2-28-2021

# Genome-wide association study of phenotypes measuring progression from first cocaine or opioid use to dependence reveals novel risk genes

Richard Sherva

Congcong Zhu

Leah Wetherill

Howard J Edenberg

Emma Johnson

*See next page for additional authors*

## Authors

Richard Sherva, Congcong Zhu, Leah Wetherill, Howard J Edenberg, Emma Johnson, Louisa Degenhardt, Arpana Agrawal, Nicholas G Martin, Elliot Nelson, Henry R Kranzler, Joel Gelernter, and Lindsay A Farrer

# Genome-wide association study of phenotypes measuring progression from first cocaine or opioid use to dependence reveals novel risk genes

Richard Sherva[1]*, Congcong Zhu[1], Leah Wetherill[2], Howard J. Edenberg[2,3], Emma Johnson[4], Louisa Degenhardt[5], Arpana Agrawal[4], Nicholas G. Martin[6], Elliot Nelson[4], Henry R. Kranzler[7], Joel Gelernter[8,9], Lindsay A. Farrer[1,10,11,12,13] (iD)

[1]Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA 02118, USA
[2]Department of Medical and Molecular Genetics and Biochemistry, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[3]Department of Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[4]Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO 63110, USA
[5]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, NSW 2052, Australia
[6]Queensland Institute of Medical Research Berghofer, Brisbane, QLD 4006, Australia
[7]Perelman School of Medicine, University of Pennsylvania and VISN 4 MIRECC, Crescenz VAMC, Philadelphia, PA 19104, USA
[8]Departments of Psychiatry, Genetics and Neuroscience, Yale School of Medicine, New Haven, CT 06511, USA
[9]Department of Psychiatry, VA CT Healthcare Center, West Haven, CT 06516, USA
[10]Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA
[11]Department of Ophthalmology, Boston University School of Medicine, Boston, MA 02118, USA
[12]Department of Epidemiology, Boston University School Public Health, Boston, MA 02118, USA
[13]Department of Biostatistics, Boston University School Public Health, Boston, MA 02118, USA

**\*Correspondence:** Richard Sherva, Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA 02118, USA. sherva@bu.edu

## Abstract

**Aim:** Substance use disorders (SUD) result in substantial morbidity and mortality worldwide. Opioids, and to a lesser extent cocaine, contribute to a large percentage of this health burden. Despite their high heritability, few genetic risk loci have been identified for either opioid or cocaine dependence (OD or CD, respectively). A genome-wide association study of OD and CD related phenotypes reflecting the time between first self-reported use of these substances and a first DSM-IV dependence diagnosis was conducted.

**Methods:** Cox proportional hazards regression in a discovery sample of 6,188 African-Americans (AAs) and 6,835 European-Americans (EAs) participants in a genetic study of multiple substance dependence phenotypes were used to test for association between genetic variants and these outcomes. The top findings were tested for replication in two independent cohorts.

**Results:** In the discovery sample, three independent regions containing variants associated with time to dependence at $P < 5 \times 10^{-8}$ were identified, one (rs61835088 = $1.03 \times 10^{-8}$) for cocaine in the combined EA-AA meta-analysis in the gene *FAM78B* on chromosome 1, and two for opioids in the AA portion of the sample in intergenic regions of chromosomes 4 (rs4860439, $P = 1.37 \times 10^{-8}$) and 9 (rs7032521, $P = 3.30 \times 10^{-8}$). After meta-analysis with data from the replication cohorts, the signal at rs61835088 improved (HR = 0.87, $P = 3.71 \times 10^{-9}$ and an intergenic SNP on chromosome 21 (rs2825295, HR = 1.14, $P = 2.57 \times 10^{-8}$) that missed the significance threshold in the AA discovery sample became genome-wide significant (GWS) for CD.

**Conclusions:** Although the two GWS variants are not in genes with obvious links to SUD biology and have modest effect sizes, they are statistically robust and show evidence for association in independent samples. These results may point to novel pathways contributing to disease progression and highlight the utility of related phenotypes to better understand the genetics of SUDs.

## Keywords

## Introduction

The opioid epidemic continues to cause human and economic devastation in the United States and elsewhere [1]. U.S. deaths due to the misuse of opioids reached a record high of more than 42,000 in 2016, which prompted the U.S. Department of Health and Human Services to declare opioid abuse to be a public health emergency in 2017. Several factors led to the epidemic, including the over-prescription of opioid analgesics, increased availability of heroin and illicit synthetic opioids [2], and possibly progressively worsening labor market opportunities, especially for those with low levels of education [3]. Despite early interventions to curb over-prescription, the epidemic continued to worsen [4]. Cocaine dependence (CD) has received less attention recently but its U.S. prevalence (likely significantly underreported [5]) remains at about 1%, resulting in a significant public health burden [6]. Cocaine users experience mortality rates four to eight times higher than the general population [7] and have an increased risk of suicide [8]. In contrast to opioid dependence (OD), pharmacological treatments for CD do not exist [9].

Moderate to high heritability estimates for both OD (43-50%) [10] and CD (65-78%) [11, 12] strongly suggest a genetic component to risk. Several OD risk genes have been identified through modestly-sized genome-wide association studies (GWAS) [13-16], but they have not been replicated and explain only a tiny fraction of the total trait heritability. Similarly, CD risk genes have been identified and there is evidence of genetic overlap between CD and other psychiatric disorders [17, 18], but collectively these results explain little trait heritability. Two relatively large OD GWAS papers have been published recently [Polimanti et al. [16] (4,503 OD cases) and Zhou et al. [19] (28,317 OD cases)]. These studies identified a single genome-wide significant (GWS) association for OD: a coding variant in the μ-opioid receptor gene *OPRM1* (opioid receptor mu 1 gene). The sample sizes available for these studies still lag far behind those available for other complex diseases with a similar public health impact, and no large scale GWAS for CD has been performed. This problem is exacerbated by the fact that controls exposed to illicit drugs (a prerequisite for developing OD or CD) represent a small subset of the total control population. They may also be limited by the lack of differentiation between dependence upon prescription opioids and heroin, which may have a different set of risk factors.

