Using Whole Exome Sequencing to Walk from Clinical Practice to Research and Back Again

Running title: Schuler et al.; Whole exome sequencing in research and the clinics

Bryce A. Schuler, BA\textsuperscript{1,2*}; Sasha Z. Prisco, BS\textsuperscript{1,2*}; Howard J. Jacob, PhD\textsuperscript{1,2,3}

\textsuperscript{1}Human and Molecular Genetics Center; \textsuperscript{2}Dept of Physiology; \textsuperscript{3}Dept of Pediatrics, Medical College of Wisconsin, Milwaukee, WI

*contributed equally

Address for Correspondence:
Howard J. Jacob, PhD
Medical College of Wisconsin
8701 Watertown Plank Rd.
Milwaukee, WI 53226
Phone: 414-456-4887
Fax: 414-456-6516
Email: jacob@mcw.edu

Journal Subject Code: [109] Clinical Genetics

Key words: Editorial, exome, genetics, human
Whole exome sequencing (WES) is currently used to identify the genetic etiologies of many diseases, especially monogenic disorders. Ng et al., in 2009, completed the first proof-of-principle study demonstrating the feasibility of using exome sequencing to identify causal variants for diseases, specifically Freeman-Sheldon syndrome. Within two years, there was a marked increase in publications presenting WES data and the pace continues to accelerate (Figure 1). In 2010, WES began to be used for clinical diagnoses, particularly for Mendelian disorders. In early 2011, Worthey et al² used exome sequencing to facilitate clinical diagnosis and modify treatment in a single case. Despite many of the successes resulting from exome sequencing, more than half of the approximately 7,000 known or suspected Mendelian disorders have not yet been discovered,³ highlighting the need for more genetic, mechanistic and clinical studies, particularly if the data are to be used clinically. Moreover, as our knowledge of the genome increases, examples of some of the complexities associated with genotypic-phenotypic relationships further substantiate the need for both additional genomic annotation and many more sequenced genomes with phenotypic information. Some of these complexities include: 1) variants in the same genes may lead to different clinical manifestations or phenotypes; 2) what appear to be similar phenotypic observations may result from different causal disease variants operating through distinct pathophysiological mechanisms; and 3) the recent ENCODE papers which suggest that up to 80% of the human genome is functional.⁴

In this issue of Circulation, Crotti and colleagues⁵ use a clever strategy to investigate the cause of cardiac arrhythmias, ventricular fibrillation, and cardiac arrest in two unrelated probands, who were negative for variants in KCNQ1, KCNH2, SCN5A, KCNE1, and KCNE2, which are the genes most commonly involved in long QT syndrome (LQTS). Specifically, they began by performing WES on the two unrelated probands and their healthy parents. As with
other reported studies, they looked only at rare variants. Analysis of the WES results for the two probands uncovered *de novo* mutations in either *CALM1* or *CALM2*, which both encode for calmodulin. At this point, the team faced the problem all investigators and clinical teams face when a new and potentially relevant variant(s) is found—is it causal? Here the team used a validation strategy which other groups have also employed. They conducted follow-up genetic screening of the calmodulin genes (*CALM1, CALM2*, and *CALM3*) on an independent cohort of 82 subjects that had congenital LQTS without a known genetic cause. Their presumed rationale was: if these genes play a role in this larger cohort, then they would be expected to contain variants. This hypothesis was correct as two individuals from this cohort also had variants in calmodulin; one individual harbored the same mutation in *CALM1* as proband 1 and the other had a novel missense mutation in *CALM1*.

The findings by Crotti et al. are an example of the genotypic and phenotypic heterogeneity that has also been observed in other studies. A previous study showed that mutations in *CALM1* may be linked to catecholaminergic polymorphic ventricular tachycardia with no evidence of prolonged QT intervals. Whereas, the individuals with calmodulin mutations in this paper presented with early onset life-threatening cardiac arrhythmias, prolonged QT intervals, and neurodevelopmental delay. Since calmodulin is involved in a variety of different cellular processes, it is feasible that variants at alternate locations in the calmodulin genes could result in an assortment of distinct clinical presentations. Had the authors not pursued a series of functional assays, the variants detected would have remained promising candidates, but would have most likely been categorized as variants of uncertain significance (VUS) by most clinical laboratories. They note that “these findings suggest an intriguing genotype-phenotype correlation among calmodulin mutations, and further suggest different
pathophysiological mechanisms.” There are many occurrences where mutations within the same gene cause different phenotypes. For example, variants in exon 10/11 of RET are responsible for multiple endocrine neoplasia type 2A (MEN 2A, an inheritable cancer syndrome characterized by a propensity to develop medullary thyroid carcinoma, pheochromocytomas, and parathyroid hyperplasia), whereas variants in exon 16 of RET predispose individuals for developing MEN 2B, which presents with neuromas on the lips and tongue, a different body habitus than MEN 2A, and no parathyroid hyperplasia. In two published cases and two additional clinical cases from our institution, the causal variant found produces a phenotype discordant from what was previously published. These examples reiterate the importance of annotating the unique functional effects of each individual variant, but could also speak to a general comfort in assigning a new function to a known gene, rather than a function to a relatively uncharacterized gene. These examples of different clinical manifestations may not be an exception, further contributing to the difficulty of substantiating a variant from its classification as a VUS to a causal variant.

