JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY VOL. 72, NO. 4, 2018

PUBLISHED BY ELSEVIER

EDITORIAL COMMENT

Genome Sequencing in Hypertrophic Cardiomyopathy*

Euan A. Ashley, MB ChB, DPhil, Chloe M. Reuter, MS, Matthew T. Wheeler, MD, PhD

As early as the 1700s, the association of cardiac enlargement with sudden death was recognized and labeled as "hypertrophy" (1), but it was not until the mid-20th century that the inheritance patterns, clinical correlates, and physiology of hypertrophic cardiomyopathy were characterized (2,3). Variation in genes encoding the cardiac sarcomere was found to be the cause of hypertrophic cardiomyopathy in the early 1990s (4), and around 15 years later, genetic testing moved from research laboratories to a clinical laboratory setting. Today, clinical genetic testing has power in identifying the true cause of hypertrophy, distinguishing, for example, Fabry disease (5) or Danon disease (6) from sarcomeric hypertrophic cardiomyopathy (7). It also powerfully facilitates family screening (8). Genetic evaluation is recommended for all patients and includes assessment by a genetic counselor, annotation of a detailed multigeneration pedigree, and sequencing of a panel of around 30 genes. Such panels use capture kits that pull down DNA from the genes of interest for next-generation, short-read sequencing. This contrasts with the earlier approach where the coding regions (exons) of the genes of interest were specifically amplified then sequenced by the dideoxynucleotide method of Sanger. Despite the large number of genes tested (some HCM [hypertrophic cardiomyopathy] panels interrogate close to 100 genes), pathogenic or likely pathogenic variants are found in a much smaller number of genes. For example, variations in the cardiac myosin heavy chain (*MYH7*) and myosin binding protein C (*MYBPC3*) genes made up 83% of the positive cases in 1 study (9). Although a mechanistic basis for this "thick filament" subcategory of HCM has recently been proposed (10,11), it remains less clear why patient and population disease variation differs markedly between these 2 genes: MYH7 exhibits extreme intolerance to disruptive missense (single nucleotide) variation, which causes disease, whereas MYBPC3 appears to generally tolerate missense variation but causes disease via haploinsufficient loss of function mechanisms (12). Such mechanisms include variants that truncate the transcript or cause splicing disruption with incorporation of pseudoexons or early termination of translation. In addition, we increasingly recognize broadly 2 clinical forms of the condition. In the clearly familial form of the disease, younger patients with commonly asymmetric and more severe hypertrophy are found to harbor causative genetic variants approximately 32% of the time (9), (a number that implies that causative genes remain to be discovered). In a form of hypertrophic cardiomyopathy presenting at an older age and associated with more discrete, often proximal, hypertrophy (13), rarely is a family history of the disease found and rarely are genetic tests positive (14). Thus, hypertrophic cardiomyopathy appears to be many distinct entities, something the science of precision medicine aspires to elucidate (15,16).

The ability to sequence almost all of the human genome is a relatively recent phenomenon. The Human Genome Project, funded with a $3 billion grant from the U.S. Department of Energy and the National Institutes of Health, produced a single monoploid human sequence over the course of 10 years. Its final acceleration was driven in part by competition with the private enterprise led by Craig Venter, whose

*Editorials published in the *Journal of the American College of Cardiology* reflect the views of the authors and do not necessarily represent the views of *JACC* or the American College of Cardiology.

From the Center for Inherited Cardiovascular Disease, Stanford University, Stanford, California. Dr. Ashley is founder of Personalis Inc. and Deepcell Inc.; and is an advisor to Genome Medical and SequenceBio. Dr. Wheeler has equity interest in Personalis Inc; and has been a consultant to MyoKardia. Ms. Reuter has reported that she has no relationships relevant to the contents of this paper to disclose.

ISSN 0735-1097/$36.00 https://doi.org/10.1016/j.jacc.2018.05.029

genome was sequenced at a cost of around US$100 million. The cost and time taken to sequence a human genome dropped dramatically over the following decade such that it is possible now to sequence a human genome for under US$1,000 in <48 h (16–18). Given the charge for many multigene panel tests remains at several thousand U.S. dollars, the question has recently arisen as to whether there might be utility in replacing such panels with, or reflexing from such panels to, genome sequencing.

