

Published in final edited form as:

Circ Cardiovasc Genet. 2012 December ; 5(6): . doi:10.1161/CIRCGENETICS.112.965152.

Action and the Actionability in Exome Variation

Calum A. MacRae, MD, PhD

Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Calum A. MacRae: camacrae@bics.bwh.harvard.edu

As next-generation sequencing emerges as a clinical tool, one of the greatest challenges facing human geneticists and clinicians is the interpretation of the vast numbers of novel variants that are uncovered in each exome or genome. The traditional evaluation of genetic variation in Mendelian disease has hinged on several widely accepted criteria for causality.¹ Before a variant can be considered potentially causal, it would normally be expected to abolish the synthesis of the encoded protein or to modify, through coding or splicing, amino acid residues of functional significance. Once this initial threshold has been attained, cosegregation with disease through a substantial number of meioses is expected. Functional significance is typically inferred from tight conservation across phyla or through specific biological assays. However, it may be difficult to relate effects in in vitro assays to the biological mechanisms responsible for disease.² Additional variants meeting these criteria in the same gene would be anticipated in other kindreds with the same phenotype. The case for causality would be further buttressed by the documentation of de novo mutations in the gene in sporadic cases and by the generation of genetically accurate animal models that recapitulate the original phenotype.³

These stringent requirements have been met to date for only a small minority of mutations annotated as definitive or high confidence in databases of human disease-related genetic variation.⁴ As the numbers of control sequences available for most genes has increased, it has become clear that many reported mutations in known disease genes are present as rare variants in normal individuals. It is now apparent that the scale and nature of interindividual variation have been substantially underestimated. Indeed, an early insight from exome sequencing has been the lack of trustworthiness of most existing human disease mutation catalogs.^{5,6} For some disease genes, all of the known mutations have been found in normal controls. How are we to build the knowledge base to interpret genomic variation? Will we be able to define what is actionable and what is not as we are inundated with genetic variation?

Even where mutations are rigorously validated and highly penetrant, the clinical outcomes in different families may vary substantially, so that using genetic data in a deterministic manner is likely to be challenging.⁷ The predictive use of a single genetic variant is largely a function of the strength of the correlation with a specific phenotype, so that often the genotype adds little as the clinician can only rely on it in the setting of high penetrance.⁸ As a result, genotypes are at present most useful to direct cascade screening in disorders with age-related penetrance. Until genotype brings along with it rigorous prognostic insight or direction to preclinical therapies, it will be difficult to add substantial value in a clinical context. For few if any causal genes in any disease area do we have sufficient information to use genotype to change management or therapy-the definition of the actionable variant.

Disclosures

None.

Much of this is a direct consequence of selection pressure. Traits with large effects on reproductive efficiency are likely to exhibit substantial buffering with consequent blunting of the relationship between genotype and phenotype.⁹ This buffering may be mediated by genetic or epigenetic factors, as well as through environmental effects on the final phenotype. The resolution of current clinical phenotypes also limits our ability to make meaningful genotype-phenotype correlations. We will need to move beyond traditional 18th- or 19th-century clinical disease entities as we begin to annotate the genome. If genetic testing is to prove worthwhile, we must generate information with rigorous predictive insights and consequences for management. Can the gap between our current knowledge and this ideal be achieved by capturing all of the genetic variation in the individual?

Recently published data from a limited number of genomes have defined the extent variation among individuals. Each human exome harbors as many as 130 to 400 rare nonsynonymous variants and around 100 true loss-of-function variants with as many as 20 genes completely inactivated.⁵ These studies have also suggested that common loss-of-function variants are unlikely to play a major role in common disease. In the current issue, Pan et al⁴ highlight the extent of variation in more than 5000 human exomes from the National Heart, Lung, and Blood Institute exome sequencing project. They note that most genetic variation is rare, with the vast majority of coding sequence variation resulting from variants that are present at allele frequencies of <1%. Interestingly, Pan et al document substantially lower rates of nonsense variation in genes associated with Mendelian disease (defined through the OMim database) than in genes without such an association. This distribution was even more extreme when cardiomyopathy genes were compared with the rest of the genome or with other Mendelian disease genes. Indeed, cardiomyopathy genes also exhibited substantially fewer missense and splice site variants, but contain modest excess of synonymous variants.

These exome sequence data offer a new baseline for the assessment of potentially pathological variants in the known human cardiomyopathy genes. The observations also suggest that these genes are remarkably intolerant of deleterious variation and so are under stringent purifying selection.⁴ Whether these genes are under more intense selection pressure than other genes that are necessary for early development is not clear. Importantly, the selection pressure restricting deleterious variation in these genes may not be operating on a phenotype that is mechanistically related to the pathophysiology of cardiomyopathy. As a consequence, the mere presence of a rare loss-of-function variant in a cardiomyopathy gene does not imply that it is causal for a cardiomyopathy. Indeed, Pan et al⁴ noted that 4 of 46 of their gold standard pathogenic variants were present in this population sample. As the authors note, the absence of relevant phenotypic data on the exome sequencing project cohort makes definitive conclusions difficult, but these estimates infer that many variants previously classified as truly causal are not in fact sufficient to result in a detectable cardiomyopathy.

