

Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects on brain

Rachel L. Kember,^{1,2} Rachel Vickers-Smith,^{1,3,4} Heng Xu,² Sylvanus Toikumo,² Maria Niarchou,⁵ Hang Zhou,^{6,7} Emily Hartwell,^{1,2} Richard C. Crist,^{1,2} Christopher T. Rentsch,^{6,7,8} VA Million Veteran Program, Lea K. Davis,⁵ Amy C. Justice,^{6,7,9} Sandra Sanchez-Roige,^{5,10} Kyle M. Kampman,^{1,2} Joel Gelernter,^{6,7} Henry R. Kranzler^{1,2}

¹ Mental Illness Research, Education and Clinical Center, Veterans Integrated Service Network 4, Crescenz Veterans Affairs Medical Center, Philadelphia, PA 19104

² Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104

³ Department of Epidemiology, University of Kentucky College of Public Health, Lexington, KY 40536

⁴ Center on Drug and Alcohol Research, Department of Behavioral Science, University of Kentucky College of Medicine, Lexington, KY 40536

⁵ Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37232

⁶ Veterans Affairs Connecticut Healthcare System, West Haven, CT 06516

⁷ Yale School of Medicine, New Haven, CT 06511

⁸ Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK, WC1E 7HT

⁹ Yale School of Public Health, New Haven, CT 06511

¹⁰ Department of Psychiatry, University of California San Diego, San Diego, CA 92110

Acknowledgments

This work was supported by Merit Review Awards from the U.S. Department of Veterans Affairs Biomedical Laboratory Research and Development Service (# I01 BX003341) and Clinical Science Research and Development Service (# I01 CX001734; the VISN 4 Mental Illness Research, Education, and Clinical Center; NIAAA grant K01 AA028292 (to RLK); and NIDA grant DA046345. The views expressed in this article are those of the authors and do not necessarily represent the position or policy of the Department of Veterans Affairs or the United States Government.

Disclosure

Dr. Kranzler is a member of advisory boards for Dicerna Pharmaceuticals and Sophrosyne Pharmaceuticals; a consultant to Sobrera Pharmaceuticals; and a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last three years by AbbVie, Alkermes, Dicerna, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. Drs. Kranzler and Gelernter are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018. The other authors have no disclosures to make.

Abstract

Despite an estimated twin heritability of ~50%, genome-wide association studies (GWAS) of opioid use disorder (OUD) have revealed few genome-wide significant (GWS) loci, with replicated findings only in European-ancestry individuals. To identify novel loci, including those in non-European ancestries, and improve our understanding of the biology of OUD, we conducted a cross-ancestry meta-analysis using the Million Veteran Program (MVP). OUD cases in MVP had at least 1 International Classification of Diseases (ICD)-9 or ICD-10 code for opioid abuse or dependence (N=31,473). Opioid-exposed controls (N=394,471) had one or more outpatient opioid prescription fills. We conducted GWAS for each major ancestral group in MVP: African Americans (AAs; N=88,498), European Americans (EAs; N=302,585), and Hispanic Americans (HAs; N=34,861), followed by a cross-ancestry meta-analysis. Ten loci were GWS in the cross-ancestry meta-analysis, 8 of them novel. In addition to the known coding variant rs1799971 in *OPRM1*, which was the lead SNP genome-wide ($p=6.78 \times 10^{-10}$), and a recently reported exonic variant in *FURIN*, we identified intronic variants in *RABEPK*, *FBXW4*, *NCAM1*, and *KCNN1*. Ancestry-specific analyses identified an additional novel locus for each of the 3 ancestry groups. A supplementary meta-analysis within EAs that included MVP and other samples identified a locus in *TSNARE1*, which was also GWS in the cross-ancestry meta-analysis of all datasets. Gene-based association analyses identified 1 gene in AAs (*CHRM2*) and 3 in EAs (*OPRM1*, *DRD2*, and *FTO*). Significant genetic correlations (r_g 's) were identified for 127 traits, including positive correlations with schizophrenia, problematic alcohol use, and major depressive disorder. The most significantly enriched cell type group was the central nervous system with gene-expression enrichment identified in brain regions previously associated with substance use disorders. With a case sample 50% larger than that of the previous largest GWAS, we identified 14 loci for OUD, including 12 novel loci, some of which were ancestry-specific. These findings increase our understanding of the biological pathways

involved in OUD, which can inform preventive, diagnostic, and therapeutic efforts and thereby help to address the opioid epidemic.

Introduction

Opioid use disorder (OUD) is a problematic pattern of opioid use that leads to significant impairment or distress¹. In the United States, a 10-fold increase in opioid analgesic prescriptions between 1990 and 2010 contributed to an epidemic of opioid misuse, abuse, and overdose deaths²⁻⁴. By 2019, 3.7% of U.S. adults reported past-year opioid misuse and 0.6% met criteria for an OUD⁵. Overdose deaths, which continue to increase annually, have reached crisis proportions⁶, reflecting the limitations of available preventive and treatment efforts.

Genetic studies can help inform our understanding of the biology underlying OUD. However, although the estimated heritability (h^2) of OUD based on twin and family studies is ~50%⁷, few genetic associations have been identified. Genome-wide association studies (GWAS) of OUD, opioid dependence (OD) or related phenotypes have yielded inconsistent results, likely due to the limited sample size of the discovery datasets and different case and control definitions⁸⁻¹¹.

The use of data from electronic health records (EHRs) linked to biobanks has permitted research consortia to assemble increasingly large GWAS samples. An EHR-based study of 1,039 OUD cases and 20,271 controls identified two genome-wide significant (GWS) loci, with common SNPs explaining 6.0% of the phenotypic variation in OUD¹². A meta-analysis of European Americans (EAs) (10,544 OUD cases and 72,163 opioid-exposed controls) and African Americans (AAs) (5,212 OUD cases and 26,876 opioid-exposed controls) based largely on data from the Million Veteran Program (MVP), identified a single GWS SNP, rs1799971 in *OPRM1*, in EAs only, with SNP-based heritability estimated at 11.3%¹³. There were no GWS findings in AAs or in a cross-ancestry meta-analysis. A study that combined data from multiple cohorts (20,858 OUD cases), including an earlier release of MVP data, identified 2 GWS loci – a variant within *OPRM1* in a cross-ancestry analysis, and an additional variant in *FURIN* in a European-ancestry meta-analysis¹⁴.

Two recent GWAS have increased sample sizes for genetic discovery by examining opioid-related phenotypes other than OUD. A GWAS of prescription opioid misuse in an EA sample from 23andMe (27,805 cases) identified 2 novel GWS loci¹⁵. A meta-analysis of European-ancestry individuals included 23,367 cases ascertained using either Diagnostic and Statistical Manual of Mental Disorders diagnoses (DSM) or frequency of opioid use (FOU) and 384,629 controls¹⁶. Using a latent trait [opioid addiction (OA)] generated from a genomic structural equation model to conduct GWAS, the group identified GWS SNPs in *OPRM1* and, in gene-based analyses, *PPP6C* and *FURIN*. The SNP h^2 for the DSM trait was 11% and 18% for the FOU trait.

These studies also identified significant genetic correlations (r_g 's) with traits well known to co-occur with OUD, suggesting widespread pleiotropy. The strongest positive r_g 's were with substance-related traits (e.g., alcohol dependence/use disorder, cannabis use disorder, drinks/week, cigarettes/day)^{12,13,15,16}, and psychiatric disorders (e.g., attention deficit hyperactivity disorder (ADHD), posttraumatic stress disorder (PTSD), major depressive disorder (MDD), schizophrenia, bipolar disorder, neuroticism)^{13,15,16}. Although these traits are known to be genetically correlated, it is unknown whether there is a shared genetic structure between OUD, other substance use, and psychiatric disorders. Negative r_g 's were seen for educational attainment^{13,15,16}, cognitive performance¹³ and subjective wellbeing¹⁶. Causal effects on OUD for some of these traits were identified via Mendelian randomization (MR) analysis. Positive causal effects on OUD were found for the following exposures: regular tobacco smoking, depressed affect, neuroticism, and cognitive performance. A negative causal effect on OUD was also seen with cognitive performance as the exposure¹³. The causal effect of OUD on these traits was unable to be examined due to the limited number of GWS variants.

The different phenotypes used in these studies reflect the difficulty of ascertaining a large, multi-ancestry, well-characterized sample for use in GWAS of opioid-related phenotypes. EHR-based traits generally use International Classification of Disease (ICD) diagnostic codes

for phenotyping OUD (e.g.,^{12,13}). Cohorts recruited from some non-clinical biobanks rely on single-item, self-report questionnaires (e.g.,¹⁵) or have combined multiple case and control definitions derived as latent variables in genomic structural equation modeling¹⁶. A key consideration in selecting OUD cases, particularly given the high prevalence of opioid use in the United States, is the stringency of the definition. More stringent case definitions increase confidence in the specificity of the diagnosis and, by reducing heterogeneity, may increase statistical power. However, they also reduce the sample size and have the potential to reduce generalizability by not capturing a disorder's full range of presentations (e.g., by misclassifying cases as subthreshold).

