CASE REPORT

Knowing Too Much, but Understanding Too Little: Making Sense of Genotype-Phenotype Correlation in Modern Day Medicine

***Hanna J. Tadros MBBCh, BAO*[1]**

[1]Pediatric Resident, Department of Pediatrics, University of Florida, Gainesville, FL

In as early as 1960, gene-linked causes of heart disease were proposed to the greater scientific community.[1] Time and time again since, multigenerational familial studies have identified genetic variation hypothesized to lead to cardiomyopathy and sudden cardiac death. This has often led to characterizing these mutations contained within one family as pathogenic. Variation, regardless of location or functional consequence, was deemed causative.

At the turn of the century, scientific endeavors to better understand the human genome were underway and near completion. These efforts were believed to lead to a better understanding of genetic variation and genotype to phenotype correlations. What we failed to foresee was the conundrum that followed, *knowing too much, but understanding too little*. Mutations with conflicting interpretations of pathogenicity, variants of undetermined significance, and benign mutations previously proven to be pathologic muddle our understanding of genetic interpretation. Furthermore, with the introduction of large variant data in the form of genetic population databases, mutations previously reported to be causative of disease were found in seemingly ostensibly healthy individuals. With increasing numbers of individuals with identified mutations, cohorting those with mutations has directly contradicted previously accepted phenotypical tendencies derived from multigenerational family studies. In addition, functional studies in animal models frequently contradicted each other, further increasing confusion within the scope of cardio-genomics.[2,3]

A prime example is variation within the Troponin Complex. Initial studies were concerned with finding any variation of the genome at specific locations. Once those were found, it became well established that variation in the Troponin complex (TNNT2-encoded Tropinin T, TNNI3-encoded Troponin I, and TNNC1-encoded Troponin C) were highly likely to lead to cardiomyopathy. However, large population databases, like the GnomAD database, exhibit that even in seemingly healthy individuals, variation in the Troponin complex exists.[4] How could a healthy individual have variation at a location where proven pathogenic mutations have been found? This would go against all scientific certainty that location of variation determines functional changes and thus, phenotype. We found ourselves with too much information, but no tools for interpretation. As such, it is not surprising that much scrutiny is being exercised regarding claiming causation when variation is identified in a cardiomyopathy patient.

Efforts within the genetic community are underway to address these exact issues. Landstrom et al. have proposed a novel approach to mapping frequencies from pathologic cohorts against those from the general population with rare variation and proved utility in identifying areas of heightened gene penetrance and pathogenicity.[5-7] Large population studies have also exhibited tendencies for certain variations to prove more pathologic than others. Theoretically, radical mutations, or those that cause deletions and "frame-shift" changes, are more likely to cause large changes in amino acid sequences. Furthermore, theories that "more" mutation leads to "more" phenotype, although not always true, is frequently the case in cardiomyopathy. In other words, compound heterozygosity, homozygosity, double heterozygosity, and even de novo mutations have been linked with more malignant phenotypes.[7-10]

In response to this explosion of genetic information, the American College of Medical Genetics (ACMG) released a document outlining guidelines for variant interpretation in 2015.[11] It recommends classifying variants by categorizing pathogenic or benign evidence, from supporting evidence (the lowest power) to very strong evidence (the highest power). This includes evidence from population, computational/predictive, functional, segregation, de novo, and allelic data. By identifying evidence for a variant, we are then able to categorize that variant as Pathogenic, Likely Pathogenic, Benign, Likely Benign, or of Uncertain Significance. Although this is highly useful in variant interpretation, much of the evidence may be interpreted subjectively and thus, this mode of interpretation will always be prone to bias.

