INTERPRETING GENETIC VARIANTS: ENABLING FINE-SCALE GENOME MAPPING AND PRECISION MEDICINE

Dr Ian M Frayling MA MB BChir PhD FRCPath Hon.FFPath(RCPI) FEBLM
Honorary Consulting Genetic Pathologist to St Mark's Hospital, London & St Vincent's Hospital, Dublin
Honorary Senior Clinical Research Fellow, Inherited Tumour Syndromes Research Group, Cardiff University
President Elect, Association of Clinical Pathologists
Member of Council & Honorary Treasurer, InSiGHT and UK Cancer Genetics Group
Honorary Adviser to Lynch Syndrome UK and Lynch Syndrome Ireland

The human genome consists of 3 billion base-pairs (bp) - letters of DNA code. An individual human may differ from the reference genome at more than 5 million of these, as well as larger variants from 2 bp to, in the case of an individual with Down syndrome, a whole extra chromosome of 48 million bp. Deciding which, if any, of these variants is clinically significant is perhaps the greatest challenge facing genomic medicine now that obtaining the sequence of an individual's genome is easy.

Twenty years ago it could take months to analyse one gene of perhaps 2,000 bp. If a variant was found it might then take the same amount of time to establish it was pathogenic, i.e. causative of disease. At this time testing of a particular gene might only be performed in one or two laboratories worldwide, for reasons of local expertise in testing and interpretation, but this has changed. It took 13 years to establish the first draft of the reference human genome sequence, published in April 2003, but such are the advances in technology that an individual's genome can now be sequenced in two days. In terms of DNA sequencing any gene can now be sequenced in any genetics laboratory, but the local expertise in interpretation remains, usually based on local research interests. However, this too is changing as it is realised that national and international pooling of data and expertise are required to interpret variants. No longer can, or indeed should variant interpretation be left up to individuals, because this is a major factor in disparity of interpretation between centres, leading to upsetting and potentially dangerous confusion to both clinicians and patients. Families span borders, so the interpretation of a variant shared across them should not be disparate. It also improves quality and thus safety if interpretation is carried out by single expert multidisciplinary groups, which are increasingly becoming international entities as the science of interpretation matures and becomes a recognised entity. [1, 2, 3]

MUTATIONS AND VARIANTS

Variants result from mutations in genes and, because genes encode proteins ultimately this results in protein variants. For example, blood groups B and O are variants of group A. One variant is referred to as a genotype, but the sum total of all an individual's variants may also be referred to as their genotype. Mutations are caused when mistakes are made in DNA replication. This is commonly due to DNA damage caused by endogenous chemicals, such as free radicals produced during normal metabolism, or exogenous chemicals such as those found in tobacco. In addition, although DNA replication is robust, it not entirely error-free, because of the very nature of the chemical structure of DNA. Hence, there is a whole army of DNA repair enzymes designed by evolution to minimise, but not eliminate mutations. If there were no mutations, there would be no variety, and without variety a species cannot evolve and ultimately it becomes extinct. If a mutation occurs in the *germline*, the specialised cells that produce eggs and sperm, then that variant may be passed on to offspring and become heritable, responsible for a variant in all cells of the body – a *constitutional* variant. If a mutation occurs in another cell of the body, a *somatic* mutation, then it may ultimately lead to a cancer. Twenty years ago the vast majority of clinical genetic tests were for possible inherited constitutional variants, but now genetic tests for somatic and constitutional variants related to cancer are in the majority.

An expressed characteristic, a phenotype, is a function of

genotype and environment. Genotypes work in combination with the environment to cause phenotypic variety: e.g. skin colour, height, and susceptibility to disease. As the environment of any one gene includes the other 20,000 genes and their variants so we get to genomics, the sum totality of all genes and variants acting in concert with the environment they all find themselves in. This may seem dauntingly complex, but it is no more complex than the cracking of the Enigma code 80 years ago, and we have the benefit of modern computers spawned by that challenge.

