Sequencing of transporter genes in cholestasis: we are still learning

Richard J Thompson

Institute of Liver Studies
King’s College London
London
UK

Correspondence:
Postal Institute of Liver Studies
King’s College Hospital
Denmark Hill
London
SE5 9RS
UK
Telephone +442032994296
Email richard.j.thompson@kcl.ac.uk

Grant support
RJT received support from NIH R01 DK094828

Conflicts of interest
RJT received support from Shire, Albireo, Alexion, Arcturus, GSK and Retrophin. None are however relevant to the content of this manuscript.

Sequencing of DNA has never been easier. It is now perfectly possible to sequence an individual’s entire genome from a few micrograms of DNA, obtained from peripheral blood or elsewhere. The cost of sequencing has also fallen dramatically. The biggest hurdle is now in making sense of the data. Variation in DNA sequence contributes a large part of the phenotypic differences between individuals. However most of the variation occurs in non-coding DNA; the overwhelming majority of which we cannot yet interpret. Variation within genes is much better understood; but even here not everything is as simple as it might seem.

“Simple” genetics might consider highly penetrant alleles, causing Mendelian phenotypes, be they dominant or recessive. That is to say, that those with a given genotype have highly predictable manifestations. At the other end of the spectrum are relatively common variants that have been associated with particular diseases. Win the latter case many unaffected individuals will have the risk allele; although it is more common in those with the disease phenotype. In between these extremes are variants which are identified as making intermediate contributions to the phenotype in question. This later group might be considered “contributory”, “predisposing” or even “modifiers” of a condition. In truth however, many variants contribute to every disease, it is just the extent of their contribution that varies. There are probably no fully penetrant alleles, or indeed any discrete categories of contribution. Worse still, because of the effect of other variants, and non-genetic factors, one variant might contribute more in one individual than in another.
All of this makes the interpretation of genetic variants surprisingly complex. There has been a rapid growth in websites and software designed to help us interpret genetic variation. There have been guidelines, most notably from the American College of Medical Genetics and Genomics (ACMG) (1), to help us. The software tools are only as good as the information available to them; this is mostly from the public domain, though some have their own databases. The ACMG guidelines are now widely employed; though they probably increase consistency, more than improving accuracy. To date all these tools have still been directed at highly penetrant or Mendelian models of disease.

We understand variation in some genes better than others. The liver disease gene with the best understood variation is probably HFE, underlying genetic haemochromatosis. The contributions to disease of the three major variants have been very well characterised (2). The risks associated with all combinations of these alleles has been defined. However this doesn’t predict those who will get haemochromatosis; only those with increased risk. In more recent times a common variant in PNPLA3 has been associated with both NAFLD and alcohol-associated liver disease. However in this case the penetrance and the contribution to disease are so low that it is of little use in making a diagnosis, though it may be of use in predicting response to treatment (3).

A growing list of genes have been associated with cholestatic liver disease. It is now 19 years since the first three were described (4-6). This means that considerable numbers of patients have been sequenced for these genes, and some prediction of the contribution of genetic variants can be made. Merely collecting genetic data is of little value however, it is only when this is combined with high quality phenotypic data can we begin to learn. In this issue Keitel and colleagues (7) describe 154 variations in ATP8B1, ABCB11 and ABCB4 in 427 individuals, all suspected of having genetic cholestasis. In this cohort of patients several previously described disease-causing variants were found. Several novel putatively disease-causing variants were identified. These are important additions to the ever-growing catalogue of variants for which there is good evidence of major pathological effect. Protein truncating mutations form the minority, though their pathogenicity is generally clear. Consensus splice site mutations, affecting the first or last 2 bases of an intron are also generally easy to classify as detrimental. The current difficulty really lies with missense changes. Such variants are frequently thought of as changes in the amino acid sequences. In truth they are alterations in genomic DNA. If transcription occurs normally they become changes in pre-mRNA. As such they may affect splicing and processing into mRNA. Only if these steps are overcome can such changes become variants in the polypeptide (8).

The molecular consequence of variants in the amino acid sequence of proteins can be analysed in several ways. The most powerful evidence in favour of pathogenicity is the observation of such a change, in trans, with another known pathogenic change. All other interpretations of function are open to criticism. Commonly applied observations include the degree of evolutionary conservation of the same protein, and between similar proteins, at both the nucleotide and amino acid level. The functional change predicted by in silico models can be helpful, but the tools available still lack enough data. In vitro models can be used to look at all the stages of mRNA and protein production, and function. However no in vitro experiment is perfect, yet! Keitel and colleagues combine the widely available
predictive tools, along with less used computer modelling. Future patients will tell us which have worked best, and the determined pathogenicity of known variants will increase or decrease accordingly.

Previous studies have shown that common variants in these genes are over represented in patients with “acquired” cholestasis. Of course such patients probably harbour variants which impair the quantity or function of the gene/protein to a lesser degree than the better described “disease-causing” mutations. The idea that “disease-predisposing” or “modifier” variants are in fact similar is highlighted by data in the current paper. Some common variants appear to be more common than is seen in the general population; in patients with no other mutations, where they would be predisposing or in patients with pathogenic mutations, where they would be modifiers. In all experiments examining allele frequencies the biggest problem is control data. The bigger, and better matched, the control set is the more credible are any differences in frequency between them and patients. The data in Keitel’s paper compare the observed frequencies with those seen in the Gnomad database (9). This is currently the largest publically-available set of allele frequencies. It is in fact an aggregation of data from other projects, and includes information from 138,000 individuals. The presence of a variant in such a large dataset does not exclude it from being highly pathogenic, as many are very rare. On the other hand many variants have very different frequencies in different populations. The data presented in Gnomad are broken down into 7 large groups and “others”. Brief inspection will show that some populations have been much more extensively investigated than others; e.g. Ashkenazi Jews and Finns. Europeans make up almost half the chromosomes sequenced, though we do now have much better data on individuals from other regions of the world than we had even one year ago. Keitel compares the overserved frequencies with those seen in the entire Gnomad dataset, and with those seen in Europeans, as that was thought to be the origin of most of their patients. As the quantity and quality of control data grows this and previous analyses, based on comparison with controls, will need re-examination.

DNA sequencing is now increasingly cheap and easy. Data analysis however remains complex. Variants in genes should be thought of as contributing to the phenotype; the most difficult decision in to what extent. The sequencing being carried out today, be it in patients and controls, will help build the datasets that will inform future research and, more importantly, diagnosis.

References


