

NEWS AND COMMENTARY

Genetic discovery in multi-ethnic populations

Matthew Traylor and Cathryn M Lewis

European Journal of Human Genetics (2016) **24**, 1097–1098; doi:10.1038/ejhg.2016.38; published online 4 May 2016

Genome-wide association studies (GWAS) have been hugely successful in identifying regions of the genome associated with traits and with disease risk. However, due to technical and statistical limitations, these studies have often been undertaken in homogeneous populations of European ancestry. This has been to the detriment of genetic discovery in other populations, in which our understanding of genetic effects is much narrower. In part this has been due to inadequate characterization of genetic ancestry across and within populations. With the advent of 1000 Genomes Phase 3 and population-specific projects such as the African Genome Variation Project,^{1,2} the resources are now available to perform such characterization. However, the appropriateness of established and novel statistical approaches to a multi-ethnic design has, until recently, remained opaque.

On page xxx of this issue, Cook and Morris³ report a statistical approach that adequately accounts for different ancestry in a multi-ethnic design GWAS. Using Type 2 Diabetes (T2D) as an example, in data from the Resource for Genetic Epidemiology on Adult Health and Aging—a large multi-ethnic population-based cohort—they show that their method both adequately accounts for population structure and can identify novel variants associated with the disease. The approach uses established techniques and can easily be implemented in standard GWAS software packages such as PLINK. Although the focus in this paper is on GWAS analyses, the method would likely also be of interest to those conducting large-scale sequencing studies. In this analysis, they

identify a novel association with T2D at the *TOMM40-APOE* locus, a region that has previously been implicated in Alzheimer's Disease, coronary heart disease, and lipid metabolism.^{4–6} This finding highlights the benefits of ancestrally inclusive study designs. In addition, Cook and Morris integrate tests of heterogeneity by axes of genetic variation (AGV)—variables derived from principal components analyses that best distinguish genetically dissimilar individuals—to formally detect heterogeneity in allelic effects between ancestry groups. They identify heterogeneity at the *TCF7L2* locus by the first AGV, corresponding to weaker effects of the SNP on T2D susceptibility in East Asians. Whether this association arises due to the different genetic background in the population, or because of different exposures, remains a pertinent question, particularly given the strong influence of the locus in Europeans.

USE OF MULTI-ETHNIC GROUPS IN GENETIC RESEARCH

We currently lack relevant terminology for the diverse genetic populations. While geneticists use convenient but clunky terms like 'of European ancestry', or 'of Asian ancestry', these can include individuals with highly diverse genetic lineage and are non-specific. The HapMap study considered that how populations are named 'has important ramifications scientifically, culturally, and ethically' (<http://hapmap.ncbi.nlm.nih.gov/citinghapmap.html>), and a recent *Science* article reminded us of the dangers in using racial terminology and classification in genetics.⁷ Cook and Morris's work emphasises this point, in that it treats genetic ancestry as a continuum and therefore better reflects the often complex genetic lineage of study participants. Indeed, classification by self-declared ancestry or by ascertainment from a specific geographic region often poorly reflects genetic reality.⁸ It is important that

statistical methods lead the way in using human genetic diversity effectively to understand the genomic underpinnings of disease. Studies assessing the contribution of variants to disease have shown that, in general, associated variants transcend ethnic groups, and play a role in disease susceptibility worldwide. There are analytical challenges to performing genome-wide association studies in the mixed ancestry and admixed populations, with care needed in both quality control and association testing,⁹ but the benefits to genetic discovery from recruiting ancestrally heterogeneous populations are considerable.¹⁰ The larger sample sizes attainable lead to novel loci being identified,¹¹ and differences in linkage disequilibrium patterns across populations open up the potential to identify causal variants at associated loci through using fine-mapping approaches.¹²

FUTURE PROSPECTS

The near future sees the release of genetic data from large multi-ethnic biobanking studies, including 500 000 participants from UK Biobank.¹³ Methods such as that outlined by Cook and Morris will no doubt help to power discovery and characterization of genetic associations in such resources. Furthermore, as the statistical methods and resources are now available to perform multi-ethnic analyses, researchers should be encouraged to recruit individuals to genetic studies regardless of ethnicity, helping to power discovery of further trans-ethnic associations. This will be important as our research focus shifts from genetic discovery to interpretation, and will aid translation of GWAS findings to clinical practice. These multi-ethnic studies will be essential to ensure that the benefits of the genetic revolution and of personalised medicine are universal. ■

CONFLICT OF INTEREST

The authors declare no conflict of interest.

M Traylor is at the Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, UK and CM Lewis is at the SGDP Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
E-mail: Cathryn.lewis@kcl.ac.uk

- 1 Gurdasani D, Carstensen T, Tekola-Ayele F *et al*: The African Genome Variation Project shapes medical genetics in Africa. *Nature* 2015; **517**: 327–332.
- 2 Auton A, Brooks LD, Durbin RM *et al*: A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.
- 3 Cook JP, Morris AP: Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility. *Eur J Hum Genet* 2016; **7**: e1001363.
- 4 Lambert JC, Ibrahim-Verbaas CA, Harold D *et al*: Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**: 1452–1458.
- 5 Teslovich TM, Musunuru K, Smith AV *et al*: Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707–713.

- 6 Nikpay M, Goel A, Won HH *et al*: A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015; **47**: 1121–1130.
- 7 Yudell M, Roberts D, DeSalle R, Tishkoff S: SCIENCE AND SOCIETY. Taking race out of human genetics. *Science* 2016; **351**: 564–565.
- 8 Tishkoff SA, Kidd KK: Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* 2004; **36**: S21–S27.
- 9 Medina-Gomez C, Felix JF, Estrada K *et al*: Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur J Epidemiol* 2015; **30**: 317–330.
- 10 Asimit JL, Hatzikotoulas K, McCarthy M, Morris AP, Zeggini E: Trans-ethnic study design approaches for fine-mapping. *Eur J Hum Genet* 2016; e-pub ahead of print 3 February 2016; doi:10.1038/ejhg.2016.1.
- 11 Li YR, Keating BJ: Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Med* 2014; **6**: 91.
- 12 Morris AP: Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* 2011; **35**: 809–822.
- 13 Sudlow C, Gallacher J, Allen N *et al*: UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**: e1001779.