

 GENOMICS

## Getting personal and regional

A major challenge of personal genomics is identifying the most medically relevant information from the wealth of data in the genome sequences of individuals. A recent study combined various bioinformatic approaches to make predictions about disease risk and pharmacological complications in a sequenced family. This project also generated three new reference sequences that can be used as a resource by other medical resequencing projects.

Dewey *et al.* carried out whole-genome sequencing of a family of four consisting of the mother, father, daughter and son. Central to the identification and interpretation of the genetic variants is the comparison with an appropriate reference sequence. However, the current human genome reference sequence that is in routine use, Hg19, is derived from a small number of anonymous donors. To overcome the potential sampling bias of Hg19 and to account for ethnicity, the authors used data from the [1000 Genomes Project](#) to generate three geographically distinct reference sequences. To

do this, they replaced ~1.6 million SNPs in Hg19 with the most common allele at each position in European, African and East Asian populations.

Use of an ethnically matched reference sequence reduced the genotyping error rate by improving the alignment and calling of genome sequencing reads and also reduced the number of variants that were identified in the family's genome sequences. This resulted in a more accurate list of higher priority variants to interpret relative to using Hg19.

So what did the genetic variants reveal? Among the most informative variants were variants of *F5* (which encodes coagulation factor V) and methylenetetrahydrofolate reductase (*MTHFR*) that were both inherited from father to daughter. Additionally, a variant of hyaluronan binding protein 2 (*HABP2*) was inherited from mother to daughter. Individually, these three variants are risk alleles for thrombophilia (a disorder of excessive blood clotting). Currently, only the father's clinical history is consistent with this condition, but the results suggest that the daughter could benefit from lifestyle changes or

preventative anticoagulants. However, analysis of variants in the daughter's drug-metabolizing enzymes also highlighted particular drugs that could result in adverse bleeding if combined.

Crucially, despite the *F5* thrombophilia variant being rare in the general population, it is present in Hg19; hence, use of this reference sequence could have rendered this variant undetectable. Overall, these results highlight that in genomic analyses, the quality of the reference can be as crucial as the samples themselves.

Darren J. Burgess

**ORIGINAL RESEARCH PAPER** Dewey, F. E. *et al.* Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. *PLoS Genet* **7**, e1002280 (2011)