In light of the deficits in our understanding of the genetic factors contributing to OD and CD risk, we have sought to derive phenotypes that are under more direct genetic control that could facilitate understanding the disorders and identification of potential drug targets and risk pathways. Our group recently identified variants near *OPRM1* associated with usual methadone dosage in a sample of individuals with OD [20]. Here, we report findings from a GWAS of two related phenotypes: the time between first reported use and first DSM-IV diagnosis of OD and CD. We analyzed each substance independently. Both cocaine and opioids are highly addictive [21]. Approximately 50% of individuals who ever use heroin develop OD [22], while about

four percent of people who try cocaine become addicted within two years, with another 16% in a prodromal stage of addiction [23]. A few studies (one conducted in the primary cohort analyzed here) have examined risk factors associated with rapid progression to dependence. Conduct disorder and childhood physical abuse predicted rapid more development of OD and CD, while and alcohol and nicotine dependence diagnoses were associated with slower progression, and African Americans (AAs) progressed to OD more rapidly than European Americans (EAs) [24].

There is also an interplay between use of medically indicated or recreationally used prescription opioids or heroin in determining who develops OD. Prescription opioids used for pain relief are generally safe when taken for a short time and as prescribed by a doctor. In a U.S. veteran population, individuals who used prescription opioids recreationally had 19 times the odds of initiating heroin use, fewer than four percent of people who abuse prescription opioids initiate using heroin within five years [25]. Despite our inability to distinguish between disease course with regard to specific opioids and their pattern of initiation, these related phenotypes have two properties that make them attractive for variant discovery: namely (1) they do not include individuals who were never exposed to opioids or cocaine and (2) they may be better indicators of an individual's genetic risk for OD/CD than case-control status because they distinguish between cases who develop the disorder slowly (i.e. are more genetically resistant) and those who develop it rapidly (i.e. are highly genetically susceptible), assuming these differences are partially under genetic control.

## Materials and methods

### Participants and diagnostic procedures

Subjects for the study were recruited from three sources. The Yale-Penn discovery sample includes 6,188 AAs and 6,835 EAs participants who were ascertained for genetic studies of dependence on opioids, cocaine or alcohol between 2000 and 2017 by advertisement and through treatment settings at Yale University School of Medicine, University of Pennsylvania, University of Connecticut Health Center, the Medical University of South Carolina, University of Pennsylvania, and McLean Hospital in Belmont, Massachusetts [26, 27]. Psychiatric interviews were conducted using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) [28, 29].

A replication sample of EAs, informative for opioids but not cocaine, was derived from the Comorbidity and Trauma Study (CATS) [30, 31]. Briefly, opioid-dependent cases age 18 or older were recruited from opioid agonist treatment (OAT) clinics in metropolitan Sydney, Australia. Persons who had recent suicidal intent or psychosis were excluded. Controls were recruited from neighborhoods geographically proximal to the cases and excluded those who used opioids recreationally more than five times. Subjects were interviewed using the Semi-Structured Assessment for the Genetics of Alcoholism-Australia (SSAGA-OZ) [32] to derive DSM-IV substance use disorders (SUDs) diagnoses.

A second independent sample used for replication came from the Collaborative Study on the Genetics of Alcoholism (COGA), a large family-based study that recruited alcohol-dependent AA and EA probands from treatment facilities across seven sites in the United States [33, 34]. Probands and their families were invited to participate. Additional individuals and their families were recruited from the same communities. All participants were administered a version of the SSAGA [32, 35] which also queried about illicit drugs including opioids and cocaine.

Institutional review boards at all sites for all three studies approved the protocol and consent forms, and all participants provided written, informed consent according to the Helsinki Declaration Code of Ethics.

### Phenotype definition

Analyses for cocaine and opioids were conducted independently; separate GWAS were performed for each outcome. Although individuals exposed to both opioids and cocaine contributed to both analyses, cocaine exposure/dependence did not inform analyses of opioids and vice versa. Time-to-dependence (TD) was defined as the interval between self-reported age at first opioid/cocaine use and the age at first self-reported endorsement of three or more DSM-IV OD/CD criteria within the same month. Individuals who used opioids or

cocaine at least once but never became dependent were censored at their age at interview. No differentiation was made between exposure to or dependence on prescription *vs.* illicit opioids, but the majority (65%) of users in the Yale-Penn sample reported heroin as their most heavily used opioid.

### Genotyping, imputation and quality control (QC)

DNA specimens in the Yale-Penn sample were genotyped on three microarrays: Yale-Penn 1 on the Illumina HumanOmni1-Quad v1.0 microarray (OMNI), Yale-Penn 2 on the Illumina Infinium Human Core Exome microarray (HCE), and Yale-Penn 3 on the Illumina Multi-ethnic Global Array (MEGA). QC of genotype data was performed as previously described [13]. Briefly, individuals and single nucleotide polymorphisms (SNPs) with a call rate < 98% and variants with minor allele frequency (MAF) < 1% were excluded prior to imputation. Pairwise identity-by-decent (IBD) was calculated with PLINK to determine genetic relatedness among individuals in the sample and individuals with a pairwise IBD estimate > 25% were assigned to the same family. Self-reported males with X chromosome heterozygosity > 20% and self-reported females with X chromosome heterozygosity < 20% were excluded. SNP genotype imputation was performed using the March 2012 1,000 Genomes reference panel (1,000 Genomes Project, 2012; http://www.1000genomes.org/) and IMPUTE2 [36] separately in AAs and EAs implemented on the Michigan imputation server (https://imputationserver.sph.umich.edu).

DNA specimens from the CATS sample were genotyped at the Center for Inherited Disease Research (CIDR) using the Illumina Human660W-Quad BeadChip. Genotype and sample QC was as described previously [9]. After initial QC, data were imputed to the 1,000 Genomes Phase 3 European ancestries reference panel using the Michigan Imputation Server. Details on the genotyping QC, and imputation of the COGA samples are described in detail elsewhere [37]. Briefly, samples were genotyped on multiple arrays in multiple labs. A subset of 47,000 common, independent, and high-quality SNPs that were genotyped across all arrays were used to assess duplicate samples, confirm the reported pedigree structure and compute ancestral principal components (PCs). After assignment of individuals in a family to a specific population, family-wise ancestry was designated according to the majority of individual family members. Genotypes were imputed to 1,000 Genomes using the cosmopolitan reference panel (Phase 3, version 5, NCBI GRCh37) using SHAPEIT2 [38] and Minimac3 [39].