Here, we conducted a cross-ancestry meta-analysis of OUD that included AA, EA, and Hispanic American (HA) subjects recruited from the MVP that maximized OUD cases by using a less stringent definition (requiring the presence of a single OUD diagnostic code) and compared them to opioid-exposed controls (N cases=31,473, N controls=394,471). In supplementary analyses, we compared our results to those using a stringent OUD phenotype in MVP, and performed a meta-analysis that combined data from the MVP, Yale-Penn (unpublished data), the Partners HealthCare Biobank¹², and the Psychiatric Genomics Consortium (PGC)¹¹. To elucidate the biology implicated by the variation that we identified through GWAS, we performed gene-based, gene-set enrichment and transcriptome-wide association analyses; examined r_g s with a wide variety of phenotypes; calculated polygenic risk scores (PRS) and performed phenome-wide association studies (PheWAS) in independent samples; and conducted Mendelian randomization analyses and genomic structural equation modeling to evaluate the causal effect and shared genetic structure between OUD, substance-related traits, and psychiatric disorders. We identified GWS findings in AA, EA and HA individuals. Further, we obtained novel insights that implicate genes expressed in the central nervous system (CNS), including specific brain regions implicated in the biology of OUD.

Methods

Overview of Analyses

We conducted an ancestry-specific GWAS using a less stringent OUD case definition in AAs, EAs, and HAs from the MVP, followed by a cross-ancestry meta-analysis. Cases had received at least one lifetime ICD Ninth Revision (ICD-9) or Tenth Revision (ICD-10) diagnosis of OUD and control subjects were opioid exposed. Further details on phenotyping are described below. This GWAS was used for all subsequent downstream analyses.

In a supplementary analysis, we performed within-ancestry meta-analyses for AAs and EAs from the MVP, Yale-Penn (unpublished data), the Partners HealthCare Biobank¹², and the PGC¹¹, followed by a cross-ancestry meta-analysis which included all samples. In a second supplementary analysis, we repeated the GWAS in MVP with the more stringent case definition used in the prior MVP OUD GWAS¹³. An overview of the analyses is provided in Supplementary Figure 1.

Million Veteran Program Cohort

As of September 2021, the MVP¹⁷ had recruited approximately 850,000 veterans at 63 VA medical centers nationwide. All participants provide written informed consent and a blood sample for DNA extraction and genotyping and give permission to access their EHR for research purposes. The MVP was approved by the Central Veterans Affairs Institutional Review Board (IRB) and all site-specific IRBs. All relevant ethical regulations for work with human subjects were followed in the conduct of the study.

Phenotypes

OUD diagnostic codes based on ICD-9/10 were obtained from the VA EHR. The main GWAS used a less stringent definition of OUD (N = 31,473), which required the presence of 1

inpatient or outpatient ICD-9/10 diagnostic code for OUD (304.0, 304.7, 305.5, F11.1, F11.2) in the EHR. The stringent definition (N = 23,459), used in the supplementary GWAS, required at least 1 inpatient or 2 outpatient ICD-9/10 OUD diagnostic codes in the EHR. Controls (N = 394,471) for all GWAS were defined as individuals with at least 1 outpatient opioid analgesic prescription fill [excluding prescriptions for OUD treatment (e.g., buprenorphine or methadone)] and no ICD-9/10 diagnosis code for OUD documented in the EHR (i.e., opioid-exposed). In all analyses, age at the time of MVP enrollment was used as a covariate. Demographics are presented in Supplementary Table 1.

Genotyping and Imputation

The genotyping of samples in the MVP is ongoing and, in this analysis, we used Release 4 imputed data. MVP samples were genotyped with a custom Affymetrix Axiom Biobank Array. Quality control of genotype data and subsequent imputation were performed by the MVP Genomics working group. Duplicate samples were removed, as were those with a sex mismatch, 7 or more relatives in MVP, excessive heterozygosity or a genotype call rate <98.5%. Variants were removed if they were monomorphic, had a missing call rate <0.8, or a Hardy-Weinberg equilibrium $p < 1 \times 10^{-6}$ both in the entire sample using a PCA-adjusted method and within 1 of the 3 major ancestry groups (AA, EA, HA). Genotypes were phased with SHAPEIT4 (v.4.1.3)¹⁸ and imputed using Minimac4¹⁹, with biallelic SNPs imputed using the African Genome Resources (AGR) reference panel by the Sanger Institute, and non-biallelic SNPs and indels imputed in a secondary imputation using the 1000 Genomes Project phase 3, version 5²⁰ reference panel. Indels and complex variants from the second imputation were merged into the AGR imputation.

Genome-wide association analyses were performed using PLINK 2.0²¹. We removed 1 individual from each pair of related individuals at random (kinship > 0.08, N = 31,010). Genetic ancestry was unified with self-identified race/ethnicity using the HARE (Harmonizing Genetic

Ancestry and Self-Identified Race/Ethnicity) method²². Quality control of imputed variants was performed within each ancestral group. Genetic variants were excluded based on minor allele frequency (MAF: AA<0.005; EA<0.001; HA<0.01), genotype call rate<0.95, and Hardy-Weinberg equilibrium $p < 1 \times 10^{-6}$ or a population-specific imputation quality (INFO) score <0.7. Covariates included sex, age at enrollment, and the first 10 genetic principal components (PCs) within each ancestry.

Datasets for Meta-analysis

Supplementary Table 2 summarizes the datasets used for meta-analysis. Summary statistics for GWAS of OUD were obtained from two previously published datasets: 1) Partners HealthCare Biobank, which used the same less stringent case definition and opioid-exposed controls in European ancestry individuals¹² 2) PGC, which used a DSM-IV OD diagnosis and opioid-exposed controls in African and European ancestry individuals¹¹. We also included the Yale-Penn 3 (YP3) unpublished dataset (Yale-Penn 1 and 2 were included in PGC analyses). In YP3, we conducted a GWAS using cases with a DSM-IV OD diagnosis and opioid-exposed controls. For AAs, there were 168 cases and 153 controls; for EAs, there were 578 cases and 219 controls. We used GEMMA to conduct association analysis to account for relatedness between the individuals. Sex, age at recruitment, and the first 10 PCs were included as covariates.

Meta-analysis and Independent Variants

Meta-analyses were conducted using a sample-size-weighted method in METAL²³. To compensate for the imbalance in the ratio of cases to controls, effective sample sizes were calculated using the formula: $4/[1/n_{\text{case}} + 1/n_{\text{control}}]$. Effective sample sizes were used in all meta-analyses and all downstream analyses. Meta-analyses were conducted across the following datasets: 1) cross-ancestry (AA, EA, and HA) meta-analysis within MVP, 2) within-

ancestry meta-analysis across datasets (AA: MVP, PGC, YP3; EA: MVP, Partners HealthCare Biobank, PGC, YP3), 3) cross-ancestry meta-analysis across all datasets (AA [MVP, PGC, YP3]; EA [MVP, Partners HealthCare Biobank, PGC, YP3]; HA [MVP]).

To identify independent variants, we performed LD-clumping within each ancestry using a range of 3000 kb, $r^2 > 0.1$, and the matched 1000 Genomes²⁰ reference panel. Following clumping, variants that were located <1Mb apart were merged into a single locus. For loci that contained multiple variants, we conducted conditional analyses using COJO in GCTA²⁴. Within each locus, we conditioned on the most significant variant. Upon conditioning, variants within the locus that remained significant ($p < 5 \times 10^{-8}$) were considered independent.

SNP-based Heritability Analyses and Partitioning Heritability Enrichment

We used LD score regression²⁵ (LDSC) to estimate OUD SNP-based heritability (h^2_{SNP}) in AAs and EAs for common SNPs mapped to HapMap3²⁶. To ensure matching of population linkage disequilibrium (LD) structure, we used pre-computed LD scores based on African and European 1000 Genomes Project Phase 3²⁰. SNPs in the major histocompatibility complex (MHC) region were excluded. Because of the high degree of genetic admixture in HAs and the smaller size of the sample, we did not estimate h^2_{SNP} in that population group.

We used LDSC to partition h^2_{SNP} in the OUD EA dataset and examined the enrichment of the partitioned h^2_{SNP} based on different functional genomic annotation models^{27,28}. In the baseline model, we examined 53 overlapping functional annotations comprising genomic, epigenomic, and regulatory features. Next, we analyzed 10 overlapping cell-type groups derived from 220 cell-type-specific annotations. Finally, enriched cell-type categories were analyzed based on annotations obtained from H3K4me1 imputed, gapped peak data generated by the Roadmap Epigenomics Mapping Consortium²⁹. For each h^2_{SNP} partitioning model, multi-allelic

and MHC region variants were excluded, and Bonferroni-correction was applied to identify significant enrichment.

Gene-based, Functional Enrichment and Pathway Analyses

We performed gene-based association testing for OUD in FUMA v1.3.6a³⁰, using MAGMA v1.08³¹, which employs multiple regression models to detect multiple marker effects that account for SNP p-values and LD between markers, using the matched-ancestry 1000 Genomes Project phase 3²⁰ panel as LD reference. We used a total of 18,707 protein-coding genes, with $p < 2.67 \times 10^{-6}$ ($0.05/18,707$) considered GWS.

To identify gene sets enriched for OUD, we used MAGMA³¹ to curate gene sets; Gene Ontology terms (obtained from MsigDB c2); and GWAS-catalog enrichment, correcting for gene size, variant density, and LD within and between genes. We also used MAGMA to test the association between gene-set properties and tissue-specific gene expression profiles using GTEx (v.7) data from 53 tissues (Bonferroni-corrected p-value threshold = 9.43×10^{-4}).

Transcriptome-wide Association Analyses

We performed transcriptome-wide association analyses using the MetaXcan framework³² and the GTEx release v.8 eQTL MASHR-M models³³. Forty-nine tissues from GTEx v.8 were analyzed comprising 12,951 samples. First, GWAS summary statistics were harmonized for the EA population based on the human genome assembly GRCh38 (hg38) and linked to the 1000 Genomes reference panel using GWAS tools (<https://github.com/hakyimlab/summary-gwas-imputation/wiki/GWAS-Harmonization-And-Imputation>). A transcriptome-wide association analysis of 49 tissues was run using S-PrediXcan³². A Bonferroni correction for statistical significance was applied within each tissue conditioned on the number of genes tested (Supplementary Table 14).