Regardless, with this new wave of genetic data, clinicians and residents, such as myself, are still left puzzled. Are variants of undetermined significance truly permissible? Should we expect more malignant phenotypes in those with "pathologic" mutations? How can we, as clinicians, make any solid conclusions from genetic testing? I believe the answer is we cannot always make definite conclusions, but we may be able to make reasonable inferences based on recommendations. As such, we should move away from calling variants "mutations" and should make a strong effort to classify variants based on ACMG criteria, which should translate into clinical relevance. Also, taking advantage of data we currently have is essential. Genetic studies have given us insight into the importance and function of significant genes in the genome. Furthermore, those with more than one variation, radical variation, and non-inherited variation may be at an increased chance of disease. Although much of what is in the middle, in the form of post-translational modifications for instance, may always remain obscure, we can make reasonable predictions about how variation within vital locations in the gene may affect gene function. And, statistically, proven pathogenic mutations within vital domains will always more than likely present with a predictable phenotype.

Although I am still fairly early in my medical career, it has become evident to me that this is not an issue solely relevant to cardio-genomics. It is also an issue in every other pediatric subspecialty. Therefore, any genetic variation should not be immediately brushed aside. We must look at these findings as scientists, understanding that this variation may be causative, but also may likely be incidental. As a pediatrician, being able to predict a child's clinical future, in the setting of conditions as severe as sudden cardiac death and debilitating cardiomyopathy likely leading to end-stage heart failure as well as avoiding reassuring parents and preventing undue stress due to incidental variations found on a chromosomal microarray, may seem like a lofty goal at present, but I hope will be an achievable reality in the not too distant future.

REFERENCES

1. Brent LB, Aburano A, Fisher DL, Moran TJ, Myers JD, Taylor WJ. Familial muscular subaortic stenosis: an unrecognized form of "idiopathic heart diseases," with clinical and autopsy observations. *Circulation.* 1960;21:167-180.
2. Dweck D, Reynaldo DP, Pinto JR, Potter JD. A dilated cardiomyopathy troponin C mutation lowers contractile force by reducing strong myosin-actin binding. *The Journal of biological chemistry.* 2010;285(23):17371-17379.
3. Baryshnikova OK, Li MX, Sykes BD. Modulation of cardiac troponin C function by the cardiac-specific N-terminus of troponin I: influence of PKA phosphorylation and involvement in cardiomyopathies. *Journal of molecular biology.* 2008;375(3):735-751.

4. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv.* 2019:531210.
5. Jones EG, Landstrom AP. Determining the Likelihood of Variant Pathogenicity Using Amino Acid-level Signal-to-Noise Analysis of Genetic Variation. *Journal of visualized experiments : JoVE.* 2019(143).
6. Landstrom AP, Fernandez E, Rosenfeld JA, et al. Amino acid-level signal-to-noise analysis of incidentally identified variants in genes associated with long QT syndrome during pediatric whole exome sequencing reflects background genetic noise. *Heart rhythm.* 2018;15(7):1042-1050.
7. Tadros HJ, Life CS, Garcia G, et al. Meta-analysis of cardiomyopathy-associated variants in troponin genes identifies loci and intragenic hot spots that are associated with worse clinical outcomes. *Journal of molecular and cellular cardiology.* 2020;142:118-125.
8. Hodatsu A, Konno T, Hayashi K, et al. Compound heterozygosity deteriorates phenotypes of hypertrophic cardiomyopathy with founder MYBPC3 mutation: evidence from patients and zebrafish models. *American journal of physiology. Heart and circulatory physiology.* 2014;307(11):H1594-1604.
9. Wessels MW, Herkert JC, Frohn-Mulder IM, et al. Compound heterozygous or homozygous truncating MYBPC3 mutations cause lethal cardiomyopathy with features of noncompaction and septal defects. *European journal of human genetics : EJHG.* 2015;23(7):922-928.
10. Kiselev A, Vaz R, Knyazeva A, et al. De novo mutations in FLNC leading to early-onset restrictive cardiomyopathy and congenital myopathy. *Human mutation.* 2018;39(9):1161-1172.
11. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2015;17(5):405-424.