SEE PAGE 419

In this issue of the *Journal*, Bagnall et al. (19) describe the results of the application of genome sequencing to the molecular diagnosis of HCM. The authors report on 2 cohorts: patients with prior genetic testing that was unrevealing and patients with no prior genetic testing. Genetic analysis focused on 58 unrelated patients, 14 affected family members, and 2 unaffected parents of a severely affected proband. One hundred and eighty-four causal and candidate left ventricular hypertrophy genes were examined, in addition to variants in the mitochondrial genome. Their results suggest that genome sequencing increases the sensitivity of detecting disease-causing variants, mostly through the discovery of intronic variants that impact splicing. They report that genome sequencing yields a molecular diagnosis in an additional 20% of patients with inconclusive prior genetic testing. Forty-two percent of HCM patients with genome sequencing as a first-line test received a molecular diagnosis.

In comparison with multigene panels, genome sequencing could be expected to identify additional causal mechanisms of disease through several avenues: 1) sequencing a broader range of genes; 2) sequencing intronic and intergenic regions; and 3) exhibiting greater sensitivity for structural variation detection (facilitated by more full and even genome coverage). It might be expected that advancing from sequencing 30 genes to 20,000 genes would greatly improve the diagnostic rate, but the gains in this domain appear relatively muted. In recent work from Cirino et al. (20), genome sequencing of 41 patients with HCM led to 1 additional diagnosis. Indeed, both that group and Bagnall et al. (19) made an additional diagnosis through uncovering *the same variant* in *PTPN11* (c.1403C>T), the gene causative of Noonan syndrome. The idea that we might be missing milder forms of Noonan syndrome and other RASopathies was explored by a group who tested 11 RASopathy genes in 1,111 individuals referred for genetic testing for hypertrophic or dilated cardiomyopathy. Disease-causing variants were identified in 4 of the 692 hypertrophic cardiomyopathy patients, suggesting a much lower yield than would be inferred from the 2 genome sequencing studies discussed here but, nevertheless, providing support for the inclusion of RASopathy genes in broader cardiomyopathy panels (21). Further support for more expansive gene interrogation is provided by an exome sequencing study in an inherited cardiovascular disease setting at Yale (22). Out of 200 participants with a spectrum of inherited cardiovascular disease, a causal variant was identified by exome sequencing in 53 (26.5%). It was estimated that only 36 (18%) of these variants would have been identified by commercially available panels. For the 28 patients presenting with HCM in that study, the overall rate of identifying a pathogenic or likely pathogenic variant was 46%, consistent with the rate reported in the Bagnall et al. (19) paper (42%) and higher than that reported for large panels alone (32%) (9). The Yale authors highlighted gains from exome sequencing, particularly in inherited arrhythmia syndromes, and especially aborted sudden cardiac death, where there was uncertainty over the primary diagnosis (arrhythmia vs. cardiomyopathy). For similar reasons, whole exome sequencing has been shown to provide value in molecular autopsy (23,24).

Despite the opportunity, there remain challenges in deploying genome sequencing as a tool to interrogate more genes: without sound evidence that putative genes are causal for disease, it is hard to move variants into the “likely pathogenic” category for any given individual. Adding further to this challenge, the case for many genes previously held to be causal appears to diminish when population site frequency spectra are taken into account (25) leading to a shortening rather than lengthening of our causal gene list. Because family cascade genetic testing requires at least a confidence of “likely pathogenic,” variants outside of the core causal genes will have limited utility until more evidence can be gathered to move the gene firmly into a disease-causing category. A critical question, then, is how to find new genes and confidently ascribe causality to them.

The most exciting finding in the current paper was the yield from interrogating deep intronic variants found to affect splicing of *MYBPC3*. Splicing is the method by which introns are removed from precursor messenger RNA and exons are ligated together to form the mature messenger RNA that is translated into protein. The process is complex and mediated by the spliceosome. Although the critical elements for spliceosome function are found immediately upstream and downstream of the exons (the canonical dinucleotides and acceptor sequences), variation deep