

Submitter Name: John Rice
Submitted email: john@zork.wustl.edu

SmokeScreen: A GWAS resource for Substance Abuse

John Rice¹, Scott Saccone¹, Lingwei Sun¹, William Howells¹, Christian Bixby², Nina Sanapareddy², Harshad Joshi², Jay Tischfield², and Andrew Brooks²

¹Department of Psychiatry, Washington University; ²Human Genetics Institute of New Jersey, Rutgers University

The first GWAS (Genome-Wide Association Study) was published in 2005, and utilized chips that provided whole genome coverage with SNPs (Single Nucleotide Polymorphisms). For complex traits, prior approaches using genetic linkage in families or candidate gene analysis had limited success, whereas GWAS studies quickly identified replicable significant hits. The threshold for significance was chosen as $0.05/1 \text{ million} = 5 \times 10^{-8}$ to avoid false positives. The resulting studies found complex traits to be polygenic with many genes of small effect. Subsequently, PRS (Polygenic Risk Score) and LDSC (LD Score Regression) techniques were introduced. These methods require only the results from a (large) discovery sample and the PRS (using many SNPs and not only GWAS significant ones) can then be applied to a smaller target sample with individual level genotypes. LDSC can be used to estimate heritability, the genetic correlation between traits, and distinguish polygenicity from population stratification.

Against this background, NIDA started the SmokeScreen Project in which existing case-control samples of drug dependence or HIV are genotyped at Rutgers and cleaned genetic and phenotypic data are submitted to dbGaP by Washington University for use by other investigators. To date, 17 studies have been released by dbGaP and many others are in progress. This resource can be used with modern analytic approaches in multiple ways to better understand drug dependence and related behaviors. Detailed procedures, possible applications, and lists of released and ongoing studies by phenotype will be presented.