



OPEN

## Multi-trait genome-wide association study of opioid addiction: *OPRM1* and beyond

Nathan Gaddis<sup>1,28</sup>, Ravi Mathur<sup>1,28</sup>, Jesse Marks<sup>1</sup>, Linran Zhou<sup>1</sup>, Bryan Quach<sup>1</sup>, Alex Waldrop<sup>1</sup>, Orna Levrant<sup>2</sup>, Arpana Agrawal<sup>3</sup>, Matthew Randesi<sup>2</sup>, Miriam Adelson<sup>4</sup>, Paul W. Jeffries<sup>3</sup>, Nicholas G. Martin<sup>5</sup>, Louisa Degenhardt<sup>6</sup>, Grant W. Montgomery<sup>7</sup>, Leah Wetherill<sup>8</sup>, Dongbing Lai<sup>8</sup>, Kathleen Bucholz<sup>3</sup>, Tatiana Foroud<sup>8</sup>, Bernice Porjesz<sup>9</sup>, Valgerdur Runarsdottir<sup>10</sup>, Thorarinn Tyrfingsson<sup>10</sup>, Gudmundur Einarsson<sup>11</sup>, Daniel F. Gudbjartsson<sup>11</sup>, Bradley Todd Webb<sup>1</sup>, Richard C. Crist<sup>12</sup>, Henry R. Kranzler<sup>12</sup>, Richard Sherva<sup>27</sup>, Hang Zhou<sup>13</sup>, Gary Hulse<sup>14</sup>, Dieter Wildenauer<sup>14</sup>, Erin Kelty<sup>15</sup>, John Attia<sup>16</sup>, Elizabeth G. Holliday<sup>16,17</sup>, Mark McEvoy<sup>16,17</sup>, Rodney J. Scott<sup>18</sup>, Sibylle G. Schwab<sup>19</sup>, Brion S. Maher<sup>20</sup>, Richard Gruza<sup>21</sup>, Mary Jeanne Kreek<sup>2,29</sup>, Elliot C. Nelson<sup>3</sup>, Thorgeir Thorgeirsson<sup>11</sup>, Kari Stefansson<sup>11,22</sup>, Wade H. Berrettini<sup>12</sup>, Joel Gelernter<sup>23</sup>, Howard J. Edenberg<sup>24</sup>, Laura Bierut<sup>25</sup>, Dana B. Hancock<sup>1</sup> & Eric Otto Johnson<sup>1,26,28</sup>✉

Opioid addiction (OA) is moderately heritable, yet only rs1799971, the A118G variant in *OPRM1*, has been identified as a genome-wide significant association with OA and independently replicated. We applied genomic structural equation modeling to conduct a GWAS of the new Genetics of Opioid Addiction Consortium (GENOA) data together with published studies (Psychiatric Genomics Consortium, Million Veteran Program, and Partners Health), comprising 23,367 cases and effective sample size of 88,114 individuals of European ancestry. Genetic correlations among the various OA phenotypes were uniformly high ( $r_g > 0.9$ ). We observed the strongest evidence to date for *OPRM1*: lead SNP rs9478500 ( $p = 2.56 \times 10^{-9}$ ). Gene-based analyses identified novel genome-wide significant associations with *PPP6C* and *FURIN*. Variants within these loci appear to be pleiotropic for addiction and related traits.

<sup>1</sup>GenOmics, Bioinformatics, and Translational Research Center, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC, USA. <sup>2</sup>The Laboratory of the Biology of Addictive Diseases, The Rockefeller University, New York, NY, USA. <sup>3</sup>Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. <sup>4</sup>Dr. Miriam and Sheldon G. Adelson Clinic for Drug Abuse, Treatment and Research, Las Vegas, NV, USA. <sup>5</sup>Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia. <sup>6</sup>National Drug and Alcohol Research Centre, University of New South Wales, Randwick, NSW, Australia. <sup>7</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia. <sup>8</sup>Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>9</sup>Department of Psychiatry, State University of New York Downstate Medical Center, Brooklyn, NY, USA. <sup>10</sup>SAA-National Center of Addiction Medicine, Vogur Hospital, Reykjavik, Iceland. <sup>11</sup>deCODE Genetics/Amgen, Reykjavik, Iceland. <sup>12</sup>Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. <sup>13</sup>Department of Psychiatry, Yale University School of Medicine, West Haven, CT, USA. <sup>14</sup>School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth, WA, Australia. <sup>15</sup>School of Population and Global Health, Population and Public Health, The University of Western Australia, Perth, WA, Australia. <sup>16</sup>Hunter Medical Research Institute, Newcastle, Australia. <sup>17</sup>School of Medicine and Public Health, The University of Newcastle, Callaghan, NSW, Australia. <sup>18</sup>School of Biomedical Sciences and Pharmacy College of Health, Medicine and Wellbeing, The University of Newcastle, New Lambton Heights, NSW, Australia. <sup>19</sup>Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, NSW, Australia. <sup>20</sup>Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. <sup>21</sup>Department of Family and Community Medicine, Saint Louis University, Saint Louis, MO, USA. <sup>22</sup>Faculty of Medicine, University of Iceland, Reykjavik, Iceland. <sup>23</sup>Department of Psychiatry, Genetics, & Neuroscience, Yale University School of Medicine, West Haven, CT, USA. <sup>24</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>25</sup>Department of Psychiatry, Washington University, St. Louis, MO, USA. <sup>26</sup>Fellow Program, RTI International, Research Triangle Park, NC, USA. <sup>27</sup>Genome Science Institute, Boston University, Boston, MA, USA. <sup>28</sup>These authors contributed equally: Nathan Gaddis, Ravi Mathur and Eric Otto Johnson. <sup>29</sup>Mary Jeanne Kreek is deceased. ✉email: ejohnson@rti.org

In 2020 the US recorded the highest 12-month count of opioid overdose deaths, > 70,000<sup>1</sup>, which represents a 40% increase since 2019, a > 250% increase since 2000<sup>2</sup>, and is 1.7 times the number of deaths caused by automobile crashes in 2020<sup>3</sup>. Approximately 4% of the US population aged 12 and older (10.1 million people) misused opioids in 2019, with 1.6 million people initiating new prescription opioid misuse<sup>4</sup>. The most recent annual estimate of the total economic burden of prescription opioid abuse and dependence in the US (2013) is over \$78 billion<sup>5</sup>, including Medicaid spending of more than \$8 billion on opioid addiction (OA) treatment<sup>6</sup>. By every metric, the opioid epidemic continues to be a tremendous burden, and the need to expand the medication-assisted treatment toolkit for OA through identification of new targets for drug development is clear<sup>7</sup>.

