



## BASIC SCIENCE FOR THE CHEST PHYSICIAN

## Basic science for the chest physician: interpretation of genetic variants

Patrick R Sosnay,<sup>1,2</sup> Garry R Cutting<sup>1,3</sup>

<sup>1</sup>McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA

<sup>2</sup>Division of Pulmonary & Critical Care Medicine, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

<sup>3</sup>Department of Pediatrics, Johns Hopkins University, Baltimore, MD, USA

**Correspondence to**

Garry Cutting, The Johns Hopkins Medical Institutions, 733 North Broadway, MRB 559, Baltimore, MD 21205; USA; gcutting@jhmi.edu

Received 22 November 2013

Accepted 27 November 2013

**ABSTRACT**

Sequencing of the human genome and introduction of clinical next-generation sequencing enable discovery of all DNA variants carried by an individual. Variants may be solely responsible for disease, may contribute to disease, or may have no influence on the development of disease. Interpreting the effect of these variants upon disease is a major challenge for medicine. Although the process is still evolving, certain methods are useful in discriminating the effect of variants upon phenotype. These methods have been employed to the greatest extent in Mendelian disorders where deleterious changes in one gene can cause disease. Here, we briefly review the relative merits of these methods, with emphasis on using a comprehensive approach modelled after the analysis of variants that causes cystic fibrosis.

Greater understanding of the influence of variation in our genome upon health and disease will help usher in the era of individualised medicine. The aetiology of common diseases is complex, in that multiple genetic and environmental risk factors combine to cause a distinct phenotype. Genome-wide association studies have identified DNA variants in numerous locations that confer risk for common pulmonary disorders, such as asthma and COPD. The mechanism by which variants cause common diseases is generally unknown. However, rare families manifesting a common disease inherited in a Mendelian fashion have facilitated the identification of genes bearing variants of high functional impact. Examples include variants in *BMPR2* that cause pulmonary arterial hypertension and variants in the promoter of *MUC5B* in patients with pulmonary fibrosis.<sup>1,2</sup> Thus, Mendelian, or so-called 'single gene', disorders provide an unparalleled opportunity to find genes that have been substantially modified by a variant in their DNA sequence. Unfortunately, the genome contains many variants that occur in or near genes, and only some of which change gene function sufficiently to cause disease. Some variants alter function in a manner that produces mild or incomplete forms of disease, while other variants cause no discernable change in phenotype. Furthermore, a variant may be an innocent hitchhiker with a pathogenic variant elsewhere in the same gene, or it may combine with other variants in the same gene to cause disease. Thus, assessment of the disease liability of a variant requires an understanding of its effect upon gene, cellular and organ function, and the genetic context in which it occurs. Elucidating the pathologic potential of variants and their relative contribution to

phenotype will provide critical insight into disease mechanisms and opportunities for intervention.

What approaches can be used to interpret the consequences of genetic variants? If the actual frequency that a variant occurs in those affected and unaffected by the disease were known, one could calculate the likelihood that a variant causes disease when present (ie, penetrance). Unfortunately, the frequency of variants across broad populations is not known, therefore, other lines of evidence must be employed to determine the extent that a variant causes disease. The usability of each method varies across diseases with different modes of inheritance, but each approach has a degree of usefulness for all single gene disorders. Each of these methods has potential shortcomings, so a strategy combining multiple modalities is proposed (figure 1), as recently demonstrated in a study of the Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*) gene.<sup>3</sup>