Phenotypic variation is acted upon by selective pressures in the environment in the fight for survival. A variant that enables an individual to produce more offspring will result in it becoming more common and *vice versa*. The G614 variant in the spike protein of SARS-CoV-2 is replacing the original D614 because it confers greater infectivity. This is Darwinian evolution acting as I write: perpetual work in progress. If a variant occurs in a somatic cell and confers a growth advantage then that may lead to a tumour and eventually cancer, usually after other mutations have occurred. It should be noted that genetically, a population is defined by the frequencies of its genetic variants.

THE IMPORTANCE OF GETTING IT RIGHT

It is said that "Any idiot can find a mutation, but only wise men can interpret them." So, not only can the consequences of a variant be severe, but so can misinterpretation. On top of this, indecision and delay affecting timeliness can also have adverse clinical effects. There is less value in interpreting a variant after a patient has succumbed to their disease, although their relatives may benefit.

There are two major parameters to consider in variant interpretation: how certain is it that it is causative of disease, its *pathogenicity*, and to what degree does it cause disease, which bears on *actionability*. These two factors are independent, but often confused. Close study of blood groups in the 1950s and since have established beyond all doubt that individuals with group A are at a greater risk of stomach cancer than group O. [4] However, the degree of risk, the Hazard Ratio, is only ◊ 1.13, so nowhere near clinical actionability (>2). In addition, as with any test, interpretation and actionability depends on the clinical context, considering the whole person. The problem is that not all variants can be assigned simply as pathogenic or not pathogenic: many lie somewhere in between, so-called "variants of uncertain significance" or VUS. In part this is often a function of the degree to which they cause disease. A weakly acting variant needs more data to be sufficiently certain as to its effect, a recognised issue in the science of variant interpretation.

HOW TO GET IT RIGHT

Many lines of evidence can be used in interpretation. A mutation that predicts cutting a protein short is generally pathogenic, but one which changes one amino acid out of a thousand or more may well not be. How a variant tracks with disease in a family is useful, as can specific clinical tests. Subtle signs, however, may only be obvious to a trained and skilled clinical eye, hence expert phenotyping by medically-trained individuals is an absolute necessity – an example of getting a sensible answer if you ask a sensible question. Clinical tests, including genetic tests, must be performed and interpreted under the direction of a medical practitioner, as the Academy of Medical Royal Colleges has advised. [5]

EXAMPLE 1: PREDISPOSITION TO CANCER

Lynch syndrome (LS) is a common condition potentially affecting up to 1/100 of the population. [6] It predisposes to bowel, womb and many other cancers from a young age, with a lifetime risk of cancer up to 85%. [6, 7] NICE guidance is that all bowel cancers should be tested to see if LS is the cause.[8] A patient is found to have the variant MLH1 c.306G>T p.(E102D), but a Genetic Counsellor sees that despite 17 entries on the international reference database it is classed as a VUS by the Variant Classification Expert Panel and seeks clarification. [9, 10, 11, 12]

The variant has a 72% (~3:1 on) chance of being pathogenic from considerations of evolution conservation, but >95% certainty (19 to 1 on) is required for clinical purposes. [12, 13] Fortuitously, UK data very recently analysed by Dr Fiona McRonald's team at Public Health England (PHE) can be used, as gathered by PHE through the National Cancer Registration & Analysis Service (NCRAS) under Section 251 of the NHS Act (2006). [14] This shows the variant has been seen in 6 of 2041 individuals with LS-type cancers in the UK, but in none of 113,654 non-Finnish Northern Europeans unaffected by cancer. [15] This gives a Hazard Ratio of 57 (CI: 1.1 – 2,800) and a probability, if random, of 0.000 000 003%. Factoring in the starting probability of 72% gives a final probability of pathogenicity of 99.999 999 998 8% (83 billion to 1 on). Hence, it can be concluded beyond all reasonable doubt that this variant is pathogenic. As a result relatives may be tested for it to see if they warrant regular surveillance for cancer or prophylactic surgery, as may the other 23 families with this variant, plus those yet to be diagnosed with it. This illustrates the enduring utility of an international expert group using data made publicly available and interpreted according to peer-reviewed criteria, based on objective probability, not subjective opinion.