Animal model and human neuroimaging studies have established a strong, albeit partial, understanding of the neurocircuitry of addiction as heuristically characterized in the Koob and Volkow model<sup>8</sup>. The primary neurocircuitry elements involved (basal ganglia, extended amygdala, and prefrontal cortex) and their molecular connections to the cycle of addiction (intoxication, withdrawal, and preoccupation) are broadly understood. However, there is clear variability in the functioning of this neurocircuitry among individuals as evidenced by only 20–30% of people who use heroin becoming addicted<sup>9,10</sup> and only 8–12% of chronic pain patients prescribed opioids developing OA<sup>11</sup>.

Genetics is a major contributor to individual variation in the risk of developing OA, with ~60% of the population variability being attributable to genetic factors<sup>12,13</sup>. This heritability estimate is comparable to other complex phenotypes, such as Alzheimer's<sup>14</sup>, age-related macular degeneration<sup>15</sup>, and height<sup>16</sup>, which have conclusively associated genetic variants. However, few robust genetic variant associations with OA have been identified<sup>17–20</sup>.

Eight genome-wide association studies (GWAS) of OA have been reported<sup>21–28</sup>, in which the number of cases varied from 104 to 10,544 for ancestry specific analyses. Six of these GWAS identified genome-wide significant (GWS) loci<sup>23,25–29</sup>. However, only the largest analysis, which combined results of European ancestry (EA) cohorts from the US Veterans Affairs Million Veteran Program (MVP), the Study of Addiction: Genetics and Environment (SAGE), and Yale-Penn (YP) cohorts (10,544 cases and 72,163 controls), identified a GWS association that replicated in an independent sample (additional YP data: 508 cases and 206 controls). The variant identified is the long-studied rs1799971 (*OPRM1*-A118G), a functional coding variant (encoding Asn40Asp) in the mu opioid receptor gene (*OPRM1*): discovery  $p = 1.51 \times 10^{-8}$ , replication  $p = 0.049$ . The rs1799971-G protective association with OA was also extended at nominal significance to buprenorphine treatment status in the UK Biobank (240 cases and 360,901 controls;  $p = 0.04$ ).

To maximize discovery, we leveraged genomic structural equation modeling (gSEM)<sup>30</sup> to combine new and existing GWAS data with varied, but closely related, phenotypes for OA to enable the largest GWAS of OA to date (23,367 cases, 384,629 controls: effective sample size 88,114). We brought together novel results from the Genetics of Opioid Addiction Consortium (GENOA) with publicly available summary statistics from the MVP-SAGE-YP<sup>27</sup>, the Psychiatric Genetics Consortium-Substance Use Disorder Group (PGC-SUD)<sup>26</sup>, and the Partners Health Group (PH)<sup>28</sup>. We examined SNP-based heritability and genetic correlation among the varied phenotypic definitions of OA across the contributing cohorts, including diagnostic and frequency of use-based cases and different types of controls: opioid exposed, unexposed, and population-based. We conducted a variant level gSEM analysis in the full complement of cohorts and a gene-based association test based on those results. Whereas gSEM accounts for the sample overlap among the GENOA, PGC-SUD, and MVP-SAGE-YP analyses, we were able to combine the samples and increase substantially available sample size compared to standard meta-analysis. Follow-up analyses included: (1) evaluation of genetic correlation with brain-related phenotypes; (2) estimation of predicted genetically driven differential expression in brain tissues; (3) colocalization of genetic association loci with cis-eQTLs; (4) evaluation of loci pleiotropy, and (5) druggability of nominated targets.

This study provides an unequivocal GWS association signal for the intron 1 locus in *OPRM1* and, through haplotype analysis, suggests that rs1799971 (A118G) may not be the driver of the locus's association with OA. We also identified two novel GWS gene-based associations with OA: *PPP6C* and *FURIN*. Both genes have been previously associated with phenotypes correlated with OA (e.g. *PPP6C* with cigarette smoking<sup>31,32</sup>, alcohol consumption<sup>31</sup>, and depressive symptoms<sup>33</sup>; *FURIN* with schizophrenia<sup>34,35</sup>, risk tolerance<sup>36</sup>, and insomnia<sup>36</sup>). This study links these genes to predicted genetically driven differential expression in brain tissues by OA. Colocalization analysis supports a shared single variant between OA association and gene expression for *PPP6C* but the results for *OPRM1* and *FURIN* are not as well defined. Collectively, these results provide extended insight into the association of *OPRM1* with OA and implicate novel genes associated with this phenotype.

## Results

**Different approaches to defining OA are highly genetically correlated.** Our gSEM for OA brings together novel GWAS data from GENOA and summary statistics from all prior GWAS of OA that included more than 1000 cases and 1000 controls of European ancestry (EA)<sup>26–28</sup>. GENOA is a new consortium comprised of investigators who attend the National Institute on Drug Abuse Genetics and Epigenetics Cross-cutting Research Team Meetings and who have GWAS data on OA (Supplementary Table 1). In this study, OA refers to a broad meaning of addiction to opioids defined by multiple approaches to measuring the phenotype. The success of both the GENOA and gSEM analyses to maximize sample size and discovery then depends on similar heritability and high genetic correlations across the different measures of OA.

We focused on EA cohorts for the genetic correlation and gSEM analyses because these approaches, which allow us to maximize sample size by bridging phenotypes and accounting for cohort overlap, require linkage disequilibrium score regression (LDSC) results to model the genetic variance–covariance matrix. LDSC, in turn, depends on an ancestry-specific reference panel, which isn't currently available for African Americans (AAs).

Among the 10 independent EA cohorts contributing to GENOA, OA was defined by Diagnostic and Statistical Manual criteria for opioid abuse or dependence (DSM-based;  $N = 17,061$ ) or by frequency of use (FOU) of

illicit opioids (e.g., injecting heroin 10 or more times in the past 30 days; FOU-based;  $N = 11,976$ ; Supplementary Table 1). SNP-based heritability for both phenotypes was strong (DSM-based:  $h^2 = 0.11$ ,  $SE = 0.03$ ; FOU-based:  $h^2 = 0.18$ ,  $SE = 0.04$ ) and their genetic correlation robust ( $r_g = 1.05$ ,  $SE = 0.16$ ; SNP-based genetic correlations are not bound by 1.0).