After applying QC filters and phenotypic criteria, the Yale-Penn discovery sample included in analyses contained 1,307 AA and 2,340 EA OD cases and 974 AA and 768 EA opioid-exposed controls and 3,554 AA and 2,712 EA CD cases and 478 AA and 915 cocaine-exposed controls. The CATS replication sample contained 1,217 EA OD cases and 88 opioid-exposed controls. Demographic information and mean TD or censoring in the three cohorts are shown in Tables 1 (CD) and 2 (OD). The COGA replication sample included 144 AA and 334 EA OD cases and 354 AA and 1,305 opioid-exposed controls, 572 AA and 759 EA CD cases and 416 AA and 1,620 EA cocaine-exposed controls.

**Table 1.** Demographic information on opioid-exposed individuals contributing to analysis of the trait time to opioid use disorder

| Cohort | AA | | | | EA | | | |
|---|---|---|---|---|---|---|---|---|
| | Age μ (SD) | Sex N (%) male | OD cases (N, mean years to dependence) | OD controls (N, mean years to censoring) | Age μ (SD) | Sex N (%) male | OD cases (N, mean years to dependence) | OD controls (N, mean years to censoring) |
| Yale-Penn | 42.2 (9.0) | 1,500 (66) | 1,307 (3.6) | 974 (14.9) | 36.7 (10.8) | 1,977 (64) | 2,340 (3.8) | 768 (13.9) |
| COGA | 56.6 (13.7) | 325 (65) | 144 (4.3) | 354 (13.8) | 49.7 (14.6) | 1,047 (64) | 334 (3.4) | 1,305 (13.6) |
| CATS | NA | NA | NA | NA | 36.5 (8.6) | 778 (59.6) | 1217 (3.8) | 88 (10.5) |

**Table 2.** Demographic information on cocaine exposed individuals contributing the time to cocaine use disorder analysis

| Cohort | AA | | | | EA | | | |
|---|---|---|---|---|---|---|---|---|
| | Age μ (SD) | Sex N (%) male | CD cases (N, mean years to dependence) | CD controls (N, mean years to censoring) | Age μ (SD) | Sex N (%) male | CD (N, mean years to dependence) | CD controls (N, mean years to censoring) |
| Yale-Penn | 42.5 (8.3) | 2,419 (60) | 3,554 (5.5) | 478 (17.8) | 38.0 (11.1) | 2,237 (62) | 2,712 (4.9) | 915 (16.3) |
| COGA | 56.9 (10.1) | 568 (57) | 572 (4.6) | 416 (13.8) | 52.1 (13.1) | 1380 (58) | 759 (3.9) | 1,620 (13.3) |

## Statistical analysis

OD and CD TD analyses were performed separately and without considering an individual's exposure or dependence status for the other drug. Analyses were stratified by genotyping platform and genetically determined ancestry and combined with inverse variance weighted meta-analysis. Variants with *P*-values < $1.0 \times 10^{-5}$ were tested in the replication samples. Association tests were performed using Cox proportional hazards regression [40] implemented in the R package Survival (https://cran.r-project.org/web/views/Survival.html). Cases had a lifetime DSM-IV OD or CD diagnosis and controls did not, although they could have endorsed two or fewer of the dependence criteria. All controls were exposed to cocaine or opioids (at least one reported use, including prescription) in their respective analyses. Imputed allele dosage was the predictor variable and models were adjusted for age (in COGA and CATS only, ten-year-age cohorts in COGA), sex, and the first five PCs (calculated within each ancestry group). The cluster option was used to account for the presence of related individuals in the samples by generating robust variance estimates. OD and CD TD analyses were performed separately and without considering an individual's status on exposure to or dependence on the other drug. Analyses in the Yale-Penn and COGA samples were stratified by population and genotyping chip (in Yale-Penn) and all results were combined via inverse variance weighted meta-analysis using METAL (https://genome.sph.umich.edu/wiki/METAL) [41]. Analyses were stratified by cohort and ancestry (where applicable) and combined with inverse variance weighted meta-analysis. Variants with *P*-values < $1.0 \times 10^{-5}$ were tested in the replication samples. The proportional hazards assumption was checked by verifying that the Schoenfeld residuals for each term in the model were independent of time. Both age at interview and age at fist cocaine/opioid use were significantly associated with TD in Yale-Penn but both violated the proportional hazards assumption and were not included in the final analysis in Yale-Penn. The top results were very similar with and without each age term in the model, but including an age by time interaction term attenuated them substantially.

The Yale-Penn results for CD and OD in EAs, AAs, and the combined meta-analyses were uploaded to the Functional Mapping and Annotation (FUMA) of GWAS portal [42], which performs functional mapping, gene based tests, and gene set enrichment analyses using GWAS summary statistics. Regional Manhattan plot were generated using Locuszoom software [43].

# Results

Across all cohorts, the mean TD was shorter for opioids than cocaine (Tables 1 and 2). In Yale-Penn, older age at interview was strongly associated with longer TD for both CD [hazard ratio (HR) = 0.97 per year, *P* = $3.51 \times 10^{-97}$] and OD (HR = 0.98 per year, *P* = $3.02 \times 10^{-27}$). EA ancestry was associated with a longer TD for CD (HR = 0.79, *P* = $3.70 \times 10^{-18}$) and a shorter TD for OD (HR = 1.43, *P* = $1.20 \times 10^{-21}$). Female sex was associated with longer TD for CD (HR = 0.97, *P* = 0.046) and shorter TD for OD (HR = 1.12, *P* = 0.001). Shorter TD was significantly correlated with a higher number DSM-IV dependence criteria endorsed (i.e. more severe dependence), with Pearson correlation coefficients of 0.54 and 0.53 in AAs and EAs, respectively, for OD and 0.44 and 0.48 in AAs and EAs, respectively, for CD. The distribution of DSM-4 CD and OD criteria counts in Yale-Penn are shown in Supplemental Figure 1.