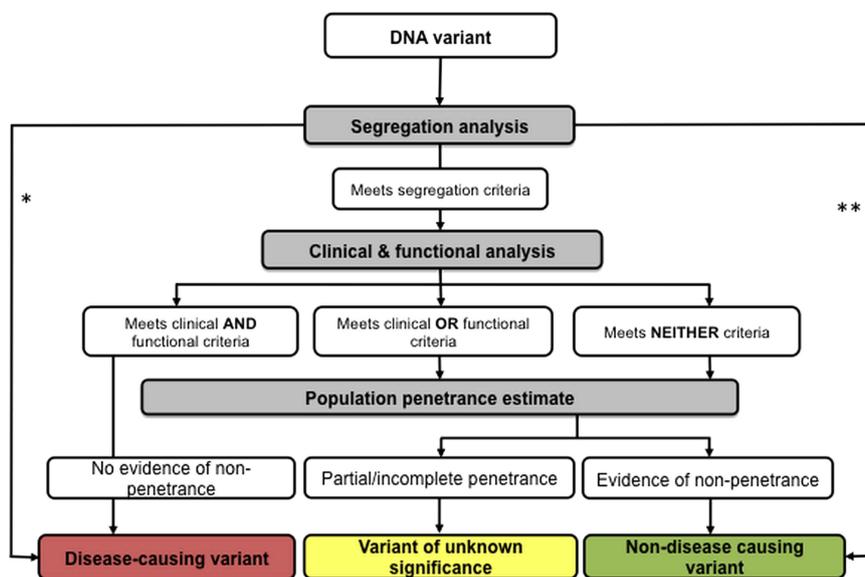
► Segregation analysis of a family pedigree with individuals in multiple generations that have genotype and phenotype information can be used to establish if variants segregate with disease within a family; that is, variants should be found in affected family members, but not in unaffected family members. The expected segregation of a variant that causes disease varies for different inheritance patterns. Variants found in an affected individual that occurred *de novo* in a gene previously associated with the disease (spontaneous changes that are not present in either unaffected parent) are highly likely to be disease-causing. Spontaneous mutations are most often observed as causes of disease with dominant, codominant, or X-linked inheritance patterns. Pedigree analysis indicating a lack of segregation with disease (such as identifying a variant in some, but not all, affected family members) is strong evidence that a given variant is not disease-causing. If the analysis of segregation in a pedigree is consistent with the known mode of inheritance it is supportive, but not conclusive, that the variant causes the disease. In this case, or in the case that an informative pedigree is not available, further evaluation of the mutation is required.

► Clinical presentation can be quantified and standardised to establish the parameters of the genotype-phenotype relationship. Specific phenotype information is needed to ensure all individuals presumed to have the disease meet uniform diagnostic criteria. Ideal measures, regardless of whether dichotomous (eg, a positive or negative methacholine challenge), or continuous (pulmonary artery pressure), are traits



CrossMark

**To cite:** Sosnay PR, Cutting GR. *Thorax* Published Online First: [please include Day Month Year] doi:10.1136/thoraxjnl-2013-204903



**Figure 1** Framework for combining multiple lines evidence to characterise DNA variants. If data are available, segregation analysis should be performed to assess whether the inheritance of the variant is consistent with the transmission pattern of disease in families. Objective measures of the clinical phenotype of individuals that carry the variant (clinical criteria) and experimentally tested gene function (functional criteria) can be used to assign disease-causing status to a variant. If a variant does not meet either clinical or functional criteria, penetrance analysis can be used to ascertain whether the variant is not disease-causing. \*Variants caused by de novo mutation found only in affected individuals with dominant or X-linked conditions can be presumed to be disease-causing. \*\*Variants that do not segregate with the disease can be ruled out as disease-causing, provided that confounding factors, such as incomplete penetrance have been ruled out.

that are known to be highly influenced by the gene under study. For example, sweat chloride concentration quantifies the effect of *CFTR* variants to a greater degree than lung function, because sweat chloride has a greater correlation with *CFTR* function than lung function.<sup>3</sup>