Because expression quantitative trait loci (eQTL) were correlated across tissues, we integrated gene expression signals for 49 tissue panels using S-MultiXcan³⁴ and tested 10,552 genes in total. Resulting p-values were Bonferroni corrected to identify significant gene associations (p-value threshold = 4.74×10^{-6}).

Drug Interactions

To identify drugs that could potentially be repurposed to treat OUD, we examined genes identified in the variant or gene-level analyses using the Drug Gene Interaction Database³⁵ (<https://www.dgidb.org>). Medications were categorized using the Anatomical Therapeutic Chemical (ATC) classification system, retrieved from the Kyoto Encyclopedia of Genes and Genomics Kyoto Encyclopedia of Genes and Genomics (KEGG; <https://www.genome.jp/kegg/drug/>).

Genetic Correlation

We used LD score regression²⁵ to calculate the r_g between a) OUD or OD datasets used for meta-analysis (AA [MVP, PGC, YP3]; EA [MVP, Partners HealthCare Biobank, PGC, YP3]) and b) OUD (MVP EA) and 40 other published traits comprising psychiatric, substance use, cognitive, and anthropometric traits selected based on a priori hypotheses (See Supplementary Table 20 for a full list), using pre-computed LD scores for HapMap3²⁶ SNPs based on the matched-ancestry 1000 Genomes Project Phase 3²⁰ reference panel. To explore additional traits in a hypothesis-free manner, we also estimated the r_g between OUD and 1,270 traits (comprising published and unpublished traits from the UK Biobank (UKBB) using the Complex-Trait Genetics Virtual Lab (CTG-VL) (<https://genoma.io>). CTG-VL integrates publicly available GWAS summary statistics and utilizes the LDSC framework to calculate r_g between complex traits and diseases of interest³⁶. A Bonferroni correction was applied within each LDSC and

CTG-VL analysis, and traits with a corrected p-value < 0.05 were regarded as significantly correlated.

Mendelian Randomization

We performed Mendelian randomization (MR) analysis using the *MendelianRandomization* package in R. Causal relationships between OUD and other traits were tested bidirectionally using three methods: Weighted Median, Inverse Variance Weighted and MR-Egger. We tested for pleiotropy using the MR-Egger intercept test. Instrumental variants were those associated with the exposure at $p < 1 \times 10^{-5}$. When the instrumental variants were not present in the outcome data, we identified the best-proxy variant ($LD > 0.8$). Variants with $MAF < 0.01$ or with no proxy with $LD > 0.8$ within 200 kb were removed. Each trait included more than 20 instrumental variables, which provides a robust estimate of causal effects. We considered causal effects as those for which at least 2 MR tests were significant after Bonferroni correction and that showed no evidence of violation of the horizontal pleiotropy test (MR-Egger intercept $p > 0.05$).

Polygenic Risk Scores and Phenome-wide Association Studies

We calculated PRS for OUD in two independent datasets (Yale-Penn and BioVU) using PRS-continuous shrinkage (PRS-CS)³⁷, followed by phenome-wide association analyses (PheWAS). In each dataset, OUD summary statistics from the matched ancestry were used to calculate PRS. Details for the analysis in each dataset are below.

Yale-Penn: We removed SNPs with INFO score < 0.7, $MAF < 0.01$ genotype call rate < 0.95, or an allele frequency difference between genotyping batches > 0.4, which left a total of 8,811,422 SNPs. We removed one individual from each pair of related individuals with $\pi\text{-hat} > 0.25$. To estimate genetic ancestry, we calculated PCs on common SNPs between Yale-Penn and 1000 Genomes Project Phase 3²⁰ using PLINK 1.9²¹. Subjects were assigned to an

ancestry based on the distance of 10 PCs from the 1000 Genomes reference populations. The resulting data set included 4,922 AAs and 5,709 EAs. We excluded binary phenotypes with fewer than either 100 cases or 100 controls, and continuous phenotypes with fewer than 100 individuals. We conducted PheWAS by fitting logistic regression models for binary traits and linear regression models for continuous traits. We used sex, age at recruitment, and the top 10 genetic PCs as covariates. We applied a Bonferroni correction to control for multiple comparisons.

BioVU: We used de-identified clinical data from Vanderbilt University Medical Center's (VUMC) Biobank (BioVU). Details on the quality control process have been described elsewhere³⁸. The genotyping information that we used was from the Illumina MEG^{EX} array. Genotypes were filtered for SNP and individual call rates, sex discrepancies, and excessive heterozygosity using PLINK v1.9²¹. Imputation of the autosomes was conducted using the Michigan Imputation Server¹⁹ based on the Haplotype Reference Consortium reference panel. PCA using FlashPCA2 combined with CEU, YRI and CHB reference sets from 1000 Genomes Project Phase 3²⁰ was conducted to determine participants of African and European Ancestry. The sample was then filtered for cryptic relatedness by removing one individual of each pair for which $\pi\text{-hat} > 0.2$. This resulted in 12,384 individuals of African ancestry and 66,903 individuals of European ancestry samples for analysis. We conducted PheWAS by fitting a logistic regression for each of the 1,335 disease phenotypes available in BioVU to estimate the odds of a diagnosis of that phenotype given the OUD PRS. Each disease phenotype (commonly referred to as "phecode"; <https://phewascatalog.org/phecodes>, Phecode Map 1.2) was classified using ICD 9 and 10 diagnostic codes to establish "case" status. For an individual to be considered a case, they were required to have two separate ICD codes for the index phenotype, and each phenotype needed at least 100 cases to be included in the analysis. The covariates included in the analyses were sex, median age of the longitudinal EHR measurements, and the

top 10 genetic PCs. The project was approved by the VUMC Institutional Review Board (IRB #s 160302, 172020, 190418).

Genomic Structural Equation Modeling

We performed Genomic Structural Equation Modeling³⁹ for OUD, 3 other substance use traits (problematic alcohol use, cannabis use disorder, ever smoked regularly), and 7 psychiatric disorders (schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, attention deficit hyperactivity disorder, Tourette's syndrome, and obsessive-compulsive disorder). We calculated a genetic covariance matrix using multivariable LDSC and the 1000 Genomes Project phase 3 European samples²⁰ as reference. An exploratory factor analysis was conducted using the genetic covariance matrix and a four latent-factor structure with varimax rotation. We used the determined structure containing paths with loading factor >0.2 to perform a confirmatory factor analysis implemented in the *GenomicsSEM* package in R. To prevent negative residual variance after estimation, we restricted the residual variance of OCD and ADHD to be greater than 0.

Results

Sample Description

Our MVP sample comprised 425,944 individuals (AA: 88,498; EA: 302,585; HA: 34,861), of whom 90.6% were male (Supplementary Table 1). The less stringent OUD definition yielded 28.8%-38.9% more cases across the ancestral groups (AA = 8,968, EA = 19,978, HA = 2,527) than the stringent definition (AA = 6,457; EA = 15,040; HA = 1,962). On average, less stringent cases had 77.2 (SD=96.9) opioid prescription fills, stringent cases had 76.5 (SD=97.6) fills, and controls had 25.0 fills (SD=48.3).

Identification of Novel Loci Associated with Opioid Use Disorder

The cross-ancestry meta-analysis of the less stringent OUD diagnosis within the MVP sample yielded 12 GWS variants, 10 of which were independent after conditioning on the lead variant within each locus (Figure 1, Supplementary Table 3). The protein-coding genes nearest these variants are *CDKAL1*, *BTNL2*, and *OPRM1* (all on chr. 6), *RABEPK* (chr. 9), *FBXW4* and *7SK* (chr. 10), *NCAM1* (chr. 11), *FURIN* (chr. 15), *KCNN1* (chr. 19), and *RNF114* (chr. 20). The most robust signal was in *OPRM1* (lead SNP rs1799971, $p=6.78 \times 10^{-10}$), which replicates the main finding of the previous MVP OUD GWAS¹³. The variant in *FURIN* is supported by prior findings at the variant¹⁴ and gene-based¹⁶ levels. In addition, there were 3 ancestry-specific loci (Supplementary Table 4): 1 each in AAs (*NNT*, chr. 5), EAs (*CDH8*, chr. 16), and HAs (*MRS2*, chr. 8). 3 loci were GWS in the cross-ancestry meta-analysis of the stringent OUD diagnosis in the MVP sample, including 1 additional locus (*TSNARE1*, chr. 8) which was also found in the EA-specific analysis of the stringent OUD diagnosis (Supplementary Tables 5 and 6), and the meta-analysis of EA subjects' data from MVP, Partners HealthCare Biobank, PGC and YP3 (Supplementary Table 8). GWS loci from all analyses are presented in Supplementary Tables 3-8. Based on all analyses, we identified a total of 14 GWS loci, 12 novel (Table 1).

SNP Heritability and Genetic Correlations Across GWAS Datasets

The estimated SNP heritability (h^2_{SNP}) in MVP for the less stringent phenotype is similar in AAs (0.11 ± 0.03) and EAs (0.12 ± 0.01). Using the stringent OUD phenotype, the estimated h^2_{SNP} was higher in AAs (0.20 ± 0.05) than EAs (0.15 ± 0.01). In contrast, the estimated h^2_{SNP} was higher in EAs (0.11 ± 0.01) than AAs (0.08 ± 0.03) in the ancestry-specific meta-analyses across datasets (Supplementary Table 9). Variation in these estimates appears to be mainly driven by changes in effective sample size, as estimates using actual sample size show little variation (Supplementary Table 9). The within-ancestry r_g between datasets is high, ranging from 0.7 (± 0.3) between the less stringent OUD MVP and Partners HealthCare Biobank datasets in EAs, to 1.2 (± 0.2) between the less stringent OUD MVP and the previous OUD MVP GWAS¹³ in EAs (which used the same diagnosis definition as the present stringent analysis) (Supplementary Table 10).

