

Submitter Name: Randall J. Ellis
Submitter email: randy.ellis@icahn.mssm.edu

Genomic prediction of alcohol and opioid use disorders using machine learning

Randall J. Ellis¹, Hang Zhou², Marco Galimberti², Henry R. Kranzler³, Joel Gelernter², Yasmin L. Hurd¹

¹Addiction Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

²Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

³Center for Studies of Addiction, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

Collectively, alcohol use disorder (AUD) and opioid use disorder (OUD) kill over 140,000 Americans annually. As substance use disorders (SUDs) are ~50% heritable, understanding the genetic contribution to the etiology of SUDs is an open question. Additionally, developing personalized approaches to predicting SUD risk based on a patient's genetics would be valuable to clinicians. Here, we describe the development of machine learning models to predict diagnosis of AUD and OUD from single nucleotide polymorphisms (SNPs) from 11,336 patients of European and African descent. Feature importance methods were used to quantify the predictive value of all SNPs, and the overlap between predictive SNPs and previously identified SNPs in genome-wide association studies was assessed. Further, patient-level feature importance values were inputted to clustering algorithms to identify subtypes of AUD and OUD based on their global SNP phenotype. In sum, these studies demonstrate the value of machine learning algorithms for predicting complex psychiatric traits from genome-wide SNP data, and of model interpretability methods to provide explanations of how these predictions are generated that can inform further scientific study into the genetic etiology of these traits.