Across the full set of GWAS results contributing to the gSEM GWAS (i.e., GENOA, MVP-SAGE-YP, PGC-SUD, and PH) there are additional OA definitions (MVP and PH used Electronic Health Record ICD-9 or ICD-10 codes for opioid use disorder [OUD]) and variation in type of controls used. GENOA cohorts used a combination of controls (opioid exposed, unexposed, and unknown exposure public controls). MVP and PH used opioid exposed controls, and the PGC-SUD results used here were based on unexposed controls. Regardless of the approach to defining OA or the type of controls, the LDSC genetic correlations across cohorts were very high (all pairwise  $r_g > 0.9$ , Supplementary Table 2). These heritabilities and genetic correlations show that the genetics contributing to OA are highly shared regardless of OA case definition or opioid exposure status of controls to which the cases are compared.

**GENOA GWAS identifies one European Ancestry specific OA association.** Conducting ancestry specific and cross-ancestry meta-analyses of the GENOA cohorts (Supplementary Figs. 1–6 and Supplementary Tables 3–5) yielded one GWS association locus on chromosome 4 among EAs (rs28386916-A,  $\beta = 0.17$ ,  $p = 9.04 \times 10^{-9}$ ). The rs28386916 variant was not associated with OA among AAs ( $\beta = -0.025$ ,  $p = 0.51$ ) and consequently was no longer significant in the EA + AA meta-analysis. Rs28386916-A is common (EUR MAF = 0.40; AFR MAF = 0.81) and was well imputed (imputation quality  $> 0.8$  across cohorts). However, it is an intronic variant located between (T)<sub>11</sub> and (A)<sub>11</sub> simple repeats within the long noncoding RNA ENST00000659878, which may undermine confidence in this result (Supplementary Fig. 7).

**Genomic structural equation model GWAS of opioid addiction identifies two other GWS loci in European Ancestry.** A single common factor gSEM (Fig. 1a) fit the GENOA, MVP-SAGE-YP, PGC-SUD, and PH summary statistics for OA well, with high Akaike information criterion and comparative fit index, and low standardized root mean squared root (SRMR) values (Fig. 1a). Testing the association of 2.4 million variants available across all cohorts (Supplementary Tables 6 and 7 for LDSC and gSEM SNP QC and derivation of available variants, respectively) with the latent genetic factor (effective sample size  $N = 88,114$ ) identified two GWS loci (Fig. 1b; Q–Q plot Supplementary Fig. 8): one on chromosome 6 with 32 GWS variants (top variant rs9478500-C,  $\beta = 0.136$ ,  $p = 2.56 \times 10^{-9}$ ; Supplementary Table 8; forest plot Supplementary Fig. 9a) and the other represented by a single variant on chromosome 16 (rs13333582-C,  $\beta = -0.219$ ,  $p = 3.58 \times 10^{-8}$ ; forest plot Supplementary Fig. 10, LocusZoom plot Supplementary Fig. 12). The LDSC intercept of approximately 1 for this model (Fig. 1b), indicates that these results are not due to uncontrolled inflation, as would be expected if the overlap in the cohorts contributing to some of the summary statistics used here was not adequately accounted for<sup>30</sup>.

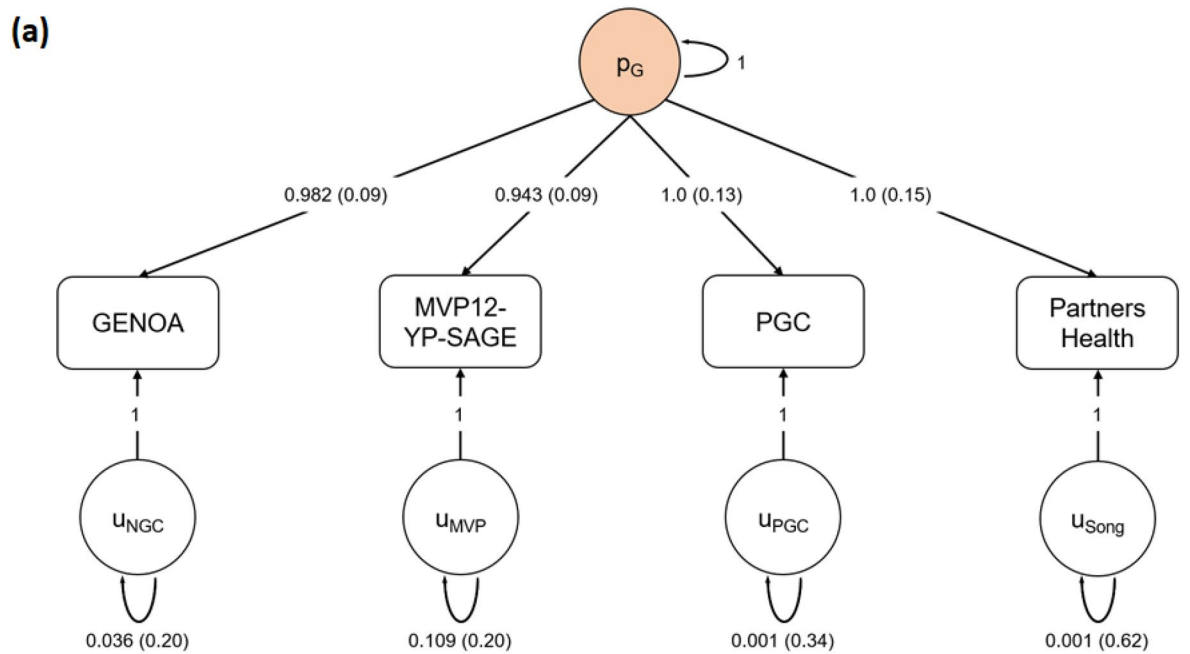
The associated locus on chromosome 6 was centered in intron 1 of the mu-opioid receptor gene *OPRM1* (Supplementary Fig. 11). The minor allele of the lead variant, rs9478500-C, was associated with increased risk of OA ( $\beta = 0.136$ ). All of the GWS variants were in high linkage disequilibrium (LD) with each other ( $r^2 > 0.88$  and  $D' > 0.93$ ; Supplementary Table 9). The previously reported missense variant rs1799971 (*OPRM1*-A118G), which was GWS for OUD in MVP-SAGE-YP<sup>27</sup>, was less statistically significant in our gSEM analysis (rs1799971-G,  $\beta = -0.115$ ,  $p = 1.94 \times 10^{-6}$ ; forest plot Supplementary Fig. 9b). In the MVP GWAS rs9478500-C was associated with OA, but with less statistical significance (MVP rs9478500-C,  $\beta = 0.09$ ,  $p = 4.31 \times 10^{-5}$ ).