Considering the similarity in h^2_{SNP} between the different OUD GWAS and the greater number of loci captured by the less stringent diagnosis in MVP, all subsequent downstream analyses are based on the less stringent OUD GWAS within the MVP sample.

Partitioning Heritability Enrichment

We performed partitioning heritability enrichment analyses in LDSC²⁷ and examined heritability enrichment for gene expression using GTEx data⁴⁰. The most significantly enriched cell type group was the central nervous system (CNS; $p = 3.34 \times 10^{-3}$, Figure 2A, Supplementary Table 11). We observed significant enrichment for OUD in brain tissues only, including the anterior cingulate cortex ($p < 4.71 \times 10^{-6}$), limbic system ($p < 3.25 \times 10^{-5}$), prefrontal cortex ($p < 5.73 \times 10^{-5}$), cerebral cortex ($p < 9.81 \times 10^{-5}$), cortex ($p < 1.11 \times 10^{-4}$), hypothalamus ($p < 1.23 \times 10^{-4}$), amygdala ($p < 1.41 \times 10^{-4}$), and hippocampus ($p < 2.04 \times 10^{-4}$)

(Figure 2B, Supplementary Table 12). There were no significant enrichments for epigenetic annotations after correction for multiple testing (Supplementary Table 13).

Transcriptome-wide Analysis

We used S-PrediXcan³² to predict the effect of genetic variation on gene expression. Significant within-tissue gene expression regulation was identified for 43 tissues, including brain, adipose, gastrointestinal, thyroid, and liver (Supplementary Figure 2, Supplementary Tables 14 and 15). Significant associations with expression in brain tissues were detected for *FURIN*, *FES*, *LRP8*, *LINC01556*, *ZNF660* and *RP1-153G14.4* (Figure 2C). Some of these genes (*FURIN*, *LINC01556*, *ZNF660*, and *RP1-153G14.4*) were also expressed in non-brain tissues, such as adipose, gastrointestinal, and thyroid tissues (Supplementary Figure 2), suggesting that OUD-related genetic variation may exert significant transcriptomic changes in the periphery as well as the CNS.

Considering the sharing of eQTLs across multiple tissues, we tested the joint effects of variation in gene expression across tissues using S-MultiXcan³⁴. Significant transcriptomic effects for OUD were detected in 8 genes, 5 of which overlapped with genes detected by S-PrediXcan (*FURIN*, *FES*, *RP1-153G14.4*, *LRP8*, and *RABEPK*) and 3 which were novel (*ZNF391*, *ZKSCAN4*, and *MAGOH*) (Supplementary Table 16).

Gene Set, Functional Enrichment, and Drug Repurposing Analyses

Gene-based analyses identified one GWS gene in AAs (*CHRM2*, $p = 9.52 \times 10^{-7}$) and three GWS genes in EAs (*OPRM1*, $p = 2.17 \times 10^{-7}$; *FTO*, $p = 9.52 \times 10^{-7}$; *DRD2*, $p = 1.67 \times 10^{-6}$) (Supplementary Figure 3), but none in HAs. Following Bonferroni correction, no biological processes or pathways were significantly enriched, although nominal associations in EAs highlighted pathways of potential relevance, including “dopamine receptors” ($p = 1.87 \times 10^{-5}$)

and “regulation of adenylate cyclase activating G-protein coupled receptor signaling pathway” ($p = 4.39 \times 10^{-5}$) (Supplementary Table 17).

Genes identified in the variant-level, gene-based, or transcriptome (brain region) analyses (N=24) are summarized in Supplementary Table 18. Examination of these genes for drug-gene interactions via the Drug Gene Interaction database identified 761 interactions between 8 genes (*CHRM2*, *DRD2*, *FES*, *FURIN*, *KCNN1*, *NCAM1*, *OPRM1*, *PRL*) and 340 unique medications (Supplementary Table 19, Supplementary Figure 4). *OPRM1* had 193 interactions, mainly with classes of analgesics, anesthetics, and drugs for constipation. *DRD2* had 376 interactions, the majority of which were with psycholeptics.

Genetic Correlations

We estimated pairwise r_g with OUD for 40 published phenotypes using LDSC²⁵. OUD showed significant r_g with 21 traits. As expected, the strongest positive correlations were with substance use traits (e.g., problematic alcohol use, cannabis use disorder, ever smoked regularly) and psychiatric disorders (e.g., bipolar disorder, major depressive disorder) and the strongest negative correlation was with educational attainment (Figure 3A, Supplementary Table 20). We also assessed r_g with OUD for 1,270 complex traits from the UKBB using CTG-VL³⁶. After multiple testing correction ($p < 3.94 \times 10^{-5}$), 106 traits demonstrated significant genetic association with OUD (Supplementary Figure 5, Supplementary Table 21). These included positive correlations with substance use related traits (e.g. current smoking, ever addicted to any substance or behavior), psychiatric traits (e.g. anxiety treatment, self-reported depression) and pain-related traits (e.g. low back pain, multisite chronic pain), and negative correlations with having secondary education qualifications and the presence of social support. Thus, overall, we found that increased risk of OUD is genetically correlated with increased liability for use of substances, psychiatric disorders, and experiencing pain, and lower likelihood of educational attainment and social support.

Mendelian Randomization

Using MR, we tested for bidirectional causal effects between OUD and the 21 traits identified as significantly genetically correlated with OUD (Figure 3A, Supplementary Figure 6). There was a causal effect of OUD on 6 traits: problematic alcohol use, drinks per week, cannabis use disorder, general risk tolerance, MDD, and cross disorder. Among the 21 traits, 9 had a causal effect on OUD, of which 2 showed a negative causal effect on OUD (cognitive performance and educational level) and 7 a positive causal effect on OUD (in descending order of magnitude: drinks per week, worry subcluster, neuroticism, the number of sexual partners, major depressive disorder, cigarettes per day and schizophrenia).

Polygenic Risk Scores and Phenome-wide Association Studies

PRS were calculated in 2 independent datasets to identify phenotypic associations of genetic liability for OUD. In the Yale-Penn sample, PRS were calculated for 4,922 African ancestry and 5,709 European ancestry individuals. No significant associations were identified for AAs (Supplementary Figure 7, Supplementary Table 22). In EAs, PheWAS identified 41 phenotypes in the opiate domain and 71 phenotypes in other phenotypic domains that were significantly associated with OUD PRS (Figure 3C, Supplementary Table 23). The most significantly associated phenotypes were “time spent obtaining/using opioids” and “opioid use disorder”. In BioVU, PRS were calculated for 12,384 AAs and 66,903 EAs. No significant associations were found for OUD PRS in AAs (Supplementary Figure 8, Supplementary Table 24). In EAs, the OUD PRS was associated with 27 phenotypes, including “substance addiction and disorders” and “mood disorders” (Figure 3B, Supplementary Table 25).

Genomic Structural Equation Modeling

We conducted genomic structural equation modeling (gSEM) to evaluate how OUD relates to the three other substance use traits and the seven psychiatric disorders identified as the most significantly associated with OUD in genetic correlation analyses. Exploratory factor analysis (EFA) involving all 11 traits supported a 4-factor model with cumulative variance of 0.639. We retained paths with a loading factor ≥ 0.2 and conducted confirmatory factor analysis (CFA). In this analysis, the 4-factor model fit the data well (comparative fit index (CFI) = 0.948, Akaike information criterion = 340.840, $\chi^2 = 276.840$, degrees of freedom = 34, standard root mean square error (SRMR) = 0.073). The 4 substance use traits all loaded on Factor 1, with a major contribution from OUD (loading = 0.84 ± 0.05) and problematic alcohol use (loading = 0.91 ± 0.3), and lower contributions from cannabis use disorder (loading = 0.58 ± 0.06) and ever smoked regularly (loading = 0.40 ± 0.03). Cannabis use disorder (loading = 0.37 ± 0.06) and ever smoked regularly (loading = 0.42 ± 0.03), together with other psychiatric disorders, also loaded on Factor 3. Major psychiatric disorders, including bipolar disorder (loading = 0.86 ± 0.04), schizophrenia (loading = 0.76 ± 0.03), and MDD (loading = 0.43 ± 0.03) loaded on Factor 2. Tourette's syndrome (loading = 0.33 ± 0.07) and obsessive-compulsive disorder (loading = 1.03 ± 0.21) loaded on Factor 4 (Figure 4, Supplementary Table 26).

Discussion

This study, the largest single-sample GWAS of OUD to date, identified 14 loci associated with the disorder, 12 of which are novel findings. Three of these loci were significant in ancestry-specific analyses only, demonstrating that inclusion of diverse ancestral samples in genetic studies of OUD permits the identification of novel genetic variants. Post-GWAS analyses in EAs revealed enrichment for OUD in the CNS, particularly the brain, and an extensive phenotypic spectrum associated with genetic liability for OUD.

