Abstract

Sustained multi-substance use (alcohol, tobacco, and opioids) is common among veterans and a major cause of morbidity and mortality. According to the Centers for Disease Control, substance use disorders collectively represent the single most preventable cause of disease, disability, and death in the United States. Numerous genetic variants have been linked specifically to alcohol, tobacco or opioid use but these explain only a small proportion of phenotypic variation. Studies aiming to find shared genetic loci and 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 that are accounted for by individual variants. Phenotypic ambiguity can result from cross-sectional assessment using 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 have assembled a consortium of scientists with unparalleled expertise in substance use clinical epidemiology, genetics and computational genomics to harness the unprecedented opportunity that the Million Veteran Program presents to advance this field. To address the specific limitations of prior work noted above, we propose to use validated phenotypes based on quantity-frequency data (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 patient 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). We will apply a two-stage genome-wide association (GWA) approach for genetic analysis.

Prior to GWA analysis, we will define and identify highly valid, longitudinal heavy substance use phenotypes in the MVP sample, employing VACS-validated algorithms in the EHR. Then, we will perform the standard two-stage GWA approach: discovery and replication. Both stages will be stratified by race/ethnicity (European American and African American); will examine the genetic associations with single, dual and multi heavy substance use phenotypes; will use principal component analysis to address population admixture; and will be meta-analyzed across population strata.