Analyses of rs9478500 conditioning on rs1799971 and vice versa using GCTA-COJO<sup>37,38</sup> showed only modest decrements in statistical significance for each variant compared to their unconditional models: rs9478500 ( $p = 2.56 \times 10^{-9}$  for the gSEM analysis vs.  $p = 2.04 \times 10^{-7}$  for the conditional analysis); rs1799971 ( $p = 1.94 \times 10^{-6}$  for the gSEM analysis vs.  $p = 1.41 \times 10^{-4}$  for the conditional analysis). These results suggest that these variants have largely independent associations. However, GCTA-COJO only accounts for LD in terms of  $r^2$ , which is low between these variants among those of European ancestry ( $r^2 = 0.035$ ). In contrast, perfect  $D'$  (1.0) between these variants indicates a strong dependency in the haplotype structure and a need to examine haplotypes.

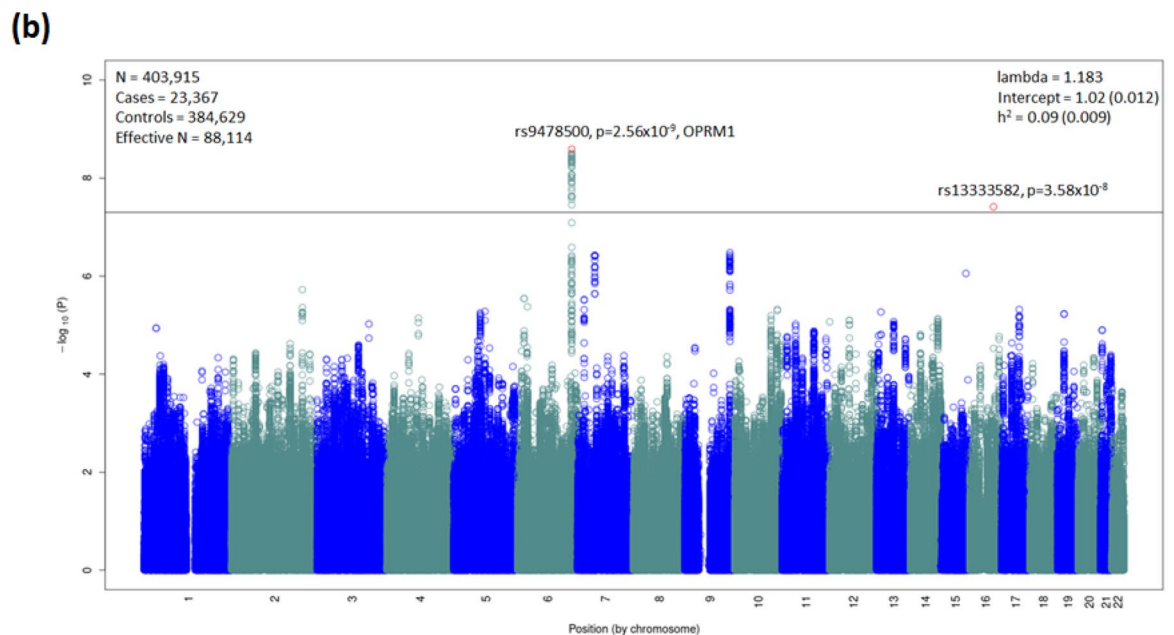
Prior candidate gene studies that examined *OPRM1* haplotypes with rs1799971 suggested that other variants may explain its equivocal association with OA<sup>39,40</sup>. Subject level genotypes for haplotype analysis were available from a subset of cohorts contributing to the gSEM analysis (Fig. 2). In this subset of cohorts, the single variant results for rs9478500 ( $\beta = 0.205$ ,  $p = 2.43 \times 10^{-9}$ ) were strong, but those for rs1799971 were weak ( $\beta = -0.058$ ,  $p = 0.135$ ). Comparison of the three haplotypes formed by rs1799971, rs9478500, and the other GWS *OPRM1* variants (Fig. 2a) further weakened evidence for an association with OA being driven by rs1799971 (Fig. 2b comparison 1:  $p = 0.52$ ,  $\beta = -0.026$ ,  $SE = 0.0397$ ) and strengthened evidence for an association with OA being driven by the effect of the non-rs1799971 variants (comparison 2:  $p = 1.63 \times 10^{-10}$ ,  $\beta = 0.2303$ ,  $SE = 0.036$ , Fig. 2b).

Intergenic variant rs13333582 on chromosome 16 was also GWS (Supplementary Fig. 12), with the minor C allele decreasing risk of OA ( $\beta = -0.22$ ). While rs13333582 passed QC, (MAF = 0.04, INFO  $> 0.8$ , for all cohorts), variants in strong LD showed weaker evidence for association (e.g., rs921982  $r^2 = 0.87$ ,  $p = 2.25 \times 10^{-3}$ ). In addition, rs13333582 was not significant in the GENOA AA analysis ( $p = 0.28$ ), and the direction of effect was opposite of that observed in the gSEM EA analysis.

The GWS association locus on chromosome 4 from the GENOA EA analysis (rs28386916-A,  $\beta = 0.17$ ,  $p = 9.04 \times 10^{-9}$ ) was not included in the gSEM analysis as it was absent from the MVP and PH GWAS. However, several variants in high LD ( $D' > 0.6$ ) with it were not significant in the gSEM analysis (minimum p-value for rs11943738— $p = 0.001$ ,  $\beta = 0.057$ , and  $SE = 0.017$ , Supplementary Table 10).



Model Fit					
Chi-square	Df	P-value	AIC	CFI	SRMR
0.857	2	0.958	16.086	1	0.020



**Figure 1.** Genomic SEM model and Manhattan plot. **(a)** A common factor ( $p_g$ ) gSEM model (using GenomicSEM) is fit with summary statistics from GENOA, MVP12-YP-SAGE, PGC, and Partners Health cohorts. Standardized estimates and standard errors are shown for each free parameter. Model fit is shown by a non-significant chi-square test, high Akaike information criterion (AIC, higher is better) and comparative fit index (CFI) equal to 1.0, and low standardized root mean squared root (SRMR) values (ideally  $< 0.05$ ). **(b)** Manhattan plot for gSEM results with summary statistics from GWAS from each cohort. Bonferroni correction was used to correct for multiple comparisons; associations with  $P < 2 \times 10^{-8}$  (indicated by horizontal black bar) were genome-wide significant (top SNP highlighted in red).

