Specific Aims

Sustained multi-substance use (alcohol, tobacco, and opioids) is common among veterans and a major cause of morbidity and mortality. Numerous genetic variants have been associated with alcohol, tobacco or opioid dependence but these explain only a small proportion of phenotypic variation. Studies of shared genetic pathways across substances have yielded inconsistent results. Among the major challenges for gene discovery are phenotypic ambiguity and inadequate statistical power to detect the small genetic effects associated with individual variants. Phenotypic inaccuracy can result from cross-sectional assessment using Diagnostic and Statistical Manual (DSM)-criteria or diagnoses, which yield lower sensitivity and specificity than longitudinal, quantity-frequency data. Inadequate statistical power stems from the difficulty in ascertaining large numbers of well-phenotyped individuals. To overcome these limitations, we propose to use validated phenotypes based on quantity-frequency data (Alcohol Use Disorder Identification Test (AUDIT-C)) for alcohol, self-reported smoking for tobacco, and prescription refills for opioid use) from a longitudinal electronic health record (EHR) in a large Veteran population. These phenotypes, previously developed and validated within the Veteran Aging Cohort Study (VACS), will be used to identify genetic variants associated with sustained heavy use of each substance and joint, multi-substance use in the Million Veteran Project (MVP). In this one-year proposal, we will apply a two-stage genome-wide association (GWA) approach for genetic analysis. In the future, we plan to replicate the findings from this project in a large population from the Psychiatric Genetic Consortium (N=400,000).

Aim 1: To define and identify highly valid, longitudinal heavy substance use phenotypes in the MVP using VACS-validated algorithms (N=300,000).

   a) Identify and compare/contrast the characteristics of heavy longitudinal alcohol, tobacco and opioid use phenotypes in the MVP sample using the VACS-validated algorithm;
   b) Develop dual and multi-substance use phenotypes based on the results from Aim 1a.

Aim 2: Discovery stage: To conduct GWA studies (GWAS) for sustained heavy use of each of the three substances (alcohol, tobacco, and opioids) and for dual and multi-substance use, as defined in aim 1 (N=200,000). Using the phenotypes defined from Aim 1, we will perform a GWAS in a discovery sample of 200,000 individuals selected from the 300,000 participant sample. European Americans (EAs), and African Americans (AAs) will be differentiated genetically and analyzed separately. Using standard methods, Hispanic individuals will be assigned genetically to either the EA or AA group. Principal component analysis (PCA) will be applied to address population admixture across and within each population.

   a) GWAS for sustained heavy alcohol, tobacco, and opioid use separately in EAs and AAs;
   b) GWAS for dual or multi-substance use for EAs and AAs;
   c) Meta-analysis across populations.

Aim 3: Replication stage: To replicate the GWAS findings from Aim 2 in the replication sample (N=100,000). We will perform association analysis to replicate the signals identified in Aim 2, separately by EAs and AAs.

   a) To replicate the findings from the Aim 2 (a) for each substance use phenotype;
   b) To replicate the findings from the Aim 2 (b) for dual or multi-substance use;
   c) To meta-analyze across populations.