Boxplots for the distributions of age at onset and the TD for each substance in Yale-Penn are shown in Supplemental Figure 2. There were 4,989 individuals exposed to both cocaine and opioids that contributed to both analyses. There were 2,908 individuals with a lifetime DSM-4 dependence diagnosis for both drugs.

## Discovery GWAS results

In the discovery sample, we identified tone region containing variants associated with TD at a GWS level (*P* < $5 \times 10^{-8}$) in the gene family with sequence similarity 78-member B (*FAM78B*) on chromosome 1 (rs61835088, *P* = $2.96 \times 10^{-8}$) for CD in the combined EA and AA sample. Each copy of minor (C) alleles at rs61835088 was associated with 0.57 fewer years between first use and CD diagnosis among those who converted. The Manhattan plots for CD and OD in Yale-Penn AAs and EAs are shown in Supplemental Figure 3. There was no evidence for inflation of the test statistics for either substance in either EAs or AAs (Supplemental Figure 4).

## Gene and gene set analyses

No significant gene-based tests were observed after correcting for the number of valid gene-based test results, which varied by population and dependence outcome. The top ranked gene ($P = 4.5 \times 10^{-6}$) in gene based tests was calcium voltage-gated channel subunit alpha1 B (*CACNA1B*) in the EA CD analysis. One gene set (membrane depolarization) showed a significant enrichment, with six of 83 genes, including *CACNA1B*, showing nominally significant gene based test results after Bonferroni correction ($P = 0.04$) for CD in EAs.

## Replication results

Three hundred and fifty one variants with suggestive *P*-values < $1.0 \times 10^{-5}$ in the discovery sample (149 for CD, and 202 for OD) were tested in the replication samples. Several of these did not yield a valid result due to low MAF or imputation quality in their respective ancestry group or substance. Valid results were obtained for 142 CD SNPs and 200 OD SNPs in COGA and 170 OD SNPs in CATS. After correcting for the number of tests performed, none of the GWS hits in the discovery met criteria for replication. Several variants were at or near nominal significance with the same effect direction as Yale-Penn in COGA/CATS and improved the association signal after meta-analysis. The strongest signal in any of the replication samples was in the OD analysis with rs8063946 ($P = 3.70 \times 10^{-4}$ in COGA AAs) in the gene FTO alpha-ketoglutarate-dependent dioxygenase (*FTO*).

## Discovery + replication results

After combining all results, the association of rs61835088 with CD TD strengthened (HR = 0.87, $P = 3.71 \times 10^{-9}$, Table 3, Figure 1). For the same outcome, the association with intergenic chromosome 21 SNP rs2825295 in AAs surpassed the GWS threshold in the combined Yale-Penn/COGA meta-analysis ($P = 2.19 \times 10^{-8}$, Table 3, Figure 1). Each minor (T) allele at rs2825295 was associated with 0.43 fewer years between first cocaine use and dependence among those who converted. The top variant in regions with *P*-values < $1.0 \times 10^{-6}$ after combining all data are shown in Tables 3 (CD) and 4 (OD in AAs). For OD TD, all but one of the top associations was in the AA portion of the sample. There was one OD TD signal in the combined EA and AA sample, rs8063946 ($P = 1.87 \times 10^{-7}$) in the gene *FTO*, which for clarity is not shown in Table 4. Near-GWS evidence of association was obtained in the combined samples with SNPs in three other genes previously associated with addiction phenotypes: rs114341823 in *GRIN2B* [44, 45], which encodes the glutamate ionotropic receptor NMDA type subunit 2B was associated with OD TD in AAs ($P = 1.45 \times 10^{-7}$); the variant in *FTO* [46-48] described above; and rs73721103, which is between alpha-aminoadipic semialdehyde synthase (*AASS*) and *LOC102724527* but is ~100 kilobases from *PTPRZ1* [49, 50], which encodes the protein tyrosine phosphatase receptor type Z1 with CD TD in AAs ($P = 2.81 \times 10^{-7}$).

**Table 3.** Top associations for time to CD

| Chr | BP | rsID | A1 | A2 | Gene | Group | Yale-Penn | | | COGA | | Yale-Penn + COGA | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | MAF | AA *P* | EA *P* | AA + EA *P* | AA *P* | EA *P* | HR | *P* | Direction |
| 1 | 166038842 | rs61835088 | T | C | *FAM78B* | All | 0.14 | 5.19 E-08 | 2.55 E-02 | 2.96 E-08 | 0.89 | 0.01 | 0.87 | 3.71 E-09 | -------- |
| 5 | 170894319 | rs112894747 | C | G | *FGF18-LOC105377721* | AA | 0.03 | 9.01 E-06 | NA | 9.01 E-06 | 0.01 | NA | 1.41 | 4.76 E-07 | +-++ |
| 7 | 90978616 | rs74426341 | C | G | *LOC105375392* | All | 0.04 | 1.46 E-06 | 5.70 E-07 | 1.46 E-06 | 0.08 | 0.62 | 1.37 | 2.52 E-07 | ?+- +++++ |
| 7 | 90983464 | rs73404786 | A | G | *LOC105375392* | EA | 0.04 | 2.75 E-01 | 5.20 E-08 | 4.19 E-07 | 0.08 | 0.74 | 1.38 | 2.37 E-07 | ++++ |
| 7 | 121807926 | rs73721103 | T | G | *AASS-LOC102724527* | AA | 0.05 | 6.57 E-01 | NA | 2.76 E-03 | 0.11 | NA | 1.30 | 4.00 E-07 | ++++ |
| 21 | 20404237 | rs2825295 | T | G | *AL157359.3-AP000431.1* | AA | 0.35 | 1.45 E-07 | 3.15 E-02 | 1.04 E-07 | 0.07 | 0.83 | 1.14 | 2.57 E-08 | ++++ |