(a)

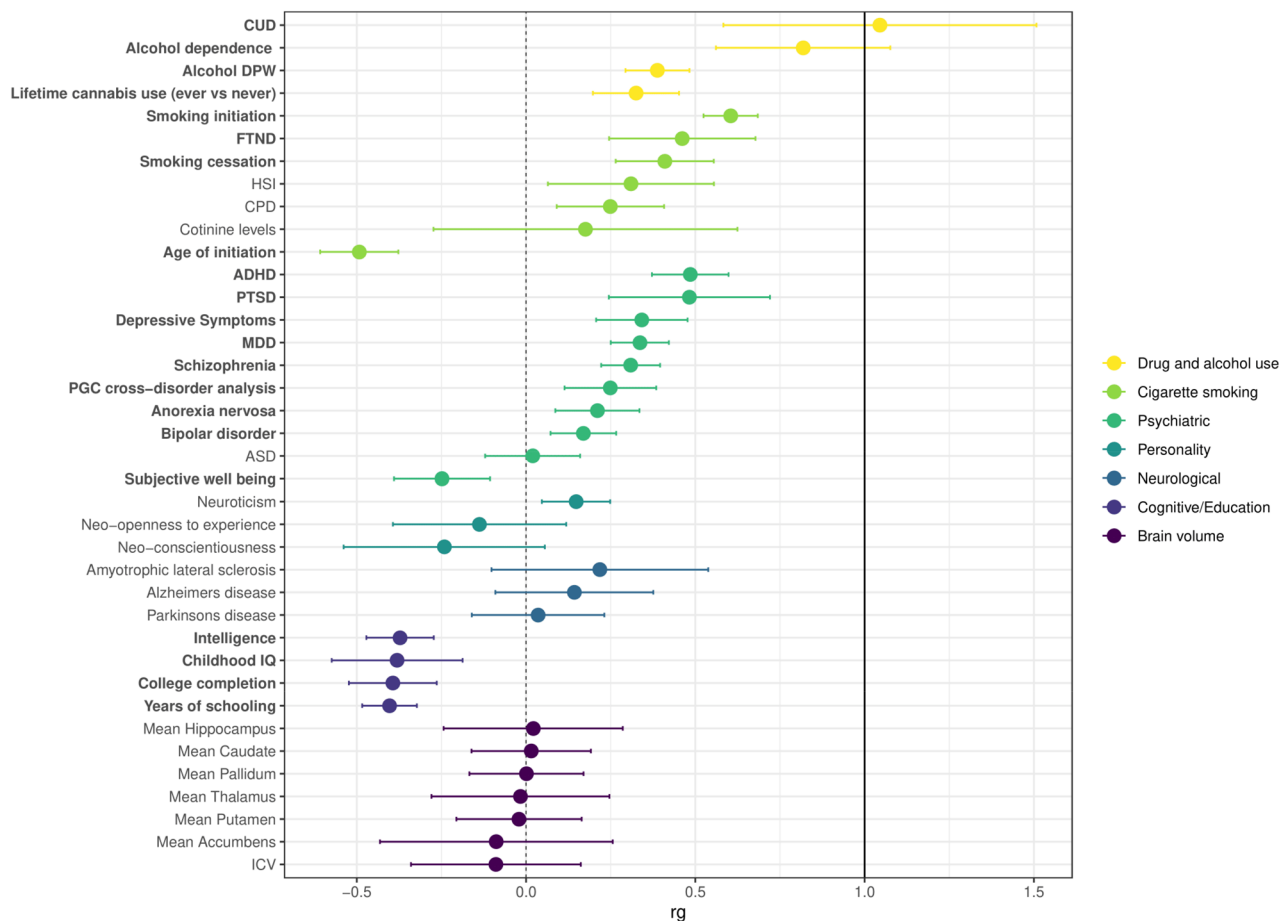
A	B	C	Variant	Chr. 6 Position
A	G	A	rs1799971	154360797
A	A	G	rs9322445	154362254
A	A	C	rs3778145	154367080
T	T	C	rs3778146	154377925
T	T	C	rs9478500	154378223
G	G	A	rs3778147	154378739
G	G	A	rs3823010	154379152
C	C	T	rs9285542	154379934
G	G	T	rs3778148	154381012
C	C	G	rs3778149	154382139
C	C	T	rs7773995	154382367
G	G	A	rs7772959	154382473
A	A	G	rs7759522	154382639
T	T	C	rs9479754	154383058
T	T	C	rs3778150	154383658
T	T	C	rs9478501	154384531
A	A	C	rs12333298	154386364
T	T	C	rs73014269	154387204
G	G	A	rs73014273	154387211
T	T	G	rs6927269	154388306
A	A	C	rs73014276	154389347
A	A	G	rs7453922	154389783
G	G	A	rs1461773	154391369
T	T	C	rs9478503	154392675
C	C	T	rs1381376	154393258
T	T	C	rs3778151	154393680
C	C	A	rs3778153	154393884
A	A	G	rs9478504	154395159
T	T	A	rs17209711	154396455
G	G	A	rs17275521	154396472
G	G	C	rs144155532	154398647
G	G	A	rs9478505	154399992
A	A	G	rs9479755	154402414

■ = major allele   ■ = minor allele

(b)

Haplotype					
	REF	ALT	Beta	SE	P-value
1	A	B	-0.03	0.04	0.52
2	A	C	0.23	0.04	1.63 x 10 <sup>-10</sup>
3	B	C	0.26	0.05	1.68 x 10 <sup>-7</sup>