- ▶ Functional assessment establishes the biologic plausibility for how a change in gene function results in disease. Variants can disrupt gene function by altering efficiency of transcription, RNA splicing, protein processing or function. Loss of function is generally observed in recessive disorders or for dominant disorders caused by the loss of one of two working copies of the gene (ie, haploinsufficiency). In loss of function disorders, there is consensus that almost all variants resulting in a premature termination codon (eg, nonsense, frame shift) are highly likely to be pathogenic.<sup>4</sup> Variants that cause the substitution of an amino acid require experimental determination that function is lost. The same method of experimental determination can be used for dominant disorders caused by gain of function effect of the protein product. In these assessments, a threshold that determines the level of function necessary for disease must be established through testing of previously well-characterised variants or extrapolated from other research (eg, animal models). Because functional studies require time and resources, considerable effort has been expended in the development of bioinformatic predictors that use protein structure information and/or common ancestral sequences to make an ‘in silico’ prediction of the effect of a variation.<sup>5</sup> Computational methods have been widely employed experimentally, but are not yet a substitute for direct experimental evaluation.
- ▶ A penetrance analysis can be performed to assess whether a variant *does not* cause disease. Essentially, one searches for the presence of a variant in a ‘healthy’ gene. For dominant disorders, genes in individuals confirmed to not have disease from the general population can be studied. For recessive disorders, two deleterious variants need to be present to cause disease. Thus, obligate heterozygotes who carry a deleterious

gene and a ‘healthy’ gene that was not passed to the offspring (such as the confirmed unaffected parents of affected offspring) can be informative, as any variant seen in the ‘healthy’ gene can be presumed to not cause disease. Highly rare variants pose a major challenge for penetrance assessment, as very large numbers of unaffected individuals are needed for analysis.

For all modes of inheritance, variants that meet segregation, clinical and functional criteria can be considered disease-causing. All diseases, even those associated with Mendelian inheritance, are comprised of traits that are subject to modification from other genes and from the environment. Therefore, it is not unexpected that some variants not meeting criteria may be neither pathogenic nor neutral, but capable of causing disease under certain circumstances (ie, variably penetrant), or capable of causing a partial form of the disease. Alternatively, some variants will be indeterminate due to absence of, or inconclusive, evidence. As an example of the usability, but also the challenges of widespread use of the interpretation of variants, a recent publication in *Thorax* that examined individuals with single organ system manifestations of cystic fibrosis (CF) within the spectrum of *CFTR* dysfunction.<sup>6</sup> As expected, only a minor fraction of individuals with incomplete CF carried two CF-causing variants. Although it may be desirable to have discrete categories of pathogenic versus neutral variants, for some disorders such as atypical forms of CF, clear demarcation between the two groups may not exist.

As genetic analysis becomes less expensive and more efficient, there will be greater opportunity to examine the genetic contributions to disease. As noted, determining the consequence of individual variants is not straightforward, even for Mendelian disorders such as CF.<sup>3</sup> Multiple modalities are needed to examine the effects of any given variant, but this divining is a necessary step to widespread use of genetic information in diagnosis and prognosis; as well as to implementing therapeutics based on DNA variation.

**Acknowledgements** The authors would like to thank Karen Siklosi Raraigh for her review of the manuscript.

**Contributors** PRS and GRC both wrote the article.

**Competing interests** None.

**Provenance and peer review** Commissioned; internally peer reviewed.

## REFERENCES

- 1 Lane KB, Machado RD, Pauculo MW, *et al*; International PPH Consortium. Heterozygous germline mutations in BMPR2, encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. *Nat Genet* 2000;26:81–4.
- 2 Seibold MA, Wise AL, Speer MP, *et al*. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–12.
- 3 Sosnay PR, Siklosi KR, VanGoor F, *et al*. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 2013;45:1160–7.
- 4 Richards CS, Bale S, Bellissimo DB, *et al*. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 2008;10:294–300.
- 5 Cline MS, Karchin R. Using bioinformatics to predict the functional impact of SNVs. *Bioinformatics* 2011;27:441–8.
- 6 Ooi CY, Durpui A, Ellis L, *et al*. Does extensive genotyping and nasal potential difference testing clarify the diagnosis of cystic fibrosis among patients with single-organ manifestations of cystic fibrosis? *Thorax* Published Online First: 21 Oct 2013. doi:10.1136/thoraxjnl-2013-203832