Name: Xingyan Wang
Email: xzw151@psu.edu
PI Name: Dajiang Liu
PI email: dxl46@psu.edu

Trans-Ancestry Fine-Mapping and Ancestry Heterogeneity Analysis for Smoking & Drinking Addiction Phenotypes using 3.4 Million Individuals with Diverse Ancestries

Xingyan Wang[1], Trans-Omics for Precision Medicine, GWAS and Sequencing Consortia of Alcohol and Nicotine Use

[1]Department of Public Health Sciences, Penn State College of Medicine, Hershey PA 17033

Smoking and drinking have been associated with chronic diseases and disorders. Recent GWAMA studies have identified 406 loci in European ancestry. While the genetic architecture in non-European ancestry remains elusive. To better address this problem, the current GSCAN study includes 3.4 million individuals with 21.47% (N=705,637) non-European ancestry. Trans-ancestry meta-analysis identified a total of 2,543 loci of which 711 are novel.

We proposed a meta-regression-based fine-mapping method. Specifically, we utilized principal components (PCs) derived from allele frequencies as proxies of continuously cohort-level ancestry. We modeled the genetic effect from each study as a mixture of models with varying numbers of PCs, which could encompass the different extent of heterogeneity for each variant. We borrowed information to learn the genetic architecture and fine-mapped causal variants by imposing a Dirichlet-Multinomial prior. Our approach improved fine-mapping resolution with fewer variants and higher coverage probability in the 90% credible set in simulations comparing to other existed methods. Additionally, our method could estimate fractions of loci with homogenous effects or ancestry-specific effects.

We applied our proposed method to the GSCAN study with five smoking/drinking phenotypes. We observed that 81% of loci have a homogenous effect, corresponding to the model with no PCs. 13% of loci are best supported by models with 1 PC, which indicates a distinction between European and Asian ancestry groups. The 90% credible sets from fine-mapping contain a median of 1.51 genes. These results could help us better understand the genetic architecture of smoking and alcohol use behaviors.