A Novel Approach to Accounting for Exposure in Association Studies of Drug Use Disorders

Dani Sisto, Michael Vanyukov, Brion Maher

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health
Department of Pharmaceutical Sciences, University of Pittsburgh

Case-control genetic association studies for drug use disorders face a unique challenge relative to other complex diseases in the selection of controls. Early genetic studies of addiction frequently contrasted drug-dependent cases and unaffected population controls. However, pathogenesis of substance use disorder (SUD) is a multi-stage process, through drug use, then abuse, and ultimately psychological and physiological dependence and compulsive pursuit of drugs, with varying influence of genes and environment on individual variation by stage. Study designs that use controls who report drug exposure but do not satisfy criteria for SUD likely deliver greater power to detect allelic associations pertaining to risk for dependence, since it is unclear if controls who have never used would develop dependence if exposure occurred. Consistent with this observation, the recent successes in addiction genetics have been in studies focusing on variation within drug-exposed populations or included controls with significant exposure, most notably exemplified by the Tobacco and Genetics Consortium GWAS. However, defining the sufficiency of exposure to maximize power has usually been done arbitrarily. Herein, we describe a novel approach relying on a modified version of Ordered Subset Analysis where we sequentially add cases and controls based in an ordered fashion based on a measured quantitative exposure variable (e.g., number of lifetime cigarettes). Unlike conventional OSA, we add cases starting with the lowest level of exposure and controls starting with the highest level of exposure. The test statistic used

is the maximum level of association attained at any subset with significance calculated by permutation of case-control status and repeated ordered subset analysis. We compare the power of the approach to detect a single SNP with approaches using unscreened controls and arbitrarily-defined exposure thresholds in simulated data with a polygenic underlying liability and varying the proportion of genetic variance "turned on" by exposure. We highlight scenarios where the novel approach performs favorably to the usual approaches and describe the computational burden of using the approach genome-wide.