Association of addiction phenotypes to predicted transcript levels

Darius Bost, Chris Bizon, Jeff Tilson, Xianyong Yin

GWAS studies unveil a wide range of variants associated with complex disease traits. Interpretation of these variants, however, is limited as it does not address the functional mechanisms involved. This deficit in identifying biological mechanisms, along with the penalty for multiple testing, hinders our ability for attributing causality from our variants to our traits. PrediXcan, a gene-based association method that prioritizes genes that are likely causal for phenotypes, has been introduced to circumvent the aforementioned hurdles in single marker analysis. PrediXcan utilizes gene regulation as a means to identify gene to trait associations. By quantifying the association between genetically regulated expression and phenotype PrediXcan allows the user to gather more biological inference from their subset of variants. PrediXcan uses reference transcriptome data to train additive models of gene expression levels. Weights were derived from the estimates of the genetically regulated expression. In regression testing which is used by Hae Kyung Im et al. there is an assumption of independence between data. This does not lend itself to studies with family data. To account for identity-by-decent we implement the use of a linear mixed model(LLM) that will produce a relatedness matrix. We implement this LLM with a Genome-wide Efficient Mixed Model Association(GEMMA) algorithm. Using the model weights estimated by PrediXcan, we predict transcript levels that we perform association test on with GEMMA. Our association is conducted with addiction phenotypes which include, but aren't limited to, alcohol use, marijuana use, and nicotine dependence.