In the majority of the known cardiomyopathy pathways, haploinsufficiency is not sufficient to cause disease, but rather specific gains of function are necessary. Although a gain of function may occur with truncated transcripts in some genes, for most cardiomyopathy genes both human and animal data suggest that true nulls are not likely to have a clinical phenotype. Missense variants and bona fide splice mutants are all well represented in the allelic spectrum of most cardiomyopathy genes, and it is now clear that it will require large exomic data sets to be able to parse signal from noise. The work of Pan et al⁴ elegantly demonstrates that the ab initio interpretation of sequence data will be fraught with problems. Yet without a rigorous mechanistic understanding of the basis of the disease, purely genetic criteria remain the most robust approach. The recognition that for some genes, there are effects at an RNA level that modify the final cellular state through effects on rates of translation, interaction with microRNAs or other mechanisms further complicates the

interpretation even of synonymous changes.¹⁰ Many factors modulate the rates of variation in genes, including the timing, duration, and other mechanics of transcription, translation, and posttranslational modification, which may correlate with other yet to be uncovered attributes of the raw primary sequence.¹¹ The increased frequency of synonymous variation in cardiomyopathy genes is difficult to explain, but may reflect positive selection at an RNA level on the basis of any number of these or other mechanisms. It is certain that we are still in the steep part of the learning curve when it comes to genome structure.

Pan et al⁴ also raise the possibility that their findings may implicate more complex patterns of inheritance, with the final phenotype in human cardiomyopathy resulting from a burden of variants in a number of genes within the disease pathway. This model is certainly reasonable for those forms of cardiomyopathy where penetrance is low and robust Mendelian transmission is not observed. However, in such situations, there are also competing hypotheses for the observed phenotypic variation, including substantial epigenetic or environmental influences, as well as stochastic processes. For the sarcomeric genes in particular, where hypertrophic cardiomyopathy is the predominant phenotype and robust, highly penetrant, autosomal dominant transmission is typical, a multi-gene variant burden is less readily reconciled with the clinical genetics. Clearly, there is significant pleiotropy both within and between families with hypertrophic cardiomyopathy. Such phenotypic variation is the rule rather than the exception in hypertrophic cardiomyopathy. In light of the data presented by Pan et al, it is unlikely to reflect the effects of common coding variation in cis or trans within the sarcomeric genes themselves. However, rigorous studies designed to define the genetic architecture of cardiomyopathy, and in particular the basis of any heritable modifiers, will be necessary to lay this question to rest.

As Pan et al highlight, the availability of whole exome and whole genome data reveal a scale of variation and unsuspected confounding that we had imagined only vaguely in the past. Although the rationale for comprehensive sequence data are clear, there may not be enough individuals on the planet to fully deconvolute all of the variation present even in a single exome. We will have to begin to grapple with new ways of thinking about the relationship between genotype and phenotype. A focus on pathways rather than on single genes is rational, but understanding the state of a given pathway in a given individual may not be possible from primary sequence data alone. The intrinsic power of genetics is rooted in the cosegregation of genotypes and phenotypes. If we are to realize the full potential of comprehensive genetic information, we must begin to expand our phenotypic repertoire in a less biased manner, and a systematic effort to develop pathway centric clinical assays may be a reasonable place to start.

Acknowledgments

Sources of Funding

Dr MacRae is supported by the NIH (HG006500 and AG037925).

References

1. Borecki IB, Province MA. Genetic and genomic discovery using family studies. *Circulation*. 2008; 118:1057–1063. [PubMed: 18765388]
2. Casimiro MC, Knollmann BC, Yamoah EN, Nie L, Vary JC Jr, Sirenko SG, et al. Targeted point mutagenesis of mouse *Kcnq1*: phenotypic analysis of mice with point mutations that cause Romano-Ward syndrome in humans. *Genomics*. 2004; 84:555–564. [PubMed: 15498462]
3. Ahmad F, Seidman JG, Seidman CE. The genetic basis for cardiac remodeling. *Annu Rev Genomics Hum Genet*. 2005; 6:185–216. [PubMed: 16124859]

4. Pan S, Caleshu CA, Dunn KE, Foti MJ, Moran MK, Soyinka O, et al. Cardiac structural and sarcomeric genes associated with cardiomyopathy exhibit marked intolerance of genetic variation. *Circ Cardiovasc Genet*. 2012; 5:602–610. [PubMed: 23074333]
5. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. 1000 Genomes Project Consortium. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012; 335:823–828. [PubMed: 22344438]
6. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. [PubMed: 23128226]
7. Ashrafian H, McKenna WJ, Watkins H. Disease pathways and novel therapeutic targets in hypertrophic cardiomyopathy. *Circ Res*. 2011; 109:86–96. [PubMed: 21700950]
8. MacRae CA. Closer look at genetic testing in long-QT syndrome: will DNA diagnostics ever be enough? *Circulation*. 2009; 120:1745–1748. [PubMed: 19841296]
9. Hartman JLT, Garvik B, Hartwell L. Principles for the buffering of genetic variation. *Science*. 2001; 291:1001–1004. [PubMed: 11232561]
10. Bartoszewski RA, Jablonsky M, Bartoszewski S, Stevenson L, Dai Q, Kappes J, et al. A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem*. 2010; 285:28741–28748. [PubMed: 20628052]
11. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011; 146:353–358. [PubMed: 21802130]