Name: Jack Pattee Email: jack.pattee@cuanschutz.edu

Adjusting for cis-regulatory variation in gene expression data can improve biological network analysis in model organisms

Jack Pattee[1], Daniel Munro[2], Abraham Palmer[3,4], Pejman Mohammadi[2], Laura M Saba[5]

[1]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; [2]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA, USA; [3]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA; [4]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA; [5]Department of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Often, studying gene co-expression patterns as networks provides insight into the etiology of biological processes. Model organisms are a rich data source for network analyses as the reduced genetic and environmental variability in these populations can facilitate co-expression modeling.

However, in some model organism data, cis-SNPs may substantially affect the gene expression levels of many nearby genes located in blocks of high linkage disequilibrium. As a result, close physical proximity between two genes may induce correlation in their expression levels even if they are not functionally related. Thus, estimated cluster assignments may represent co-localization rather than a shared biological process. We investigated this problem by comparing gene clustering data from five brain regions in heterogeneous stock rats. To determine the biological relevance of the estimated clusters, we conducted overrepresentation tests for genes associated with small molecule perturbations in the Signature Commons (SigCom) LINCS database. Adjusting for cis-SNPs in clustering generated a higher proportion of clusters overrepresented for genes associated with small molecule perturbations, indicating that such clusters are more likely to represent shared biological function. In summary, adjusting for cis-SNPs in co-expression network analyses can reduce the number of clusters driven by co-localization while preserving clusters driven by shared biological function. Supported by a Skaggs Scholars Grant Award from the ALSAM Foundation, by NIDA (P50 DA037844; P30 DA044223), and by NIH R01GM140287.