Because the effect sizes of common variants contributing to highly polygenic phenotypes such as OUD are small, large sample sizes are required to identify GWS loci. The largest OUD GWAS prior to the current study greatly increased the effective sample size ($N_{\text{effective}} = 88,115$) by meta-analyzing the results of studies that used a range of case and control definitions¹⁶. Here, we performed GWAS using the stringent definition of OUD used by Zhou et al.¹³ ($N_{\text{cases}} = 23,459$, $N_{\text{effective}} = 88,569$) and a less stringent definition requiring the presence of only 1 ICD-9/10 diagnostic code for opioid abuse or dependence ($N_{\text{cases}} = 31,473$, $N_{\text{effective}} = 116,590$). Although the less stringent definition lowers the specificity of the case phenotyping (i.e., individuals are more likely to be mislabeled as having OUD), it increases the number of cases by more than 8,000 and reveals 8 more GWS variants than the stringent definition. These results support prior conclusions that the potential variability introduced by broadening phenotypic definitions in genetic studies of OUD is outweighed by substantial increases in sample size¹⁶; however owing to clinical differences between substance use disorders and other psychiatric disorders, this may not be generalizable. In contrast, our meta-analysis of the MVP data with other datasets reduced the number of GWS loci identified, potentially because the smaller additional datasets increased the variability in the effect size of variants, reducing associations overall.

The most significant locus, *OPRM1*, encodes the mu-opioid receptor, which binds morphine and other opioids and has been the focus of many functional and candidate gene studies of opioid-related phenotypes^{41–43}. In a previous GWAS comprised principally of subjects from MVP, OUD was significantly associated only with *OPRM1* in EAs¹³, with the lead SNP being the non-synonymous, exon 1 variant rs1799971 (A118G). In neither that study, nor the present study, was the SNP associated with OUD in AAs, presumably because the minor (G) allele frequency in this population group is considerably lower than in EAs⁴⁴. Even so, it is difficult to explain why meta-analysis with AAs does not increase the statistical strength of the association of OUD with this variant if it is truly the lead functional variant, even if based on introgressed EA alleles alone. A recent meta-analysis of opioid addiction in European-ancestry individuals, which included data from the previous MVP OUD GWAS¹³ and 4 other cohorts also identified *OPRM1* as a risk gene, although the lead SNP at that locus was rs9478500 in intron 1¹⁶.

We identified a second peak in *OPRM1*, with the lead SNP rs3778151, a variant in intron 1 that is in high LD with rs9478500 ($r^2 = 0.56-0.90$)⁴⁵, the variant associated with opioid addiction in the recent meta-analysis¹⁶. Although LD clumping in our sample showed these two loci to be independent, when we conditioned the analysis on rs1799971, rs3778151 was no longer significant, suggesting that the variants are structurally independent, but functionally redundant. A prior study in EA alcohol- or drug-dependent cases and controls also identified two independent LD blocks in *OPRM1*⁴⁶. The findings were interpreted as a partial explanation of the inconsistent findings for *OPRM1*, particularly rs1799971, in candidate gene studies of alcohol and drug use disorders⁴⁶. Our findings suggest that although LD clumping may differentiate the two *OPRM1* loci, a more definitive test of independence – namely, a conditional analysis – fails to show their independence, consistent with the presence of a single locus at *OPRM1*.

Our cross-ancestry analysis showed an association between a SNP in *FURIN* and OUD, with transcriptome-wide analyses showing significant downregulation of gene-expression for *FURIN* in brain-related tissues. These findings support the reported associations of OUD with *FURIN* both in gene-based analyses^{14,16} and in a variant-level meta-analysis of OUD¹⁴. *FURIN* is associated with schizophrenia⁴⁷ and harbors a schizophrenia-associated cis-eQTL⁴⁸. Although *FURIN* encodes a protease that cleaves some endogenous opioids, notably proenkephalin⁴⁹, the enzyme has not previously been linked to the effects of exogenous opioids or mu-opioid receptor signaling. Given these findings, further research on the mechanism underlying the gene's effects on risk of OUD is warranted.

Our analysis also identified 12 novel GWS loci. Two of these, which are in *RABEPK* and *NCAM1*, were GWS in a multi-trait analysis using MTAG of OUD with cannabis use disorder and alcohol use disorder¹⁴. Here, we show associations directly with OUD. *RABEPK* is adjacent to *PPP6C*, a gene previously implicated in a gene-level analysis of OUD¹⁶ that has also been linked to reward-related phenotypes like obesity and smoking^{50,51}. Whereas SNPs within *RABEPK* contributed to the *PPP6C* signal in a previous study⁴⁵, our lead SNP could be tagging functional variants in *PPP6C*. Alternatively, our study included multiple ancestral groups, and 50% more OUD cases, which could have affected the location of the GWAS peak and enabled us to localize the peak more accurately than in⁴⁵. *NCAM1* and *KCNN1* have been implicated in the neuropharmacology of opioid-related phenotypes. The mouse homolog of *KCNN1* is differentially expressed in the nucleus accumbens following chronic morphine exposure⁵². The gene is also downregulated in the rodent prelimbic cortex after exposure to cues associated with morphine withdrawal⁵³, suggesting a connection to learning and memory. *NCAM1* also appears to be involved in the response to morphine exposure. Tolerance in rodents due to repeated morphine injection can be prevented by treatment with an antisense oligodeoxynucleotide that targets *Ncam1*⁵⁴. *NCAM1* variants have also been significantly

associated with alcohol dependence⁵⁵, drug dependence⁵⁶, smoking initiation⁵¹, cannabis use⁵⁷, and alcohol consumption⁵¹.

Several other loci identified in our cross-ancestry meta-analysis contain GWAS hits for other traits, suggesting the possibility of widespread pleiotropy of loci associated with OUD. *CDKAL1* and *BTNL2* have been associated with metabolic traits such as type 2 diabetes, body mass index⁵⁸ and obesity-related phenotypes⁵⁰. *FBXW4* and *CDH8* have prior associations with cognitive traits such as educational attainment and mathematical ability⁵⁹, and *TSNARE1* has a prior association with schizophrenia⁶⁰.

Partitioning heritability enrichment analyses showed that CNS cells were the only significantly enriched group. We found significant enrichment for OUD in brain tissues only, including the anterior cingulate cortex, limbic system, prefrontal cortex, hypothalamus, amygdala, and hippocampus, regions previously associated with the underlying neurobiology of the disorder⁶¹. These findings underscore the neural basis of OUD and reinforce the conceptualization of substance use disorders, which are often chronic and relapsing, as brain diseases. This notion was novel when proposed nearly 25 years ago⁶² and although today it is a view widely held by neuroscientists and clinicians, it is not universally understood by politicians or the general public. Improving our understanding of the biological basis of OUD could promote a science-based response to the opioid epidemic.

Consistent with prior findings, OUD showed strong genetic correlations with multiple substance use disorders, psychiatric disorders, cognitive traits, and risk behavior^{13,15,16,63}. Mendelian randomization analyses demonstrated causal effects of OUD on problematic alcohol use and cannabis use disorder, and a bidirectional causal effect of drinks per week. Given the high frequency with which substance use disorders co-occur, and the causal associations shown here, treatment efforts that aim broadly to reduce substance use are recommended: reducing use of substances comorbid with OUD would, according to this model, also reduce use of opioids. Genetic liability for psychiatric traits, including neuroticism and schizophrenia, was

also causally associated with OUD, with a bidirectional causal effect of MDD on OUD. Our findings, along with those of others¹³, suggest that OUD has a common biological pathway with schizophrenia and MDD. Despite the significant genetic correlations and causal associations between OUD and psychiatric disorders, genomic structural equation modelling indicated a common genetic factor representing broad genetic liability for substance use disorders that is distinct from those underlying the psychiatric disorders. The factor structure among psychiatric disorders seen here is consistent with previous findings⁶⁴ and shows that cannabis use disorder and smoking, unlike OUD, load onto both the substance use disorder factor and the factor underlying MDD, ADHD, autism spectrum disorder, and Tourette's syndrome.

PheWAS of the genetic liability for OUD in the Yale-Penn sample, which was ascertained for substance use disorders, reproduced the broad association with other substance use. In a clinical dataset using EHR data, the genetic liability for OUD was associated with multiple traits in every phenotypic domain tested, demonstrating the widespread effects of OUD liability on bodily systems. Some of the associations may be due to phenotypic correlation. For instance, associations were found with viral hepatitis and human immunodeficiency virus (HIV), potential proxies of injection drug use, and with chronic pain and back pain, potential proxies for the use of analgesic medications. Negative associations with obesity, type 1 diabetes, and skin cancer may reflect under reporting or diagnosis in individuals with OUD, or they could also reflect true biological relationships.

Limitations to the present study should be noted. Although it includes AA, EA, and HA individuals, participants of European ancestry comprise more than 60% of the total sample, which in large part drove the results of the cross-ancestry analysis. This disparity in sample size is also reflected in analyses of the individual ancestral groups, in which the smaller AA and HA groups provided less statistical power and yielded fewer significant loci. The lower power of the AA GWAS is also reflected in the lack of associations in PRS analysis in AAs. Future GWAS of OUD should focus on expanding sample sizes for non-European-ancestral populations to

capture loci that are relevant to specific population groups. The sample for this study is >90% male, reflecting the sex distribution of Veterans in the United States. Risk variants relevant only to women may thus have been overlooked due to the lower statistical power. Because the MVP dataset lacks information on the initiation of opioid use among patients diagnosed with OUD, we could not differentiate patients who developed the disorder only after being prescribed opioid analgesics from those whose first opioid use involved recreational use of analgesics or heroin. Differences in the initiation of opioid use could reflect different genetic risk factors contributing to non-overlapping intermediate phenotypes (e.g., pain threshold/susceptibility in analgesic use vs. risk taking in recreational use).

In summary, we have identified 14 genetic loci associated with OUD, the majority novel. Many of the loci contain genes with prior associations with substance use or psychiatric disorders, suggesting widespread pleiotropy. The use of a less stringent definition of OUD allowed a 25% larger number of OUD cases than a stringent definition in the MVP sample. Downstream analyses validate this approach by demonstrating plausible enrichment of OUD in brain regions, genetic correlations with other substance use disorders and psychiatric disorders, and association between OUD PRS and opioid dependence in an independent sample. Our findings provide insight into the biological underpinnings of OUD, which could inform preventive, diagnostic, and therapeutic efforts and thereby help to address the opioid epidemic.

