Nature Reviews Genetics | https://doi.org/10.1038/s41576-018-0047-5 | Published online: 09 August 2018

COMPLEX TRAITS

Lessons from 1 million genomes

> imperfect genetic correlation of the phenotype across cohorts

A person's educational attainment is a model behavioural phenotype for genetic analysis owing to the very large available sample sizes. A study published in *Nature Genetics* now reports a genome-wide association study (GWAS) for educational attainment in more than 1 million individuals and identifies 1,271 independent lead single-nucleotide polymorphisms (SNPs) associated with this phenotype.

In total, 71 independent cohorts were brought together in a sample-size-weighted meta-analysis to increase the number of analysed individuals from the previous 405,072 to now 1,131,881. The 162 SNPs known from the hitherto largest study showed good reproducibility in the newly added data sets. Analysis of the X chromosome in a subset of 694,894 individuals uncovered 10 lead SNPs, and the SNP heritability due to the X chromosome was estimated at 0.3%, which is lower than what would be expected for a comparable autosome. A separate analysis of the X chromosome in women and men (including only data from the UK Biobank) demonstrated almost identical SNP heritability estimates between sexes.

Credit: Ellen Isaacs/Alamy

Using a Bayesian statistical framework, posterior probabilities of SNP effect sizes were calculated and the median effect per allele of lead SNPs was found to correspond to 1.7 weeks of schooling.

Autosomal SNPs were annotated using DEPICT software and a striking enrichment of SNPs near genes implicated in the central nervous system, specifically with neurophysiological functions and synaptic plasticity, was observed. Of note was the absence of enrichment of glia cell and myelination gene sets. The authors point out that this might be considered an argument against the hypothesis of individual differences in cognition being driven by differences in axonal transmission speed.

To achieve polygenic prediction of education, various polygenic scores were calculated and the predictive power was found to be greater the more SNPs were included, that is, the less stringent the significance threshold for inclusion was. The best score reached a predictive accuracy of 12.7% and 10.6%, respectively, in the two tested independent cohorts — a better predictor than household income but less predictive than the educational attainment of the parents.

Despite the staggering numbers that this GWAS reports, both in terms of sample size and identified SNPs, challenges remain. The authors discuss the need for inclusion of more diverse ancestries and raise the issue of imperfect genetic correlation of the phenotype across cohorts owing to educational systems differing locally and over time, something that has to be taken into account in power calculations and study design. With a phenotype as highly dependent on the environment as educational attainment, the authors further caution that polygenic score effects may represent in part environmental influences shared with the parents.

Michelle Trenkmann, Associate Editor,
Nature Communications

ORIGINAL ARTICLE Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018)

FURTHER READING Plomin, R. & von Stumm, S. The new genetics of intelligence. *Nat. Rev. Genet.* **19**, 148–159 (2018)