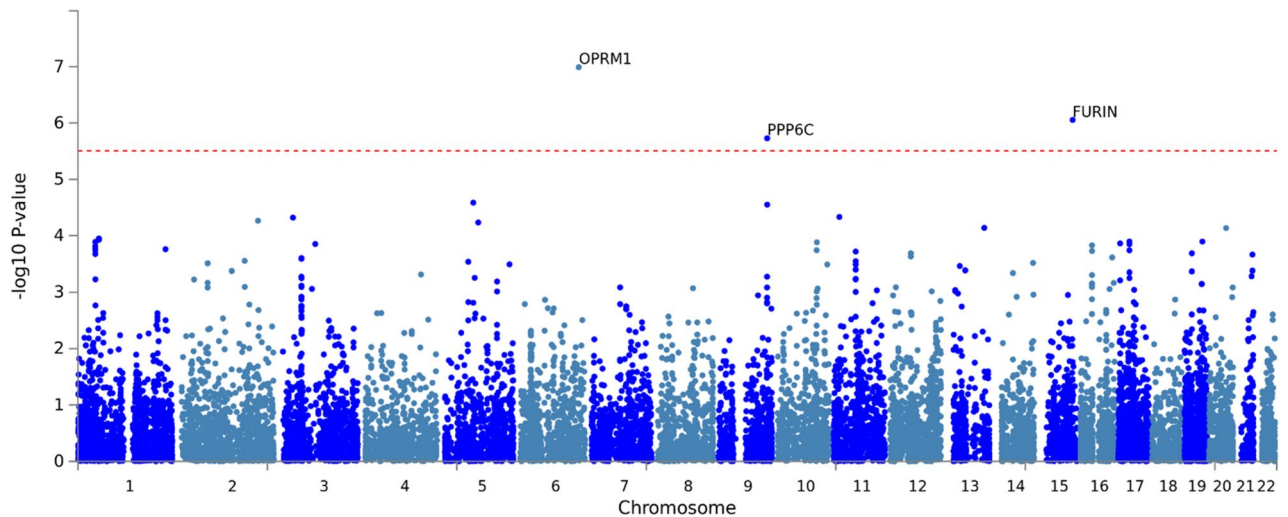
**Figure 2.** Association of major haplotypes for genome-wide significant *OPRM1* variants with OA. (a) The 3 major haplotypes for genome-wide significant *OPRM1* variants. Haplotype A is the predominant haplotype (frequency ~0.69 among contributing cohorts) and consists of major alleles for all variants. Haplotype B (frequency ~0.13 among contributing cohorts) consists of the minor allele for rs1799971 and the major allele for all other variants. Haplotype C (frequency ~0.16 among contributing cohorts) consists of the major allele for rs1799971 and minor allele for all other variants. The cohorts for whom we had the raw data to conduct the haplotype analyses were: UHS, VIDUS, ODB, Yale-Penn, CATS and Kreek (Supplementary Table 1). (b) Association of *OPRM1* haplotypes with OA. Haplotype C is associated with increased risk of OA when compared to Haplotype A or Haplotype B, whereas Haplotype B does not have a significant impact on OA relative to Haplotype A. The single variant results using the cohorts contributing to the haplotype analyses were: rs1799971 beta = -0.058,  $p = 0.135$ ; rs9478500 beta = 0.205,  $p = 2.43 \times 10^{-9}$ .



**Figure 3.** Genetic correlations of opioid addiction (OA) with 38 other brain-related phenotypes. Correlations were calculated using linkage disequilibrium (LD) score regression with the gSEM OA GWAS meta-analysis results, compared with results made available via LD Hub or study investigators (see Supplementary Table 24 for original references). Phenotypes were grouped by disease/trait or measurement category, as indicated by different colorings. Dots indicate the mean values for genetic correlation ( $r_g$ ); error bars show the 95% confidence intervals; the dashed vertical black line corresponds to  $r_g = 0$  (no correlation with OA), and the solid vertical black line corresponds to  $r_g = 1.0$  (complete correlation with OA). Phenotypes with significant correlation with OA are bolded (1 degree of freedom Chi-square test; Bonferroni adjusted  $p$ -value  $< 0.05$  after accounting for 38 independent tests). Exact  $p$ -values are provided in Supplementary Table 10). CUD, Cannabis use disorder; DPW, drinks per week; FTND, Fagerström test for nicotine dependence; HSI, heaviness of smoking index; CPD, cigarettes per day; ADHD, attention deficit hyperactivity disorder; PTSD, post-traumatic stress disorder; MDD, major depressive disorder; ASD, autism spectrum disorder; ICV, intracranial volume.

**OA is genetically correlated with 21 other brain-related traits.** LDSC-based genetic correlation ( $r_G$ ) analyses between OA (gSEM results) and 37 brain related traits (Fig. 3, Supplementary Table 11) yielded 21 significant results (Bonferroni threshold of  $p < 1.35 \times 10^{-3}$ ). The highest  $r_G$ s were with cannabis use disorder ( $r_G = 1.0452$ ,  $p = 9.40 \times 10^{-6}$ ) and alcohol dependence ( $r_G = 0.8183$ ,  $p = 4.82 \times 10^{-10}$ ); smoking traits ( $r_G = -0.4928$  to  $0.6043$ ,  $p = 1.82 \times 10^{-49}$  to  $0.44$ ) and other psychiatric disorders ( $r_G = -0.2482$  to  $0.4849$ ,  $p = 0.79$  to  $1.82 \times 10^{-49}$ ) were more modest. Expected inverse genetic correlations were also evident for age of initiation of cigarette smoking ( $r_G = -0.4928$ ,  $p = 5.97 \times 10^{-17}$ ) and cognitive/educational traits ( $r_G = -0.3722$  to  $-0.4033$ ,  $p = 1.00 \times 10^{-4}$  to  $1.40 \times 10^{-22}$ ). There were no genetic correlations between OA and brain volume traits.

**Gene-based MAGMA GWAS of gSEM summary statistics for OA corroborates *OPRM1* and identifies novel genes.** Gene level analyses of gSEM GWAS results using MAGMA<sup>41</sup> showed significant associations with *OPRM1* and two novel genes out of 15,977 genes tested ( $p < 3.13 \times 10^{-6}$ ; Fig. 4; Q-Q plot Supplementary Fig. 13; full results in Supplementary Table 12). These novel gene-based associations include Protein Phosphatase 6 Catalytic subunit gene (*PPP6C*) and Furin Paired Basic Amino Acid Cleaving Enzyme gene (*FURIN*). The association between *PPP6C* and OA utilized 57 variants including several in LD that approached variant-level genome-wide significance ( $p = 3.37 \times 10^{-7}$  to  $2.14 \times 10^{-5}$ , Fig. 1b). This peak extended across three genes: *PPP6C*, *SCAI*, and *RABEPK* (Supplementary Fig. 14) but only *PPP6C* reached gene-based genome-wide significance (*SCAI*  $p = 0.0016$ ; *RABEPK*  $p = 2.82 \times 10^{-5}$ ).



**Figure 4.** Gene-level Manhattan Plot. GWAS results were summarized at the gene-level using MAGMA. Bonferroni correction was used to correct for multiple comparisons; associations with  $P < 3 \times 10^{-6}$  (indicated by horizontal red dotted line) were genome-wide significant.

