

Which differences make us different?

To discover what makes us different genetically, including our genetic predisposition to disease, the first major challenge is to find individual variations in our DNA sequence. Although sequencing the complete genome of one person can be done, it is still a mammoth task. One strategy to reduce the amount of sequencing involved is to study variation across the genome of an individual, but focus on just the protein-coding portion. A recent paper that describes such an approach highlights challenges for the development of personalized genomic research.

The authors reasoned that looking at all the exons of the genome — the 'exome' — could be especially informative, as this is where the majority of mutations that cause Mendelian

diseases are found.

If a large proportion of human functional variation lies within the exome, sequencing exons could become the focus for developing individualized

medicine — so a key component of the study was to explore which types of exonic genetic variation are most likely to have phenotypic effects.

By comparing an individual exome sequence with the NCBI reference sequence, 12,500 coding variants that affect protein sequence were found — the majority of which were non-synonymous SNPs (nsSNPs). The authors used an algorithm to predict which of these nsSNPs are most likely to alter protein function based on, for example, the type of amino-acid change they cause and where they lie in the protein. This analysis predicted that ~1,500 nsSNPs (14% of the total) would have deleterious effects on proteins, confirming previous estimates made from studies of smaller numbers of genes. However, as these apparently important nsSNPs are rare variants, the researchers point out that it will be challenging to correlate them with phenotypes.

In addition to SNPs, the exome contained 739 coding insertions and deletions (indels). These were enriched for indels with a base-pair length that is divisible by three, presumably because

these are less deleterious in a coding region as they cause an insertion or a deletion in the amino-acid sequence, rather than a frameshift. Many of the indels are located at exon boundaries or protein termini, and of the remainder those that cause frameshifts tend to be in hypothetical proteins, suggesting overall that a substantial fraction of coding indels are actually functionally neutral. This detailed indel sequencing also resulted in some corrections to the NCBI reference sequence, for example, by correcting small introns.

This exome study highlights how long lists of coding variants from genomic studies need to be filtered to extract those that are most likely to have effects on protein function. In terms of how the field of personalized genomics can move forward, this work also raises many questions, such as how can the phenotypic consequences of rare alleles be identified, and what are the relative contributions of coding and non-coding variants to disease susceptibility?

Mary Muers

ORIGINAL RESEARCH PAPER Ng, P. C. *et al.*
Genetic variation in an individual human exome.
PLoS Genet. 4, e1000160 (2008)