EXAMPLE 2: PHARMACOGENETICS

2012: A 52y old NHS Consultant is prescribed an selective serotonin re-uptake inhibitor (SSRI) antidepressant, but suffers a severe adverse drug reaction (serotonin syndrome toxicity) and unable to tolerate treatment has to take 2 years off work. Testing for the genes CYP2C19 and CYP2D6 that produce the enzymes responsible for metabolising SSRIs is unavailable in the UK. A Dutch laboratory is able to report absent CYP2D6, but also variant CYP2C19*17 conferring high activity, inconsistent with the phenotype. 2018: An expert from the Karolinska Institute is consulted and reveals it has since been discovered that a rare novel loss-of-function variant CYP2C19*4B is composed of both *17, a controlling region variant which turns the gene on, and *4, a variant at the start codon which prevents protein production (Figure 1). [16] As the earlier pre-2012 assay was unable to detect *4B, because it could not detect *4 in the presence of *17, the mystery is resolved.

Fortunately, if the patient should suffer depression in the future they can be treated with Vilazodone, a new SSRI introduced after 2012 due to the high prevalence of side-effects of conventional SSRIs due to the high prevalence of low activity variants of CYP2C19 in persons of BAME origin (20-40% vs 2% in Northern Europeans). In addition, as the patient will not activate Clopidogrel they can be warned that they must be given an alternative anticoagulant if they should require a cardiac stent, to avoid death from iatrogenic thrombosis. [17] They

Figure 1

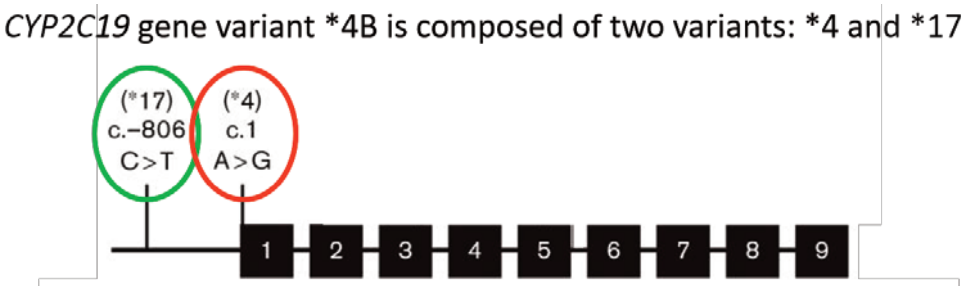


Illustration of the *CYP2C19* gene, showing variant *4B is composed of both *17, a gain-of-function variant in the controlling region (highlighted in green), and *4 (red), a loss-of-function variant which abrogates the start codon so preventing protein production. Thus, *4B is a loss-of-function variant.

can also be advised that as they have zero CYP2D6 activity then they require smaller doses of opiates and proton pump inhibitors (PPI), minimising side effects from those medicines as well. In contrast, persons of BAME origin frequently have high CYP2D6 activity, and so are *inter alia* at risk of inadequate analgesia when treated with opiates.

This illustrates how variants can act in concert, interpretation can change with time as techniques improve and knowledge is acquired, and knowledge of population-specific variant frequencies is needed for precision medicine, all confirming: "You can get your genome for less than $1000, but you need a $100 million Institute to interpret it." Severe adverse drug reactions are common: everyone is waiting for the drug that will affect them. They are dangerous, cause considerable morbidity and mortality, poor compliance with treatment and incur great expense to the NHS. However, they can be foreseen and avoided by genetic testing, but knowledge, availability and application in the NHS is so far minimal.