Figure legends

Figure 1: Manhattan and quantile-quantile plot for cross-ancestry meta-analysis of OUD (N case=31,480, N control=394,484). Effective sample size weighted meta-analyses were performed in METAL. The nearest protein-coding gene (<1Mb) in each locus is labelled.

Figure 2: Enrichment of OUD in the brain. A. Partitioning heritability enrichment analyses using LDSC show enrichment for OUD in the central nervous system. B. Heritability enrichment for gene expression using GTEx data show enrichment for OUD in brain regions previously associated with addiction. C. Predicted gene expression using S-PrediXcan identify genes with differential expression in brain regions.

Figure 3: Phenotypic spectrum associated with OUD. A. Genetic correlation analyses show multiple traits significantly genetically correlated with OUD (red bar – positively correlated, blue bar – negatively correlated). Mendelian randomization analyses identify causal associations between OUD and other traits (arrows, red – positive causal association, blue – negative causal association). B and C. PheWAS results in BioVU (B) and Yale-Penn (C) datasets. All phenotypes significant at FDR $p < 0.05$ are plotted. In B, all phenotypes which pass Bonferroni correction are labelled. For readability, in C, only the top 3 traits within each group which pass Bonferroni correction are labelled. Circle size denotes effect size.

Figure 4: Genomic SEM analysis of OUD with other substance use traits (OUD: opioid use disorder; PAU: problematic alcohol use; CUD: cannabis use disorder; SMK: ever smoked regularly) and psychiatric disorders (SCZ: schizophrenia; BIP: bipolar disorder; MDD: major depressive disorder; ASD: autism spectrum disorder; ADHD: attention deficit hyperactivity disorder; TS: Tourette's syndrome; OCD: obsessive compulsive disorder). Four factors were identified. Factor loadings for each trait are depicted by arrows between the trait and the factor. Correlation between factors is indicated by arrows between the factors. Residual variance for each trait is indicated by the U-circles. Standard errors are depicted in parentheses.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Association, 2013). doi:10.1176/appi.books.9780890425596.
2. Strang, J. *et al.* Opioid use disorder. *Nat. Rev. Dis. Primer* **6**, 3 (2020).
3. Vowles, K. E. *et al.* Rates of opioid misuse, abuse, and addiction in chronic pain: a systematic review and data synthesis. *PAIN* **156**, 569–576 (2015).
4. Centers for Disease Control and Prevention (CDC). Vital signs: overdoses of prescription opioid pain relievers and other drugs among women--United States, 1999-2010. *MMWR Morb. Mortal. Wkly. Rep.* **62**, 537–542 (2013).
5. Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health. <https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR090120.htm>.
6. Wilson, N. Drug and Opioid-Involved Overdose Deaths — United States, 2017–2018. *MMWR Morb. Mortal. Wkly. Rep.* **69**, (2020).
7. Kendler, K. S., Jacobson, K. C., Prescott, C. A. & Neale, M. C. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am. J. Psychiatry* **160**, 687–695 (2003).
8. Gelernter, J. *et al.* Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. *Biol. Psychiatry* **76**, 66–74 (2014).

9. Cheng, Z. *et al.* Genome-wide Association Study Identifies a Regulatory Variant of RGMA Associated With Opioid Dependence in European Americans. *Biol. Psychiatry* **84**, 762–770 (2018).
10. Nelson, E. C. *et al.* Evidence of CNH3 involvement in opioid dependence. *Mol. Psychiatry* **21**, 608–614 (2016).
11. Polimanti, R. *et al.* Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium. *Mol. Psychiatry* **25**, 1673–1687 (2020).
12. Song, W. *et al.* Genome-wide association analysis of opioid use disorder: A novel approach using clinical data. *Drug Alcohol Depend.* **217**, 108276 (2020).
13. Zhou, H. *et al.* Association of OPRM1 Functional Coding Variant With Opioid Use Disorder: A Genome-Wide Association Study. *JAMA Psychiatry* **77**, 1072–1080 (2020).
14. Deak, J. D. *et al.* Genome-wide association study and multi-trait analysis of opioid use disorder identifies novel associations in 639,709 individuals of European and African ancestry. 2021.12.04.21267094 <https://www.medrxiv.org/content/10.1101/2021.12.04.21267094v1> (2021) doi:10.1101/2021.12.04.21267094.
15. Sanchez-Roige, S. *et al.* Genome-wide association study of problematic opioid prescription use in 132,113 23andMe research participants of European ancestry. *Mol. Psychiatry* 1–9 (2021) doi:10.1038/s41380-021-01335-3.
16. Gaddis, N. *et al.* Multi-trait genome-wide association study of opioid addiction: OPRM1 and Beyond. 2021.09.13.21263503

<https://www.medrxiv.org/content/10.1101/2021.09.13.21263503v1> (2021)

doi:10.1101/2021.09.13.21263503.

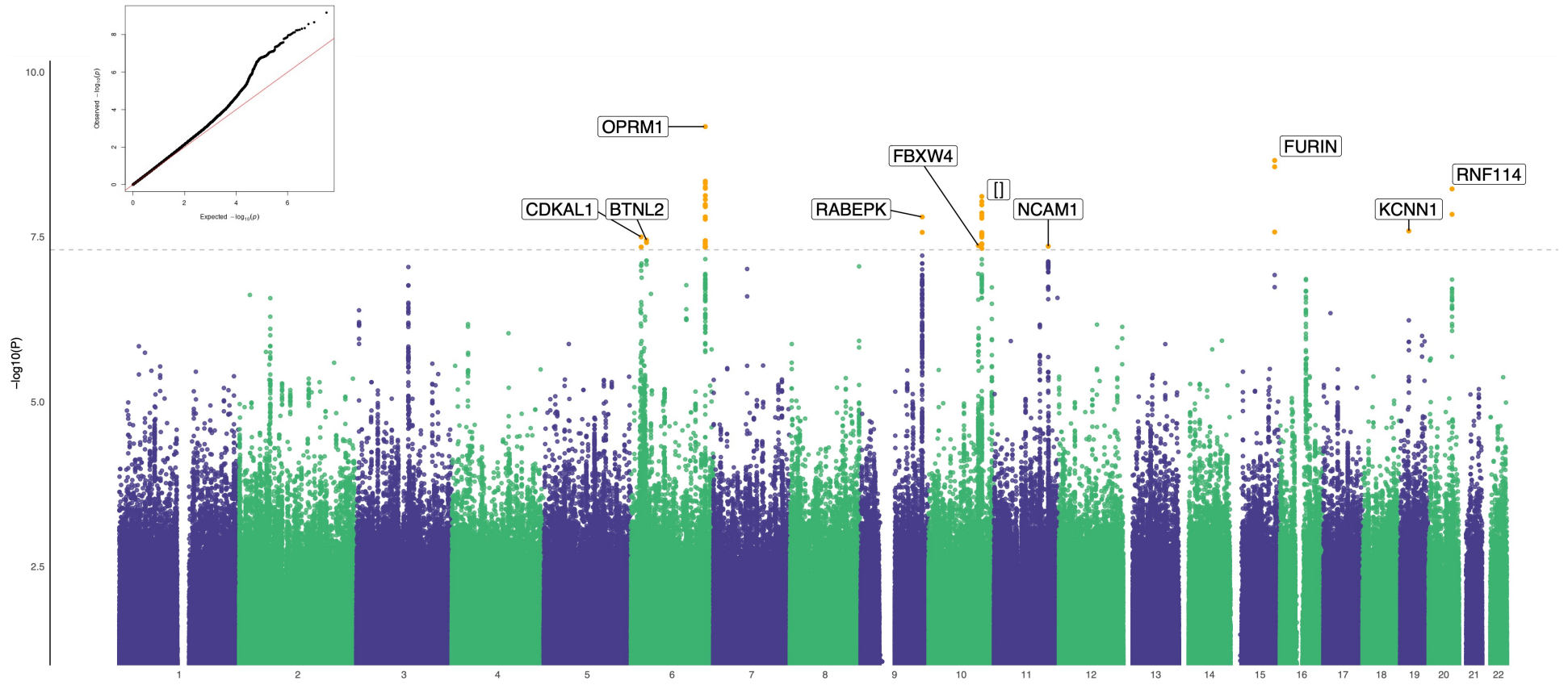
17. Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
18. Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
19. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
20. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
21. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, (2015).
22. Fang, H. *et al.* Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am. J. Hum. Genet.* **105**, 763–772 (2019).
23. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma. Oxf. Engl.* **26**, 2190–2191 (2010).
24. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
25. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
26. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).

