Submitter Name: Sandra Sanchez-Roige Submitted Email: sanchesroige.ucsd.edu

Characterizing cases and controls for opioid use disorders using electronic health records: a phenomic and genomic exploration in over 1 million individuals

Maria Niarchou[1], Ellen Tsai[1*], Annika Faucon[1*], Mariela V Jennings[2], Hyunjoon Lee[3], Kritika Singh[1], Justin Tubbs[3,4,5,6], Tian Ge[3,4,5,6], Jordan W Smoller[3,4,5,6], Lea K Davis[7,8,9] Brandon Coombes[10], Travis T Mallard[3,4,5.6], Vanessa Troiani[11], Sandra Sanchez-Roige[1,2,12]

[1]Vanderbilt Genetics Institute, Vanderbilt University Medical Center;
[2]Department of Psychiatry, University of California San Diego;
[3]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital;
[4]Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital;
[5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard Department of Psychiatry, Harvard Medical School;
[6]Department of Psychiatry, Harvard Medical School;
[7]Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center;
[8]Department of Biomedical Informatics, Vanderbilt University Medical Center;
[9]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center;
[10]Department of Quantitative Health Sciences, Mayo Clinic;
[11]Department of Autism and Developmental Medicine, Geisinger;
[12]Institute for Genomic Medicine, University of California San Diego
*shared second authors

In the midst of a global opioid crisis, uncovering the genetic underpinnings of opioid use disorder (**OUD**) is critical. However, efforts to understand etiology via genome-wide association studies (**GWAS**) have been hampered by insufficient phenotypic data in cases and controls and a reliance on highly ascertained samples. Although optimal characterization of OUD cases and controls remains challenging, the integration of extensive electronic health record (**EHR**) and genomic data offers new promise for advancing GWAS of OUD. Leveraging data from over 1 million patients across two healthcare systems, in this project we are iteratively determining the number of OUD diagnostic (**ICD**) codes needed for accurate identification of OUD well-known comorbidities, and evaluating the impact of including exposed vs. unscreened controls via phenome-wide association studies (**PheWAS**) and GWAS. Preliminary PheWAS results indicate that the inclusion of ≥1 OUD ICD codes recapitulates known OUD comorbidities (e.g. substance use disorders; pain). These associations are more pronounced in the unscreened group (i.e., 727 vs 644 in the exposed group) and of higher magnitude (mean OR(SE)=3.83±0.05 vs OR=2.52±0.52). In contrast, differences are less apparent in GWAS; associations with coding *OPRM1* SNP rs1799971 (p=8.83E-03 vs p=1.83E-02) and genetic correlations with clinically ascertained cohorts (r_g=0.83±0.26 vs r_g=0.86±0.29) are comparable across groups. Replication analyses are underway in *All of Us*. This study represents the first effort to optimize case/control definitions using EHR data and PheWAS with the ultimate goal of enhancing OUD genetic research.