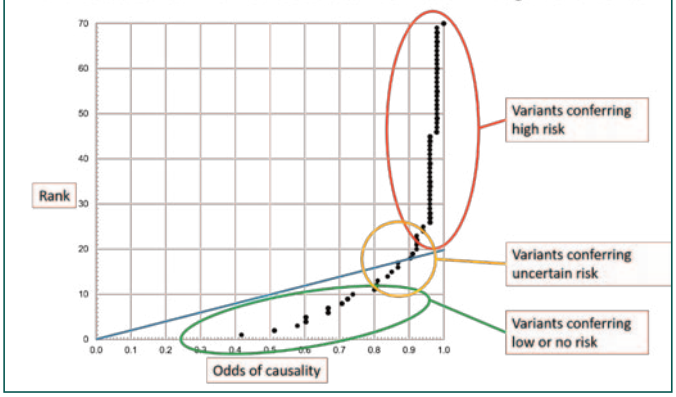
EXAMPLE 3: NEUROFIBROMATOSIS TYPE 1 AND BREAST CANCER RISK

Neurofibromatosis type 1 is a disfiguring and disabling condition that affects 1/3000 individuals, predisposing to multiple benign and malignant tumours. A raised risk of breast cancer (NF1-BC) is seen, often young onset and more malignant, but breast screening is controversial. Is there a genotype-phenotype correlation that could guide management? Gathering data on *NF1* gene variants seen in patients with NF1-BC reveals that a minority of variants (probably those conferring a gain of function) predispose strongly, whereas one variant seen commonly (whole gene deletion) which causes total loss of protein function is not seen in NF1-BC, revealing variant heterogeneity (Figure 2). [18] This illustrates the hazards of generalising variant effects at the level of a gene, but at the level of individual variants the fine detail may guide precision medicine.

CONCLUSIONS

Gone are the days of individuals working alone. The interpretation of genetic variants requires large multidisciplinary groups working on freely-available data, contributed to by all for the greater good. Patient groups support this approach.[19] However, this necessitates provision for and support *in perpetuum* of large inter/national multidisciplinary expert groups and their associated databases, such as the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) initiative (which is currently only supported by charitable and research funding from abroad) and CanVIG-UK. [20, 21]

Figure 2



References

1. https://www.hgvs.org/
2. https://clinicalgenome.org/
3. https://www.ncbi.nlm.nih.gov/clinvar/
4. Aird I, Bentall HH, Roberts JAF. (1953) Relationship Between Cancer of Stomach and the ABO Blood Groups. *British Medical Journal.* 1(4814):799-801.
5. http://www.aomrc.org.uk/statements/genomic-medicine-in-the-nhs/
6. Frayling IM, Frankel WL, Arends MJ, Nagtegaal I, Tomlinson IPM, Morreau H, Brosens Lodewijk AA. (2018) WHO Classification of Tumours 5th Edition – Tumours of the Digestive System. Chapter 15, Section 1.1.3 Lynch syndrome, pp.515-521. Eds: Carneiro F, Arends MJ. International Agency for Research on Cancer (IARC), Lyon, FR.
7. https://www.insight-group.org/
8. https://www.nice.org.uk/guidance/dg27
9. http://insight-database.org/
10. https://www.ncbi.nlm.nih.gov/clinvar/submitters/500189/
11. https://www.insight-group.org/criteria/
12. Thompson, BA., et al. 92014) "Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database." *Nature Genetics* **46.2**:107-115.
13. Thompson, BA., et al. (2013) "Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions." *Human Mutation* **34**: 255-265.
14. https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras
15. https://gnomad.broadinstitute.org/
16. Scott, SA., et al. (2012) "PharmGKB summary: very important pharmacogene information for cytochrome P450, family 2, subfamily C, polypeptide 19." *Pharmacogenetics and Genomics* **22**.2:159. https://europepmc.org/article/med/22027650
17. https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/fda-drug-safety-communication-reduced-effectiveness-plavix-clopidogrel-patients-who-are-poor
18. Frayling, IM., et al. (2019) "Breast cancer risk in neurofibromatosis type 1 is a function of the type of *NF1* gene mutation: a new genotype-phenotype correlation." *Journal of Medical Genetics* **56**: 209-219.
19. https://www.lynch-syndrome-uk.org/
20. https://www.insight-group.org/variants/databases/
21. Garrett, A., et al. (2020) "Cancer Variant Interpretation Group UK (CanVIG-UK): an exemplar national subspecialty multidisciplinary network." *Journal of Medical Genetics* Published Online First: 13 March 2020. doi: 10.1136/jmedgenet-2019-106759 https://www.cangene-canvaruk.org/canviguk