The gene-level association between *FURIN* and OA was based on a single variant (rs17514846-A,  $\beta = -0.08$ ,  $p = 8.82 \times 10^{-7}$ ). While other *FURIN* variants were available in individual cohorts, the gSEM GWAS, and thereby the MAGMA analysis, included only one due to the gSEM method's requirement that variants be present in every contributing cohort. A standard logistic regression meta-analysis of *FURIN* variants across the subset of GWAS cohorts without overlapping participants (GENOA, MVP, and PH) retained additional variants excluded from the gSEM analysis and identified 3 additional variants in strong LD with rs17514846 ( $r^2 > 0.64$ ,  $D' = 1.0$ ); all four variants were associated with OA (Supplementary Fig. 15; Supplementary Table 13), the weakest association being for rs17514846 ( $p = 1.67 \times 10^{-6}$ ) and the strongest being for rs11372849, which was GWS (rs11372849-TC,  $\beta = -0.074$ ,  $p = 4.11 \times 10^{-8}$ ; forest plot Supplementary Fig. 16).

#### Predicted genetically driven gene expression in brain tissue expands neurobiologically relevant evidence for OA-associated genes.

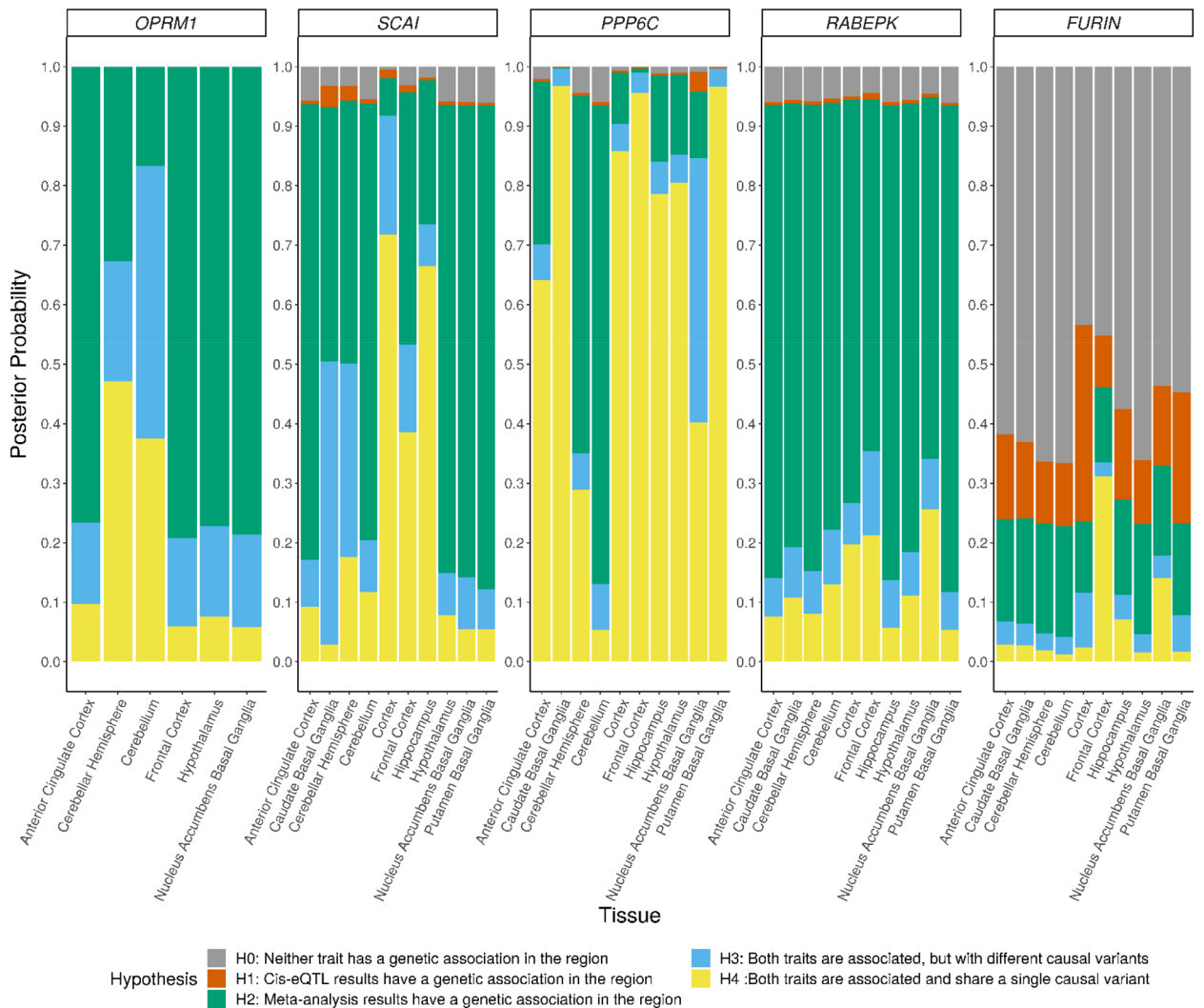
We applied S-PrediXcan<sup>42</sup> using GTEx version 8 eQTL gene models (<http://predictdb.org/>) with the gSEM GWAS summary statistics as input to estimate genetically driven differential gene expression in human brain tissues associated with OA. Fourteen gene-tissue combinations surpassed correction for the total number of gene models and brain tissues (156,215 tests) with a false discovery rate (FDR) less than 0.05 (Table 1; all results presented in Supplementary Table 14). Predicted genetically driven *OPRM1* expression was significantly associated with OA in cerebellum. Only four brain tissues had gene models for *OPRM1* (cerebellum, cerebellar hemisphere, hypothalamus, and nucleus accumbens; Supplementary Table 14). In contrast, *PPP6C* was predicted to be differentially expressed in nine of 12 available brain tissue models. Nearby *SCAI* was the only other gene to show statistically significant genetically driven expression associated with OA, doing so across four brain tissues. *FURIN* was nominally associated ( $p = 9.67 \times 10^{-5}$ ) in hippocampus but did not surpass the  $FDR < 0.05$  threshold. *RABEPK* was not predicted to be differentially expressed by OA (best  $p = 0.055$  in caudate).