The order of the symbols corresponds to Yale-Penn 1, Yale Penn 2, Yale-Penn 3, COGA. If AAs and EAs are combined, the AA symbol precedes the EA symbol. A1: effect allele; A2: non-effect allele; MAF: minor allele frequency in the subset of the Yale-Penn cohort indicated in the Group column; *P*: *P*-value; Direction: effect direction of A1
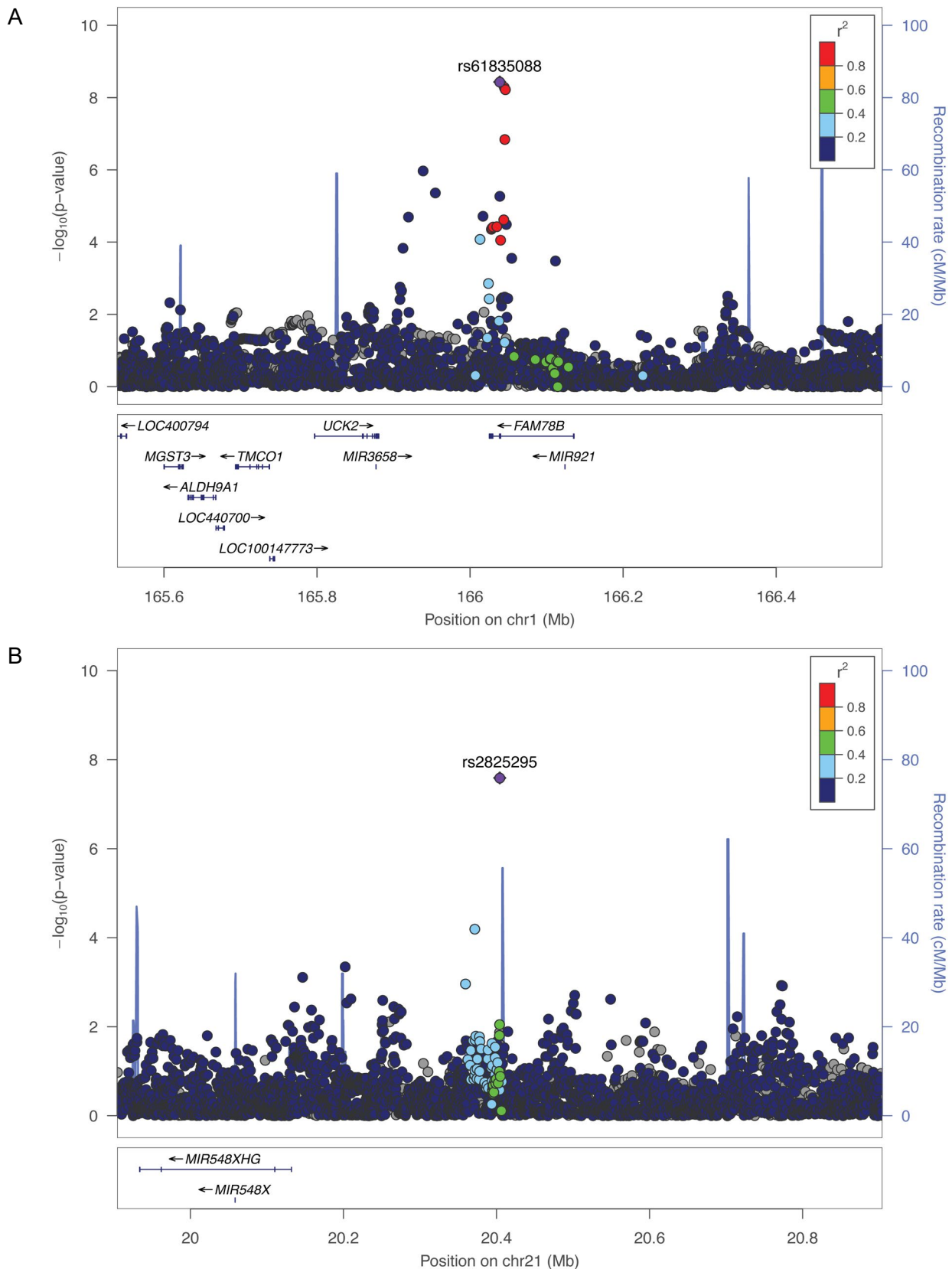
**Figure 1.** Regional Manhattan plots for the *FAM78B* region on chromosome 1 (A) and the intergenic region on chromosome 21 (B) for time to CD in AAs in the Yale-Penn cohort. SNPs are color coded according to the correlation coefficient (*r²*) in the 1,000 Genomes African reference panel with the top-ranked SNP. Light blue line indicates the observed recombination rate (right-side y-axis)

## Functional annotation

None of the variants we identified result in amino acid changes and any function they might have is likely regulatory, or as a result of linkage to a functional variant. Several of the top CD TD SNPs showed evidence for regulatory potential in various databases. According to the Genotype Tissue Expression database (https://

gtexportal.org/home/) rs61835088 is a significant eQTL for *FAM78B* in tibial nerve tissue. This variant was also an eQTL for that gene in prefrontal cortex and blood according to QTLbase (http://mulinlab.org/qtlbase) [51]. Several variants, including rs61835088, rs73404786 were significant methylation QTLs (mQTL) in prefrontal cortex. Rs73721103 was a significant eQTL for the gene ring finger protein 133 (*RNF133*) in blood. Only one OD TD SNP showed evidence for regulatory potential: rs11728570 as a significant mQTL in blood. Neither SNP Function Prediction (https://snpinfo.niehs.nih.gov/) nor Braineac (http://braineacv2.inf.um.es/) provided any additional information about potential functionality of the top results.

