*Nature Reviews Genetics* | Published online 22 Feb 2016; doi:10.1038/nrg.2016.18

Oleksiy Maksymenko/Alamy

COMPLEX TRAITS

# Integrating gene variation and expression to understand complex traits

Genetic variants can influence complex human traits in many ways, often through the modulation of gene expression. Now, researchers have integrated genetic and transcriptional variation to conduct a transcription-wide association study (TWAS), linking gene expression with complex traits.

Association studies that measure both genetic variation and gene expression in individuals with measured traits tend to be costly to perform, and often have smaller sample sizes, with a corresponding reduction in power of association, than studies that measure genetic variation and traits alone. Gusev *et al.* have attempted to work around this limitation by focusing on a small set of reference individuals for whom data on both gene expression and single-nucleotide polymorphisms (SNPs) were available. By correlating genotype and expression to phenotypic traits from large-scale genome-wide association studies (GWAS), the authors derived the *cis*-genetic component of gene expression that is relevant to the trait. The team then built on this approach to develop a way of imputing expression–trait association statistics directly from GWAS summary statistics.

Genome-wide SNP and gene expression data from the Metabolic Syndrome in Men Study, the Young Finns Study and the Netherlands Twin Registry enabled the authors to estimate the association between expression and complex traits for approximately 7,000 genes enriched for heritable expression. Conceptually, this can be seen as the correlation between the genetic component of gene expression and the genetic component of a complex trait, or the colocalization of signal between expression and trait. The team compared their method to other recently proposed tests for these associations (LDSC and COLOC, respectively) and saw superior performance with less 'noise' in their approach.

To validate their approach, the authors analysed previous smaller GWAS data sets for traits associated with height or lipid levels and 'rediscovered' loci that had been identified in more recent and much larger GWAS. Turning to summary-level data from recent GWAS on lipid measurements, height and BMI, the group's TWAS approach identified 69 new associations for potential follow-up functional studies. Using the Hybrid Mouse Diversity Panel, they found that 14 of these identified genes were significantly associated with obesity phenotypes.

No statistical method is without its limitations and assumptions. The authors are unable to rule out that expression is mediated by phenotype rather than vice versa, or that the associations do not arise from independent effects of phenotype and expression at the same SNP. Finally, their TWAS cannot account for rare variants or nonlinear relationships.

However, by focusing on the link between a gene and a complex trait, the TWAS method of Gusev *et al.* lowers the burden of multiple testing and captures the heterogeneous signal of multiple SNPs, focusing the analysis on the transcriptional basis of complex traits. This new type of analysis can be applied to previous GWAS summary-level data to identify new associations between genes and traits, thereby enabling a better understanding of the genetic basis underpinning complex characteristics.

*Ross Cloney, Associate Editor,*
Nature Communications

**ORIGINAL ARTICLE** Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* http://dx.doi.org/10.1038/ng.3506 (2016)