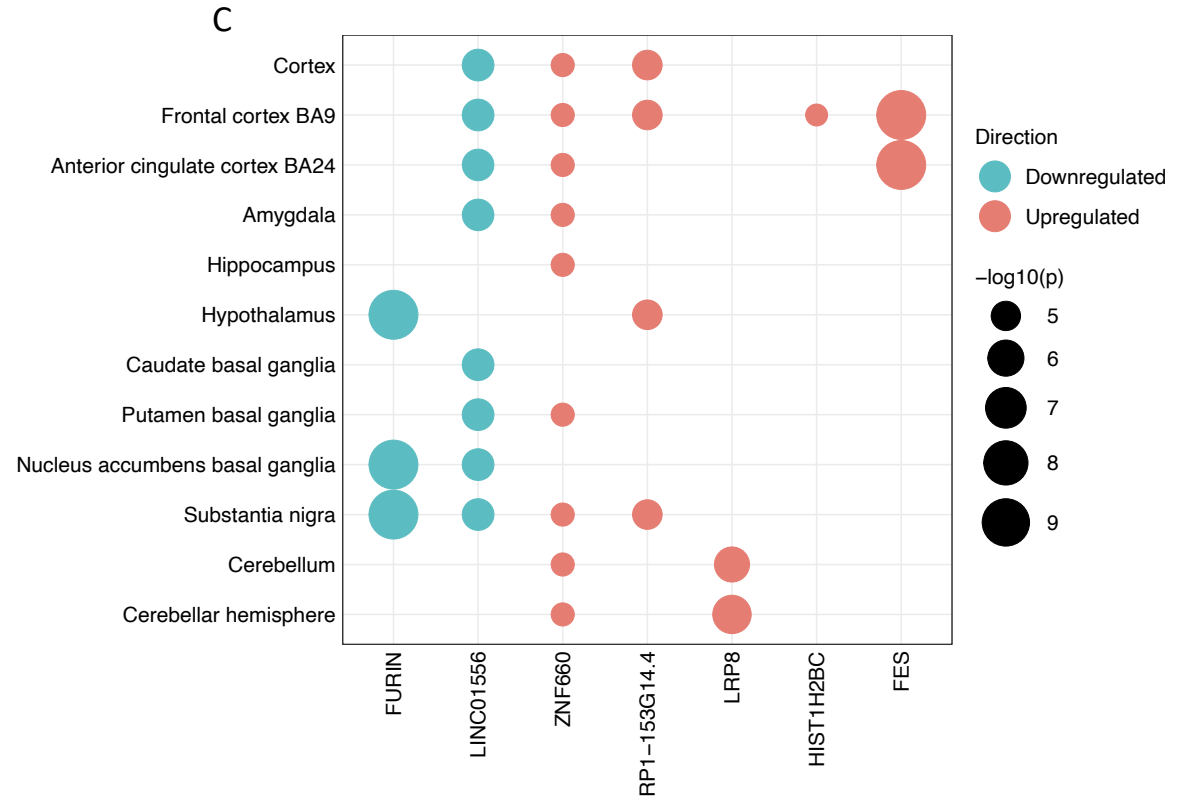
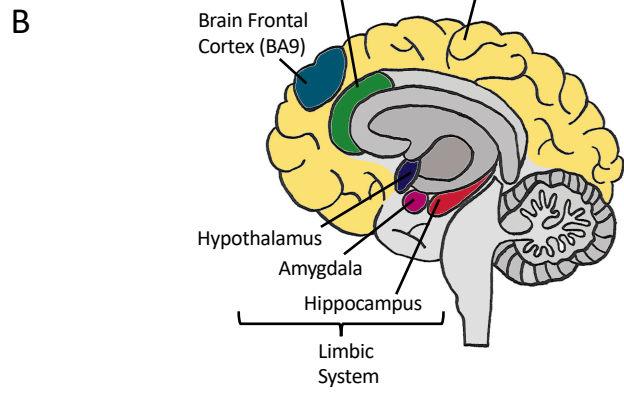
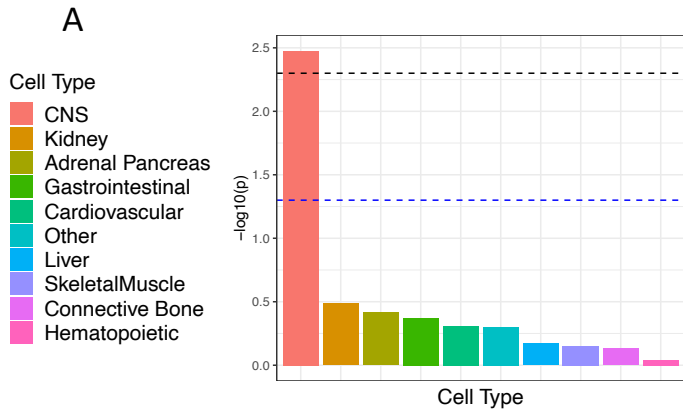
27. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
28. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
29. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
30. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
31. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
32. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
33. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **22**, 49 (2021).
34. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLOS Genet.* **15**, e1007889 (2019).
35. Freshour, S. L. *et al.* Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsourcing efforts. *Nucleic Acids Res.* **49**, D1144–D1151 (2021).
36. Cuellar-Partida, G. *et al.* *Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses*. <http://biorxiv.org/lookup/doi/10.1101/518027> (2019)
doi:10.1101/518027.

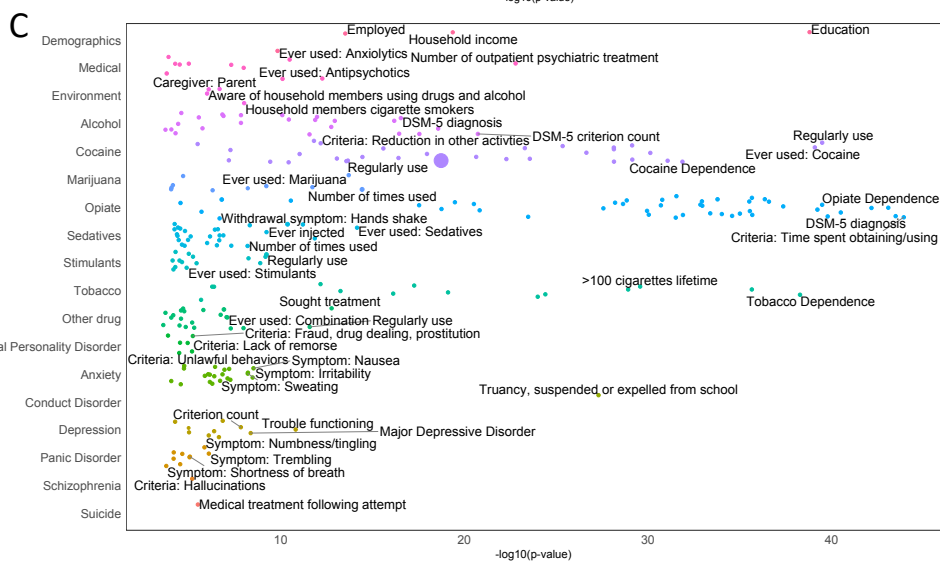
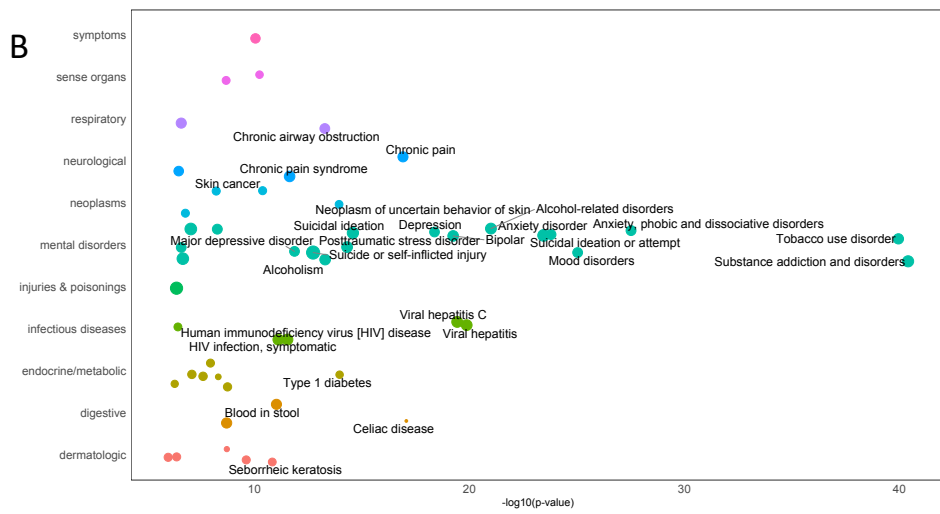
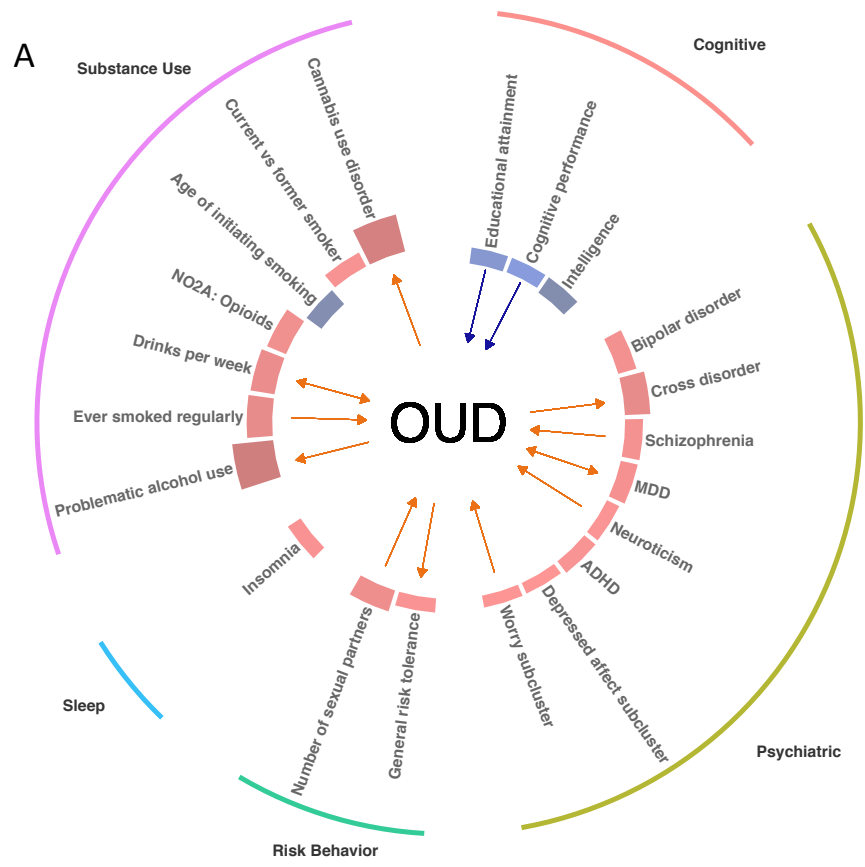
37. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
38. Dennis, J. K. *et al.* Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. *Genome Med.* **13**, 6 (2021).
39. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
40. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
41. Crist, R. C. & Berrettini, W. H. Pharmacogenetics of OPRM1. *Pharmacol. Biochem. Behav.* **123**, 25–33 (2014).
42. Crist, R. C., Reiner, B. C. & Berrettini, W. H. A review of opioid addiction genetics. *Curr. Opin. Psychol.* **27**, 31–35 (2019).
43. Moninka, H., Lichenstein, S. & Yip, S. W. Current Understanding of the Neurobiology of Opioid Use Disorder: an Overview. *Curr. Behav. Neurosci. Rep.* **6**, 1–11 (2019).
44. Gelernter, J., Kranzler, H. & Cubells, J. Genetics of two mu opioid receptor gene (OPRM1) exon I polymorphisms: population studies, and allele frequencies in alcohol- and drug-dependent subjects. *Mol. Psychiatry* **4**, 476–483 (1999).
45. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
46. Zhang, H. *et al.* Association between two μ -opioid receptor gene (OPRM1) haplotype blocks and drug or alcohol dependence. *Hum. Mol. Genet.* **15**, 807–819 (2006).
47. Biological insights from 108 schizophrenia-associated genetic loci | Nature. <https://www.nature.com/articles/nature13595>.

