

 HUMAN GENETICS

Strength in numbers in the low-frequency spectrum



Numerous genome-wide association studies in recent years have identified scores of common variants that associate with human complex traits and disease. Rare and low-frequency variants, which are collectively common, have been much less well studied, mainly for methodological reasons. Fortunately, whole-genome sequencing (WGS) and large cohorts are now beginning to address this discrepancy. Recent papers, primarily in *Nature* and *Nature Communications*, now report the main results of the UK10K project, a Wellcome Trust-funded effort designed to characterize rare and low-frequency variants in the UK population and to evaluate their contribution to medically relevant traits and disease. Collectively, the papers identify specific variants in this frequency spectrum that associate with various phenotypic traits; but, importantly, they also give us invaluable genomic tools in terms of an improved reference panel and new imputation insights.

The UK10K project was designed in two parts. The cohort part was intended to reveal the contribution of genetic variation to a range of 64 traits in 3,781 healthy individuals; this cohort was subjected to 7×, low-read-depth WGS. The other part of the project focused on exomes in some 6,000 individuals and, by high-read-depth sequencing (80×), aimed to identify causal variants involved in rare disease, severe obesity and neurodevelopmental disorders. The main paper by the UK10K consortium provides an overview of the project's strategy and findings, whereas other papers focus on associations between genetic variants and specific traits:

Zheng *et al.* for bone mineral density and fracture; Taylor *et al.* on thyroid function; and Timpson *et al.* for circulating lipid levels. Additionally, Geijs *et al.* describe web tools for accessing the association results and genome-wide summary statistics, and the individual-level genotype and phenotype data are available under managed access conditions from [UK10K Data Access](#).

A project of this size and complexity is certain to yield a plethora of useful insights, some of which are specific to individual traits, whereas others have broader implications. For example, the study revealed that the penetrance of recognized variants for specific disorders — that is, the proportion of individuals harbouring a genetic variant that are affected by the associated phenotypic disorder — is likely to be overestimated in many studies. The consortium authors suggest that, as a benchmark, future studies should also report estimates of population frequencies for a given reported variant.

An important legacy of the project is a new haplotype reference panel, which significantly increases coverage of rare and low-frequency variants compared to other existing panels, including the 1000 Genomes panel.

Imputing genotypes from WGS reference panels is widely used to augment data available from genome-wide single-nucleotide polymorphism (SNP) microarrays. In another companion paper, Huang *et al.* show that improvements to imputation can be made by combining WGS panels and re-phasing them after initial genotype calling. In addition, they

show that increasing sample sizes is likely to be the most efficient way of discovering new loci driven by common variants. This is because known common variants can be exhaustively imputed using existing panels.

Other methodological insights are reported in the UK10K overview paper. It has been known for some time that the association tests most often used for common variants are not optimal for variants at the opposite end of the frequency spectrum. Although many questions remain, the UK10K project provides important insights into the relative utility of different methodologies when looking for rare and low-frequency variant associations.

In agreement with previous reports, the UK10K study found that variants predicted to have greater phenotypic effects tend to be rare or of low frequency. That said, the results clearly indicate that there are few low-frequency variants with very large effect that make a substantial contribution to population trait variation. This means that WGS in very large cohorts will be needed in order to explore the contribution of variants at this frequency spectrum to complex traits. This, in turn, means that the future of human genetics will have to be even more collaborative.

Magdalena Skipper, Executive Editor,
Nature Partner Journals

ORIGINAL RESEARCH PAPERS UK10K

Consortium. The UK10K project identifies rare variants in health and disease. *Nature* <http://www.dx.doi.org/10.1038/nature14962> (2015) | Zheng, H. F. *et al.* Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* <http://www.dx.doi.org/10.1038/nature14878> (2015) | Taylor, P. N. *et al.* Whole-genome sequence-based analysis of thyroid function. *Nat. Commun.* **6**, 5681 (2015) | Timpson, N. J. *et al.* A rare variant in APOC3 is associated with plasma triglyceride and LDL levels in Europeans. *Nat. Commun.* **5**, 4871 (2014) | Geijs, M. *et al.* An interactive genome browser of association results from the UK10K cohorts project. *Bioinformatics* <http://dx.doi.org/10.1093/bioinformatics/btv491> (2015) | Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015)