**Table 4.** Top associations for time to OD in AAs

| Chr | BP | rsID | A1 | A2 | GENE | Yale-Penn | | COGA | Yale-Penn + COGA | | |
| | | | | | | MAF | *P* | *P* | HR | *P* | Direction |
|-----|-----|------|----|----|------|-----|-----|------|-----|------|-----------|
| 2 | 8653036 | rs113142991 | A | C | *LINC00299-LINC0184* | 0.07 | 1.81E-07 | 8.30E-01 | 1.57 | 6.48E-07 | +++- |
| 4 | 59826282 | rs4860439 | T | C | *RP11-506N2.1* | 0.31 | 1.02E-07 | 6.50E-01 | 0.80 | 1.06E-07 | ---- |
| 4 | 95747400 | rs11728570 | T | C | *BMPR1B* | 0.37 | 1.08E-05 | 4.90E-02 | 0.82 | 1.07E-06 | ---- |
| 9 | 118707243 | rs115391335 | T | C | *LINC00474-LOC105376236* | 0.03 | 5.55E-07 | 1.80E-01 | 1.72 | 1.11E-07 | +++- |
| 10 | 49574826 | rs1881736 | A | T | *MAPK8* | 0.04 | 1.47E-05 | 7.60E-03 | 0.63 | 3.73E-07 | --+- |
| 12 | 14086651 | rs114341823 | T | C | *GRIN2B* | 0.03 | 8.75E-07 | 9.20E-02 | 1.70 | 1.77E-07 | ++++ |
| 12 | 53305274 | rs114198947 | A | G | *KRT8* | 0.03 | 2.68E-06 | 8.80E-01 | 1.67 | 7.35E-07 | ++++ |

The order of the symbols corresponds to Yale-Penn 1, Yale Penn 2, Yale-Penn 3, COGA. A1: effect allele; A2: non-effect allele; MAF: minor allele frequency in the AA subset of the Yale-Penn cohort; *P*: *P*-value; Direction: effect direction of A1

# Discussion

Here we present the results from GWASs of CD and OD related phenotypes measuring the time interval between first use of cocaine or opioids and a DSM-IV dependence diagnosis on the respective drug. These analyses, made possible by the extensive phenotyping performed on three independent cohorts, identified two GWS associations, one specific to AAs and one in the full sample, between relatively common variants and enhanced susceptibility to CD that were not detected at GWS in GWAS that used the bivariate case-control status for dependence as outcomes. Both of these were strengthened after combining meta-analysis with the replication sample. To our knowledge, this is the first time these trait definitions have been studied using GWAS. Although none of the GWS associations in the discovery sample were replicated after correcting for the number of variants tested in the COGA and CATS cohorts, 13 (9%) of the CD TD SNPs and 11 (5.5%) of the OD TD SNPs were nominally significant in one of those samples with the same effect direction.

The function of *FAM78B*, associated with progression from cocaine use to dependence, is not well understood but it is highly expressed in the brain [52] and interacts with several other proteins with diverse functions [53]. According to the STRING protein-protein interaction database (as https://string-db.org) [54], the proteins with which it is predicted to interact include ones related to mitochondrial function (ATP5F1), neurofilament network integrity (SNCG), and NOTCH signaling (XXYLT1).

Other genes that were among the top hits, but not GWS, have functions with potential links to SUDs. The genes *GRIN2B* and *FTO*, associated with progression from opioid use to dependence; and *PTPRZ1*, associated with cocaine TD have all previously been linked to SUD traits. Different variants than the one identified in this study in *GRIN2B* have been associated with OD [45] and chronic ketamine use [44]. Its protein product directly binds calcium/calmodulin-dependent protein kinase 2-alpha (encoded by *CAMK2A*), which can lead to synaptic long-term potentiation by facilitated CAMK2 response to synaptic calcium [55]. We previously identified this pathway as an important modulator of OD risk [13] and this result strengthens the evidence that an individual's propensity to establish and reinforce substance-specific neural circuits may be an important factor driving their genetic predisposition to SUDs, and that calcium homeostasis may at least partially drive that propensity. *FTO* has also been associated with addictive behaviors. Although initially identified as a regulator of eating and obesity traits [56, 57], though probably not through an effect of FTO itself but rather as a consequence of variants in the gene that affect expression of *IRX3* and *IRX5* [56, 58].

*FTO* has also been associated with alcohol dependence [46-48] and connectivity in a dopamine-dependent reward circuit of meso-striato-prefrontal regions of the human brain [59]. *PTPRZ1* binds to two neurotrophic cytokines [pleiotrophin (PTN) and midkine (MK)] that, along with other functions, contribute to the extinction of cocaine and amphetamine-seeking behaviors [49, 50] and limit morphine withdrawal syndrome [60]. The variants we identified, however, were ~100 KB away from the gene and may not strongly suggest a role for this gene in SUD risk. The top gene identified through gene based tests, *CACNA1B* in EAs for CD TD, encodes a pore-forming subunit of the presynaptic neuronal voltage-dependent calcium channel. Genes from this family of neuronal regulators of calcium and potassium levels were associated with their respective traits in our previous OD and CD GWAS papers [13, 18]. This, along with the fact that the membrane polarization gene set was significantly enriched among our results provides further evidence for the role of genetically determined differences in synaptic membrane potential acting as mediators of addiction risk.

We also looked up the results for the top TD SNPs in our previously published CD [18] and OD [13] GWAS of case-control status where controls were exposed to cocaine or opioids. There was modest evidence for association of each TD GWS variant identified with risk of dependence (not TD) on the respective substances, although not at the GWS level. Rs61835088 ($P$ = 0.007) and rs2825295 ($P$ = 0.001) were associated with binary CD diagnosis, and rs4860439 ($P$ = 1.11 × $10^{-6}$) and rs7032521 (4.10 × $10^{-7}$) with OD diagnosis. The *OPRM1* OD variant (rs1799971) identified by Polimanti et al. [16] and Zhou et al. [19] was nominally associated with OD TD in Yale-Penn EAs ($P$ = 0.05).

### Limitations

The primary limitation of this work is the relatively small sample size compared to other GWAS of complex disease. Also, the Yale-Penn sample had exclusion criteria for major psychiatric comorbidities including schizophrenia and suicidal ideation, which could have led to the exclusion of people with severe, possibly highly 'genetically-driven' CD or OD. Third, we did not detect GWS associations with our TD phenotypes with the top genes identified through the corresponding binary trait analyses, or vice versa. While we hypothesize that we may be detecting biologically distinct pathways and/or achieved better power (given the sample sizes) by using survival models, an alternative hypothesis is that our models were more prone to error and false positives.

In conclusion, although our findings were derived from a relatively small sample compared to those used for GWAS of other complex psychiatric diseases, we identified significant associations in genes never before implicated in SUD risk, one of which, in contrast to the top findings from published dependence diagnosis GWAS papers, showed evidence for association in both AA and EA ancestry groups. Although the effect size is small and the effect of the variant has little if any implication for the prediction or treatment of SUDs, it may point to a novel SUD-relevant pathway. Measuring gene and gene network expression changes after perturbing *FAM78B* in cell lines or mouse models would be a logical next step to determine if such a pathway exists. This work also highlights the value of analyzing related phenotypes for addictive disorders given the scarcity of risk genes identified through dependence diagnosis GWAS.

