⮕ **DISEASE GENETICS**

UK Biobank — a new era in genomic medicine

> " a powerful way to link genetic and epidemiological factors to disease risk "

Population-based 'biobank' studies are a powerful way to link genetic and epidemiological factors to disease risk. The UK Biobank is a long-term prospective cohort study that includes genetic data associated with extensive phenotypic and health-related information. Two publications in *Nature* now report on the first main phase of the project.

The first paper by Bycroft et al. includes a summary of the high-resolution genetic and phenotypic data on the full cohort. The project enrolled ~500,000 volunteers aged 40–69 years from assessment centres across the UK from 2006 to 2010. Participants contributed detailed questionnaires, physical measurements and samples, and consented to be followed up with repeat measurements and linked health records. This approach enabled a focus on adult diseases; by using UK national health service (NHS) registries, the team have recorded over 14,000 deaths and 79,000 cancer diagnoses to date.

The majority of samples were genotyped on the Applied Biosystems UK Biobank Axiom array, which was developed specifically for this project. The array includes 825,927 single-nucleotide polymorphism (SNP) and insertion and deletion (indel) markers. The array design was optimized for genome-wide coverage to facilitate imputation in European populations and to include coding variants across a range of allele frequencies, including rare and low frequency, as well as regions with known or suggested roles in disease.

To produce the genotyped and imputed data set, which increased the number of testable variants to ~96 million, the authors developed several new computational methods. They also developed a new file format for improved data compression, thereby facilitating the distribution of the data sets. Strong data sharing policies make this resource even more impactful to health research; UK Biobank is making their full data sets available, as well as results conducted by researchers accessing these data.

Participant diversity was not fully representative of the UK's population diversity, as expected from the recruitment process. 94% of participants self-reported as white, and 88.26% as British and white. 84% of the full cohort showed similar ancestry based on both self-reporting and genetic information, and data on these individuals is offered as a more homogenous ancestry data set for analyses. Further studies are needed to examine fine-scale population structure within this data set.

The authors also examined relatedness on the basis of genetic data, estimating that UK Biobank includes 22,666 sibling pairs and that 30% of participants were related at third degree or closer to at least one other participant, which is higher than expected by chance and may reflect the nature of the recruitment and sampling bias.

In a second paper, Elliott et al. report the brain imaging data from the first ~10,000 UK Biobank participants, providing a resource for joint analyses of neuroimaging measures and genetics. The authors conducted genome-wide association studies (GWAS) for 3,144 different measures of brain structure and function, including structural volumes, lesion size and the connectivity and microstructure of the brain's white matter.

The team reported 148 clusters of associations between genetic variants and imaging traits and showed that many of these traits are heritable. They also observed genetic correlations with neurodegenerative, psychiatric and personality traits. The associations include candidate genes involved in the transport and storage of iron, which can play a role in neurodegenerative disorders. Imaging on 100,000 participants is planned to be completed in 2020.

This nearly 15-year-long experiment of building a national biobank in the UK has begun to bear fruits and is accelerating international genetics and health research, as reflected by the more than 100 preprints in bioRxiv and more than 500 publications using UK Biobank data even before these first main papers on the project's data sets were published. We hope to see this experiment replicated many times over, in locations across the world, and with the inclusion of more diverse populations.

Orli G. Bahcall, Senior Editor,
Nature

ORIGINAL ARTICLE Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018) | Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018)
RELATED LINK UK Biobank: http://www.ukbiobank.ac.uk

Credit: Pal Szilagyi Palko/EyeEm/Getty