Alternatively, what appears to be the same clinical phenotype may actually be caused by variants in different genes and be a result of different pathophysiological mechanisms. Crotti et al. cites the use of a gene panel for the “most frequently mutated genes in LQTS,” providing an example of the convention of classifying pathologies like LQTS by their phenotypic endpoint rather than their molecular mechanism. The probands in this study presented with LQTS-like features and were compared with a congenital LQTS cohort. Without exome sequencing, the probands might have been clustered with other idiopathic congenital LQTS cases. However, with WES and the discovery of causal disease variants, the probands can now be classified as a distinct subset of congenital cardiac arrhythmias. It is also interesting to note that 80 individuals
within the congenital LQTS cohort described by Crotti et al. still lack a genetic diagnosis for their LQTS since they were negative for pathological variants in both the known LQTS genes and the calmodulin genes. These cases likely have alternate genetic etiologies such as variants in uncharacterized genes or noncoding mutations. As more genomes are sequenced and phenotypic annotation for each gene is improved, defining the molecular defect will improve the accuracy of the diagnosis, provide additional prognostic information, and hopefully help develop future therapies.

In order for the discoveries of pathological variants to reach their full clinical utility, research efforts will need to focus on understanding the molecular mechanisms contributing to these disease processes. For example, the discovery of pathological variants in calmodulin has raised a variety of interesting and novel research questions. As previously discussed, variants in the calmodulin genes are not commonly associated with LQTS and confer a variety of phenotypes, which stresses the need to further characterize the different molecular mechanisms that could lead to these distinct pathologies. Additionally, the presence of three calmodulin genes and the severity of the phenotype that results from a heterozygous variant in one of these genes raise questions regarding gene-dosage, the function of each of the calmodulin genes, and the evolutionary significance of the calmodulin gene family. Clearly more research is needed to elucidate the mechanisms by which calmodulin dysfunction leads to cardiac arrest, refine current disease classifications, and learn more about the genomic etiologies of other diseases. In the interim, will these variants or others in these genes be considered as causal in other clinical cases?

Finally, future genetic studies need to account for the existence of genotypic and phenotypic heterogeneity. This study by Crotti et al. exemplifies the importance of classifying
cases with very similar phenotypes (i.e. ventricular arrhythmias with neurological complications) together in order to identify causal variants. As another example, the same group of researchers who identified candidate genes for Freeman-Sheldon syndrome (Ng et al.) sequenced the exomes of ten unrelated individuals affected with Kabuki syndrome. The researchers applied the same variant prioritization scheme applied previously, but could not identify a likely candidate gene that had variants observed in all of the individuals. When they repeated the analysis with a subset of the affected individuals and accounted for phenotypic heterogeneity by stratifying and ranking cases based on “the presence of, or similarity to, the canonical facial characteristics of Kabuki syndrome, and the presence of developmental delay and/or major birth defects,” they were able to identify that causal variants in MLL2 are linked to Kabuki syndrome. This suggests that genetic and phenotypic heterogeneity underlies the clinical presentation of this disorder and again substantiates the argument for careful classification of diseases and the need for additional functional studies.

While there has been great advancement in the discovery of novel causal variants for various disorders with exome sequencing, there are nevertheless limitations with the use of exome sequencing. Exome sequencing does not sequence all exons, does not include noncoding regions (which include many evolutionarily conserved regulatory regions that have been previously associated with disease), and does not detect structural variants or chromosomal rearrangements. In addition, the recent flurry of publications by the ENCODE consortium reports that 80% of the human genome is functional. In our opinion, WES is an intermediate stop on the cost curve towards whole genome sequencing (WGS). As the price drops for WGS and investigators and clinicians identify more causal variants that fall outside the exome region, there will be an increased demand for the use of WGS.
Crotti et al. began with a clinical research question; they found that variants in calmodulin might lead to cardiac arrhythmias, ventricular fibrillation, and cardiac arrest. They confirmed this hypothesis by looking in a larger cohort and completing mechanistic studies to demonstrate that variants in calmodulin are likely disease candidates. The question now is whether other clinicians and investigators will use this information to make additional diagnoses. The authors clearly outline several open research questions that need to be answered. As WES becomes more widely used for patient-centered work (Figure 1), there will be increased pressure on clinicians to make a diagnosis and determine the functional significance of specific variants. This transformation will be heavily reliant on our ability to annotate function to genomic variation and classify pathologies with a high degree of fidelity. As more and more genomes are sequenced, we will have an increased ability to screen for rare variants. How we ascertain whether a particular variant is truly causal remains to be determined, but the strategy used here provides another approach to help resolve this challenge. As we try to resolve this question, we also need to address the secondary findings that each genome holds. With further genomic resolution, we can more effectively integrate WES and WGS into research and clinical practices.

Conflict of Interest Disclosures: None.

References:


**Figure Legend:**

**Figure 1.** Number of Exome Sequencing publications by year. “Human Exome Sequencing” and “Human Patient Exome Sequencing” were searched in PubMed. Reviews, perspectives, and methodology papers were excluded.
Using Whole Exome Sequencing to Walk from Clinical Practice to Research and Back Again
Bryce A. Schuler, Sasha Z. Prisco and Howard J. Jacob

Circulation. published online February 6, 2013;
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/early/2013/02/05/CIRCULATIONAHA.113.001284