## Abbreviations

AAs: African Americans

*AASS*: alpha-aminoadipic semialdehyde synthase

*CACNA1B*: calcium voltage-gated channel subunit alpha1 B

CATS: the Comorbidity and Trauma Study

CD: cocaine dependence

COGA: the Collaborative Study on the Genetics of Alcoholism

EAs: European Americans

*FAM78B*: family with sequence similarity 78-member B

*FTO*: FTO alpha-ketoglutarate-dependent dioxygenase

GWAS: genome-wide association studies

GWS: genome-wide significant

IBD: identity-by-decent

MAF: minor allele frequency

mQTL: methylation QTLs

OD: opioid dependence

*OPRM1*: opioid receptor mu 1 gene

PC: principal components

QC: quality control

SNP: single nucleotide polymorphism

SSADDA: Semi-Structured Assessment for Drug Dependence and Alcoholism

SSAGA: Semi-Structured Assessment for the Genetics of Alcoholism

SUD: substance use disorder

TD: time-to-dependence

## Supplementary materials

The supplementary materials for this article are available at: https://www.explorationpub.com/uploads/Article/file/100132_sup_1.pdf.

## Declarations

### Author contributions

JG, LAF, HRK, HJE, EN, NGM, and LD contributed conception and design of the study; CZ, EJ, and LW, and RS performed the statistical analysis; RS wrote the first draft of the manuscript; all authors contributed to manuscript revision, read and approved the submitted version.

### Conflicts of interest

Dr. Kranzler is a member of an advisory board for Dicerna Pharmaceuticals and a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative (ACTIVE Group), which over the past three years was supported by Alkermes, Amygdala Neurosciences, Arbor Pharmaceuticals, Ethypharm, Indivior, Lundbeck, Mitsubishi, and Otsuka. 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. In the past three years, Dr. Degenhardt has received untied educational grant funding from Indivior and Seqirus. No other authors reported potential conflicts of interest.

### Ethical approval

This study was approved by the Institutional Review Boards of Boston University, Yale University and the University of Pennsylvania.

### Consent to participate

Informed consent to participate in the study was obtained from all participants.

### Consent to publication

Not applicable.

### Availability of data and materials

Summary statistics for GWAS analyses will be provided by the corresponding author upon request.

## References

1.  GBD 2016 Alcohol and Drug Use Collaborators. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Psychiatry. 2018;5:987-1012.

2.  Rummans TA, Burton MC, Dawson NL. How good intentions contributed to bad outcomes: the opioid crisis. Mayo Clin Proc. 2018;93:344-50.

3.  Case A, Deaton A. Mortality and morbidity in the 21st century. Brookings Pap Econ Act. 2017;2017: 397-476.

4.  Degenhardt L, Grebely J, Stone J, Hickman M, Vickerman P, Marshall BDL, et al. Global patterns of opioid use and dependence: harms to populations, interventions, and future action. Lancet. 2019;394:1560-79.

5.  Clark CB, Zyambo CM, Li Y, Cropsey KL. The impact of non-concordant self-report of substance use in clinical trials research. Addict Behav. 2016;58:74-9.

6.  Compton WM, Thomas YF, Stinson FS, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. Arch Gen Psychiatry. 2007;64:566-76.

7.  Degenhardt L, Singleton J, Calabria B, McLaren J, Kerr T, Mehta S, et al. Mortality among cocaine users: a systematic review of cohort studies. Drug Alcohol Depend. 2011;113:88-95.

8.  Roy A. Characteristics of cocaine dependent patients who attempt suicide. Arch Suicide Res. 2009;13: 46-51.

9.  Chan B, Kondo K, Freeman M, Ayers C, Montgomery J, Kansagara D. Pharmacotherapy for cocaine use disorder-a systematic review and meta-analysis. J Gen Intern Med. 2019;34:2858-73.

10. Mistry CJ, Bawor M, Desai D, Marsh DC, Samaan Z. Genetics of opioid dependence: a review of the genetic contribution to opioid dependence. Curr Psychiatry Rev. 2014;10:156-67.

11. Kendler KS, Karkowski LM, Neale MC, Prescott CA. Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. Arch Gen Psychiatry. 2000;57:261-9.

12. Kendler KS, Prescott CA. Cocaine use, abuse and dependence in a population-based sample of female twins. Br J Psychiatry. 1998;173:345-50.

13. Gelernter J, Kranzler HR, Sherva R, Koesterer R, Almasy L, Zhao H, et al. Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. Biol Psychiatry. 2014;76:66-74.

14. Cheng Z, Zhou H, Sherva R, Farrer LA, Kranzler HR, Gelernter J. Genome-wide association study identifies a regulatory variant of RGMA associated with opioid dependence in European Americans. Biol Psychiatry. 2018;84:762-70.

15. Nelson EC, Agrawal A, Heath AC, Bogdan R, Sherva R, Zhang B, et al. Evidence of CNIH3 involvement in opioid dependence. Mol Psychiatry. 2016;21:608-14.

16. Polimanti R, Walters RK, Johnson EC, McClintick JN, Adkins AE, Adkins DE, 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. 2020;25:1673-87.

17. Cabana-Dominguez J, Shivalikanjli A, Fernandez-Castillo N, Cormand B. Genome-wide association meta-analysis of cocaine dependence: shared genetics with comorbid conditions. Prog Neuropsychopharmacol Biol Psychiatry. 2019;94:109667.

18. Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, Kranzler HR, et al. Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. Mol Psychiatry. 2014;19:717-23.

19. Zhou H, Rentsch CT, Cheng Z, Kember RL, Nunez YZ, Sherva RM, et al. Association of OPRM1 functional coding variant with opioid use disorder: a genome-wide association study. JAMA Psychiatry. 2020;77:1072-80.

20. Smith AH, Jensen KP, Li J, Nunez Y, Farrer LA, Hakonarson H, et al. Genome-wide association study of therapeutic opioid dosing identifies a novel locus upstream of OPRM1. Mol Psychiatry. 2017;22:346-52.