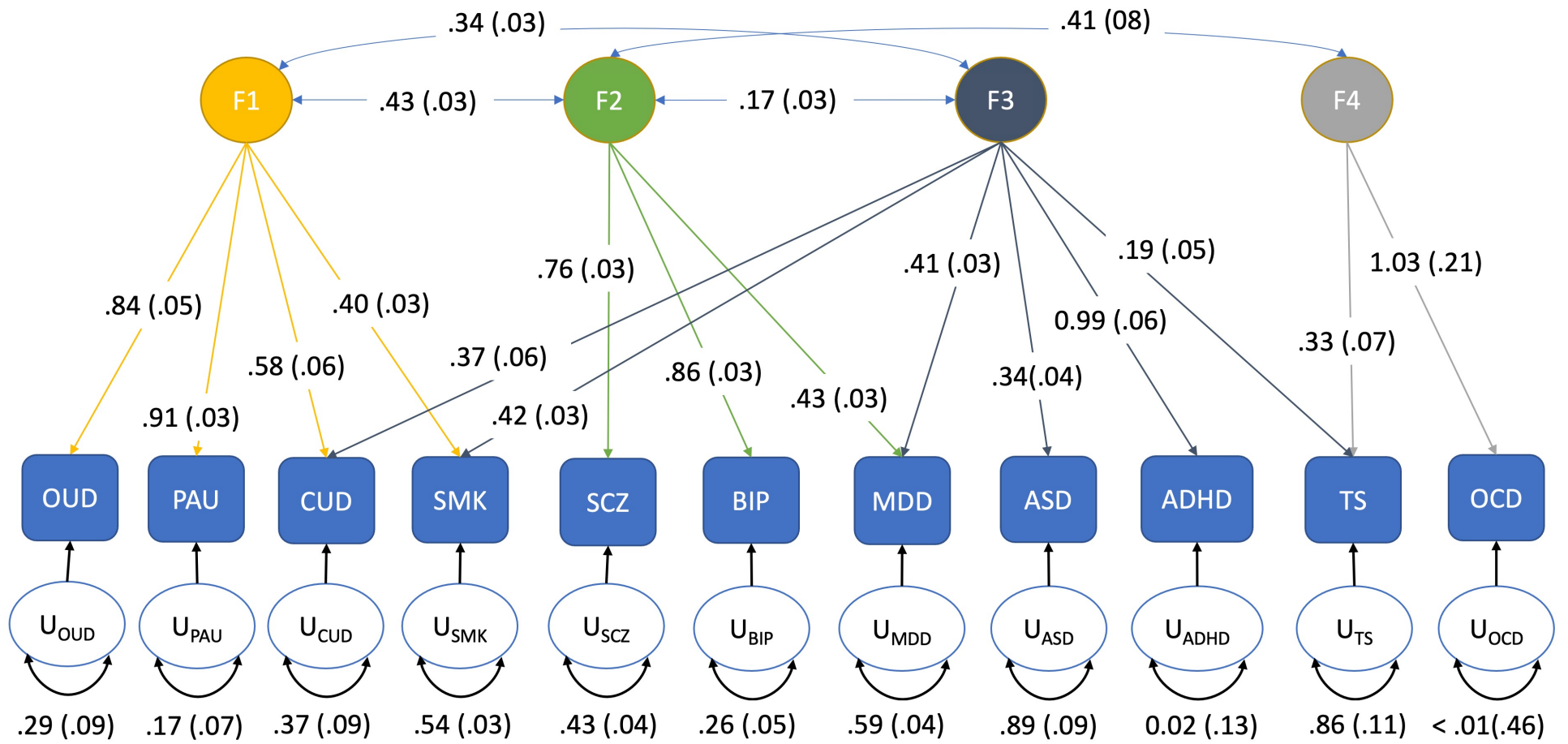
48. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
49. Breslin, M. B. *et al.* Differential processing of proenkephalin by prohormone convertases 1(3) and 2 and furin. *J. Biol. Chem.* **268**, 27084–27093 (1993).
50. Christakoudi, S., Evangelou, E., Riboli, E. & Tsilidis, K. K. GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer. *Sci. Rep.* **11**, 10688 (2021).
51. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
52. Grice, D. E. *et al.* Transcriptional profiling of C57 and DBA strains of mice in the absence and presence of morphine. *BMC Genomics* **8**, 76 (2007).
53. Yu, L. *et al.* Activity in projection neurons from prelimbic cortex to the PVT is necessary for retrieval of morphine withdrawal memory. *Cell Rep.* **35**, 108958 (2021).
54. Fujita-Hamabe, W., Nakamoto, K. & Tokuyama, S. Involvement of NCAM and FGF receptor signaling in the development of analgesic tolerance to morphine. *Eur. J. Pharmacol.* **672**, 77–82 (2011).
55. Yang, B.-Z. *et al.* Association of haplotypic variants in DRD2, ANKK1, TTC12 and NCAM1 to alcohol dependence in independent case control and family samples. *Hum. Mol. Genet.* **16**, 2844–2853 (2007).
56. Yang, B.-Z. *et al.* Haplotypic variants in DRD2, ANKK1, TTC12, and NCAM1 are associated with comorbid alcohol and drug dependence. *Alcohol. Clin. Exp. Res.* **32**, 2117–2127 (2008).

57. Pasma, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21**, 1161–1170 (2018).
58. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**, 680–691 (2020).
59. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
60. Amare, A. T. *et al.* Bivariate genome-wide association analyses of the broad depression phenotype combined with major depressive disorder, bipolar disorder or schizophrenia reveal eight novel genetic loci for depression. *Mol. Psychiatry* **25**, 1420–1429 (2020).
61. Koob, G. F. Neurobiology of Opioid Addiction: Opponent Process, Hyperkatifeia, and Negative Reinforcement. *Biol. Psychiatry* **87**, 44–53 (2020).
62. Leshner, A. I. Addiction is a brain disease, and it matters. *Science* **278**, 45–47 (1997).
63. Karlsson Linnér, R. *et al.* Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat. Neurosci.* **24**, 1367–1376 (2021).
64. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469-1482.e11 (2019).









Chr	Position (GRCh37/hg19) of lead SNPs	Nearest genes	GWAS analysis
5	43846681	<i>NNT</i>	AA (Less stringent)
6	21362610, 21478361	<i>CDKAL1, SOX4</i>	Cross-ancestry (Less stringent), HA (Less stringent), Cross-ancestry (Meta)
6	32383573	<i>BTNL2</i>	Cross-ancestry (Less stringent)
6	154360797, 154380719, 154382139, 154393680, 154396472	<i>OPRM1</i>	Cross-ancestry (Less stringent), EA (Less Stringent), Cross-ancestry (Stringent), EA (Stringent), Cross-ancestry (Meta), EA (Meta)
6	24394925	<i>MRS2</i>	HA (Less stringent)
8	143312933, 143316970	<i>TSNARE1</i>	Cross-ancestry (Stringent), EA (Stringent), EA (Meta)
9	127873473, 127959540, 127980426	<i>SCAI, RABEPK</i>	Cross-ancestry (Less stringent), Cross-ancestry (Meta), EA (Meta)
10	103414885	<i>FBXW4</i>	Cross-ancestry (Less stringent), EA (Less Stringent), EA (Meta)
10	110504365	[]	Cross-ancestry (Less stringent)
11	112869404	<i>NCAM1</i>	Cross-ancestry (Less stringent)
15	91410009, 91406146, 91426560	<i>FURIN</i>	Cross-ancestry (Less stringent), EA (Less Stringent), Cross-ancestry (Stringent), EA (Stringent), Cross-ancestry (Meta), EA (Meta)
16	61631362	<i>CDH8</i>	EA (Less stringent), EA (Stringent)
19	18093588	<i>KCNN1</i>	Cross-ancestry (Less stringent), AA (Less stringent), Cross-ancestry (Meta), AA (Meta)
20	48540277, 48583726	<i>RNF114</i>	Cross-ancestry (Less stringent), Cross-ancestry (Meta)