigrees make the needle [rare variants] in the haystack bigger' and thus easier to find. Further, the extensive identity-by-descent (IBD) between individuals helps to distinguish true rare variant calls from sequencing errors and provides a means to verify novel alleles in multiple related individuals.

Robert Elston (Case Western Reserve University, USA) was the most insistent about the current appeal of family studies in a special session recognizing his 80th birthday. He went so far as to say that 'somehow, for the last decade or so, we were misguided into thinking that families were not necessary, and we have seen epidemiologists having a ball with case-control studies and honestly believing that they are doing genetic research!' Elston did say that family studies may one day in the distant future be dispensable, but maintained that, for now, it is crucial to study variants in the context of inheritance, rather than simply as DNA.

From interpretation bottlenecks to genetic bottlenecks

In addition to the renewed interest in family studies, several speakers highlighted the value of studying isolated populations. Jeffrey O'Connell (University of Maryland, USA) described a study of complex traits in the Amish population, in which he found a steady increase in the inbreeding coefficient in the last 200 years. He showed that, on average, a pair of individuals is as genetically similar as first cousins once removed. With such a strong genetic bottleneck, rare variants that segregate in the European population may increase by orders of magnitude in the Amish population, enabling robust statistical inference about their roles. To stress this point, he concluded his talk with a reminder that 'we study the Amish not because they are different but because they are us.'

William Scott (University of Miami, USA) and Cornelia van Duijn (Erasmus University Rotterdam, the Netherlands) presented an integrative approach for identifying pathogenic variants of complex disorders in isolated populations. Their method starts with linkage analysis to find large segments that segregate with a given phenotype, followed by whole exome sequencing to pinpoint the pathogenic variant in the linkage interval. This

technique showed mixed results. They were able to uncover a rare pathogenic variant in a study of depression but found no coherent signal in a study of Parkinson's disease, suggesting a potential role for non-coding variants.

Other presenters tried to reconcile the advantages of both traditional case-control and family studies. Hua Zhou (UCLA, USA) discussed combining genome-wide association mapping with pedigrees for quantitative trait locus analysis. Similarly, Richard Spritz (University of Colorado Denver, USA) presented an approach that integrates GWAS with the sequencing of siblings under a linkage peak. Elizabeth Thompson (University of Washington, USA) discussed using IBD within and between pedigrees, echoing Robert Elston's emphasis on the need for information on relatedness. She concluded that, eventually, pedigree and population studies will be equivalent, in the sense that we can use techniques for analyzing IBD to obtain the same information.

Conclusion

In the past few years, we have witnessed the emergence of large-scale sequencing projects to study common diseases, such as the NHGRI's ClinSeq study, NHLBI's Exome Sequencing Project, and The Personal Genome Project. The renewed interest in large pedigrees and isolated populations for complex trait studies at this year's meeting was refreshing. Several speakers highlighted the advantages of such designs in interpreting the role of rare genetic variations. In addition to facilitating the ascertainment of multiple individuals with the same rare variant, the substantial IBD in the samples promotes imputation and increases confidence in the sequencing results. Further, these designs afford a set of complementary tools, including linkage analysis and heritability measurements, that can accelerate genetic investigation.

By definition, personalized medicine entails drawing conclusions based on the study of a single genome from the general population. Despite the tremendous advantages of family and isolated population study designs, we should also remember that they do not entirely reflect the general population. For instance, the substantial IBD between participants in these studies increases the likelihood of overestimating the effect of a variant due to epistasis, a point that was emphasized in the opening talk at ASHG 2010 by Eric Lander (Broad Institute, USA). Another potential complication is the sampling of individuals from narrow environmental conditions, which is more prone to confounding gene-environment interactions. Evidently, nothing comes free in human genetics. Each study design has its own limitations and advantages, necessitating an integrative approach to bridge the interpretation gap and effectively handle today's population-scale datasets.

Meeting tweets are available online at #ashg2012.

Abbreviations

GWAS, genome wide association study; IBD, identity-by-descent.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This publication was supported by the National Defense Science & Engineering Graduate Fellowship and by a National Human Genome Research Institute grant R21HG006167. YE is an Andria and Paul Heafy Family Fellow.

Author details

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. ²Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA 02139, USA. ³Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ⁴Department of Molecular Biology and Diabetes Unit, Massachusetts General Hospital, Boston, MA 02114, USA.

Published: 19 December 2012

doi:10.1186/gm398

Cite this article as: Zielinski D, *et al.*: Back to the family: a renewed approach to rare variant studies. *Genome Medicine* 2012, **4**:97.