21. Anthony J, Warner L, Kessler R. Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: basic findings from the National Comorbidity Survey. Exp Clin Psychopharmacol. 1994;2:244-68.

22. Jones CM. Heroin use and heroin use risk behaviors among nonmedical users of prescription opioid pain relievers-United States, 2002-2004 and 2008-2010. Drug Alcohol Depend. 2013;132:95-100.

23. Reboussin BA, Anthony JC. Is there epidemiological evidence to support the idea that a cocaine dependence syndrome emerges soon after onset of cocaine use? Neuropsychopharmacology. 2006;31:2055-64.

24. Sartor CE, Kranzler HR, Gelernter J. Rate of progression from first use to dependence on cocaine or opioids: a cross-substance examination of associated demographic, psychiatric, and childhood risk factors. Addict Behav. 2014;39:473-9.

25. Associations of nonmedical pain reliever use and initiation of heroin use in the United States [Internet]. Rockville: Substance Abuse and Mental Health Services Administration; c2013 [cited 2020 Dec 1]. Available from: https://www.samhsa.gov/data/sites/default/files/DR006/DR006/nonmedical-pain-reliever-use-2013.htm

26. Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, et al. Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. Am J Med Genet B Neuropsychiatr Genet. 2005;136B:45-52.

27. Sherva R, Wang Q, Kranzler H, Zhao H, Koesterer R, Herman A, et al. Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. JAMA Psychiatry. 2016;73:472-80.

28. Pierucci-Lagha A, Gelernter J, Chan G, Arias A, Cubells JF, Farrer L, et al. Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). Drug Alcohol Depend. 2007;91:85-90.

29. Malison RT, Kalayasiri R, Sanichwankul K, Sughondhabirom A, Mutirangura A, Pittman B, et al. Inter-rater reliability and concurrent validity of DSM-IV opioid dependence in a Hmong isolate using the Thai version of the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). Addict Behav. 2011;36:156-60.

30. Nelson EC, Lynskey MT, Heath AC, Wray N, Agrawal A, Shand FL, et al. Association of OPRD1 polymorphisms with heroin dependence in a large case-control series. Addict Biol. 2014;19:111-21.

31. Nelson EC, Lynskey MT, Heath AC, Wray N, Agrawal A, Shand FL, et al. ANKK1, TTC12, and NCAM1 polymorphisms and heroin dependence: importance of considering drug exposure. JAMA Psychiatry. 2013;70:325-33.

32. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol. 1994;55:149-58.

33. Edenberg HJ. The collaborative study on the genetics of alcoholism: an update. Alcohol Res Health. 2002;26:214-8.

34. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, et al. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet. 1998;81:207-15.

35. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. Addiction. 1999;94:1361-70.

36. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44:955-9.

37. Wetherill L, Lai D, Johnson EC, Anokhin A, Bauer L, Bucholz KK, et al. Genome-wide association study identifies loci associated with liability to alcohol and drug dependence that is associated with variability in reward-related ventral striatum activity in African- and European-Americans. Genes Brain Behav. 2019;18:e12580.

38. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2011;9:179-81.

39. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284-7.

40. Cox DR. Regression models and life-tables. J R Stat Soc Series B Stat Methodol. 1972;34:187-220.

41. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190-1.

42. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.

43. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336-7.

44. Fan N, An L, Zhang M, He H, Zhou Y, Ou Y. GRIN2B gene polymorphism in chronic ketamine users. Am J Addict. 2020;29:105-10.

45. Xie P, Kranzler HR, Krystal JH, Farrer LA, Zhao H, Gelernter J. Deep resequencing of 17 glutamate system genes identifies rare variants in DISC1 and GRIN2B affecting risk of opioid dependence. Addict Biol. 2014;19:955-64.

46. Polimanti R, Zhang H, Smith AH, Zhao H, Farrer LA, Kranzler HR, et al. Genome-wide association study of body mass index in subjects with alcohol dependence. Addict Biol. 2017;22:535-49.

47. Goodyear K, Lee MR, Schwandt ML, Hodgkinson CA, Leggio L. Hepatic, lipid and genetic factors associated with obesity: crosstalk with alcohol dependence? World J Biol Psychiatry. 2017;18:120-8.

48. Sobczyk-Kopciol A, Broda G, Wojnar M, Kurjata P, Jakubczyk A, Klimkiewicz A, et al. Inverse association of the obesity predisposing FTO rs9939609 genotype with alcohol consumption and risk for alcohol dependence. Addiction. 2011;106:739-48.

49. Gramage E, Perez-Garcia C, Vicente-Rodriguez M, Bollen S, Rojo L, Herradon G. Regulation of extinction of cocaine-induced place preference by midkine is related to a differential phosphorylation of peroxiredoxin 6 in dorsal striatum. Behav Brain Res. 2013;253:223-31.

50. Gramage E, Putelli A, Polanco MJ, Gonzalez-Martin C, Ezquerra L, Alguacil LF, et al. The neurotrophic factor pleiotrophin modulates amphetamine-seeking behaviour and amphetamine-induced neurotoxic effects: evidence from pleiotrophin knockout mice. Addict Biol. 2010;15:403-12.

51. Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. Nucleic Acids Research. 2019;48:D983-91.

52. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics. 2014;13:397-406.

53. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. Nature. 2017;545:505-9.

54. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47:D607-13.

55. Bayer KU, De Koninck P, Leonard AS, Hell JW, Schulman H. Interaction with the NMDA receptor locks CaMKII in an active conformation. Nature. 2001;411:801-5.

56. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med. 2015;373:895-907.

57. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316:889-94.

58. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507:371-5.

59. Sevgi M, Rigoux L, Kuhn AB, Mauer J, Schilbach L, Hess ME, et al. An obesity-predisposing variant of the FTO gene regulates D2R-dependent reward learning. J Neurosci. 2015;35:12584-92.

60. Gramage E, Vicente-Rodriguez M, Herradon G. Pleiotrophin modulates morphine withdrawal but has no effects on morphine-conditioned place preference. Neurosci Lett. 2015;604:75-9.