An atlas of the genetic architecture of gene expression traits across the entire human body

Heather E. Wheeler[1,2], GTEx Consortium, Kaanan P. Shah[3], . . ., Nancy J. Cox[4], Dan L. Nicolae[3], Hae Kyung Im[3]

[1]Department of Biology and [2]Department of Computer Science, Loyola University Chicago, [3]Section of Genetic Medicine, Department of Medicine, University of Chicago, [4]Division of Genetic Medicine, Vanderbilt University 2015-11-23 15:36:22

Using the hybrid polygenic-sparse approach of BSLMM (Bayesian Sparse Linear Mixed Model) [1], we show that the local architecture of gene expression is sparse (high PGE) for most heritable genes in both DGN and GTEx. Using the elastic net [2], we observed improved cross-validated expression prediction for $\alpha \geq 0.5$ across tissues, confirming the sparsity result. This result demonstrates that sparse effects can be identified with sample sizes in the hundreds rather than the thousands and is supported by many prior studies with sample sizes near 100 that identified replicable eQTLs near the transcription start sites of genes[1]. Conversely, for traits that are highly polygenic, e.g. height, BMI, schizophrenia, and bipolar disorder, thousands to tens of thousands of samples are needed to identify significant genetic signals [**???**]. Therefore, the distal contributions to expression h^2 are likely to be more polygenic because they could not be accurately estimated here with sample sizes in the hundreds.

1. Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. Visscher PM, editor. PLoS Genetics. Public Library of Science (PLoS); 2013;9: e1003264. doi:10.1371/journal.pgen.1003264

2. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). Wiley-Blackwell; 2005;67: 301–320. doi:10.1111/j.1467-

9868.2005.00503.x