#### Some OA associations colocalize with genetically driven gene expression.

To estimate the likelihood that the genetic loci associated with OA share a causal variant with the expression quantitative trait loci (eQTLs) for our nominated genes (*OPRM1*, *PPP6C*, and *FURIN*), we applied coloc<sup>43</sup> to our gSEM GWAS results and the GTEx eQTL results for these genes. Because the variants underlying the GWS association for *PPP6C* physically extend into *SCAI* and *RABEPK* (Supplementary Fig. 14), we included these genes in the analysis. We evaluated colocalization across the superset of 10 brain tissues which showed genetically driving differential expression for at least one gene in the S-PrediXcan analysis (Supplementary Table 14). *OPRM1* is expressed at relatively low levels in the GTEx brain tissues (Supplementary Fig. 17a). Only six of 10 brain tissues showed variant associations with *OPRM1* expression in GTEx and could be included in the coloc analysis. The posterior probabilities for four tissues of the six tissues tested for *OPRM1* favored the hypothesis that only a genetic association with OA at this locus is present (Fig. 5H2, Supplementary Table 15). However, in the cerebellum, where *OPRM1* is most highly expressed and for which S-PrediXcan predicted differential expression by OA, the greatest posterior probabilities favored hypotheses for both the OA-associated locus and *cis*-eQTL traits being associated, but with different causal variants (H3) or a shared single causal variant (H4). Among the three genes at the *PPP6C*-centered locus, *PPP6C* shows the highest levels of gene expression in brain tissues (Supplementary Fig. 17b–d) and the greatest support for colocalization of OA-associated variants with *cis*-eQTLs for *PPP6C* (Fig. 5, Supplementary Table 15). In contrast, the analysis for *RABEPK* uniformly indicated that the OA-associated variants do not colocalize with the *RABEPK cis*-eQTLs. The analyses for *SCAI* showed mixed results, indicating the OA-associated variants have a moderate probability of colocalization with *SCAI cis*-eQTLs

Gene	Tissue	Across tissue FDR
<i>OPRM1</i>	Cerebellum	0.009
<i>SCAI</i>	Cerebellum	0.009
<i>SCAI</i>	Frontal cortex	0.009
<i>SCAI</i>	Hippocampus	0.009
<i>PPP6C</i>	Hippocampus	0.009
<i>PPP6C</i>	Anterior cingulate cortex	0.01
<i>PPP6C</i>	Cerebellar hemisphere	0.01
<i>PPP6C</i>	Putamen basal ganglia	0.01
<i>PPP6C</i>	Caudate basal ganglia	0.01
<i>SCAI</i>	Cortex	0.01
<i>PPP6C</i>	Cortex	0.01
<i>PPP6C</i>	Frontal cortex	0.01
<i>PPP6C</i>	Hypothalamus	0.01
<i>PPP6C</i>	Nucleus accumbens basal ganglia	0.01

**Table 1.** Fourteen gene-brain region combinations exhibiting predicted genetically driven differential gene expression in human brain regions associated with OA (across tissue FDR < 0.05) in analysis of gSEM GWAS summary statistics with S-PrediXcan analysis using GTEx version 8 eQTL gene models.



**Figure 5.** Colocalization of GWAS loci and QTLs for selected genes across 10 brain tissues. Posterior probabilities of supporting hypotheses regarding the association of each trait with SNPs in a region were calculated using coloc. For *OPRM1*, SNP-gene cis-eQTL associations were reported in GTEx Analysis v8 for only 6 of the 10 tissues.



for some tissues (cortex, frontal cortex, and hippocampus) but not for most, and not as high a probability as for *PPP6C*. For *FURIN*, the null hypothesis of neither trait being associated in this region has the highest posterior probability across all tested brain tissues (Fig. 5, Supplementary Table 15), which is consistent with (1) a single variant (rs17514846) driving the GWS gene-based association with OA, (2) no significant evidence for differential gene expression in the S-PrediXcan analyses, (3) limited evidence for this variant as an eQTL in brain tissues (Supplementary Table 16).

**Drug repurposing analyses suggest druggability for *OPRM1*, *PPP6C*, and *FURIN*.** To identify potential for new pharmacological treatments of OA through drug repurposing or compound development, we examined *OPRM1*, *PPP6C*, and *FURIN* across multiple drug repurposing databases (the Drug Gene Interaction Database v.3.0 [DGIDb]<sup>44</sup>, Connectivity Map [CMap]<sup>45</sup>, PHAROS [<https://pharos.nih.gov/>]<sup>46</sup>). *OPRM1* is a known target of more than one-hundred drugs and compounds, including illicit drugs, abused therapeutics (e.g., oxycodone), and OA treatments (e.g., methadone and buprenorphine) (Supplementary Table 17a–c). In contrast, *PPP6C* is not a target of any known drug or compound but has a 94% likelihood that its protein has ligand properties based on its chemistry<sup>47</sup>. *FURIN* is the target of one approved drug, pirfenidone, which is indicated for treatment of idiopathic pulmonary fibrosis. There are more than 80 compounds identified targeting *FURIN*, most developed as inhibitors targeting *FURIN* function in infectious diseases (Supplementary Table 18a–c).

**Testing previously reported GWAS associations supports three variants & two genes: rs1799971, rs62103177, rs640561, *BEND4*, and *PTPRF*.** Among the previously reported associations, the strongest association in the current results was observed for rs1799971 ( $p = 1.94 \times 10^{-6}$ , see Supplementary Table 20 for previously GWAS associated variants and genes)<sup>27</sup>. The only other previously reported variant tested in our gSEM was rs62103177 in the *KCNQ2* gene<sup>23</sup>, which was nominally associated with OA in our EA cohort ( $p = 0.0024$ ). To capture variants reported to be association with OA that were not included in the gSEM results, we extended our lookup to standard logistic regression meta-analysis of the subsets of EA GWAS cohorts that were not included in the originally reported findings. The *CNIH3* variant, rs10799590, reported by Nelson et al.<sup>24</sup> and the PGC-SUD reported<sup>48</sup> variant, rs201123820, were not statistically significant in our standard meta-analysis ( $p = 0.49$  and  $p = 0.63$ , respectively). Of the two PH reported variants<sup>28</sup>, only rs10014685 was present in independent cohorts, but it was not significant in either (deCODE  $p = 0.89$ ; UHS  $p = 0.36$ ). Examining a recently reported GWAS of prescription opioid misuse (POU)<sup>49</sup>, we see a genetic correlation between OA and POU ( $r_g = 0.74$ ,  $p = 2.24 \times 10^{-12}$ ) and extend their association of rs640561 to OA (rs640561-T,  $\beta = -0.061$ ,  $p = 0.009$ ). Finally, we examined our gene-based GWAS results for evidence supporting previously reported genes and found no support for *GRM8*<sup>22</sup> ( $p = 0.655$ ), *CNIH3*<sup>24</sup> ( $p = 0.174$ ), *CCDC42*<sup>50</sup> ( $p = 0.307$ ) or *SPDYE4*<sup>50</sup> ( $p = 0.856$ ). However, we found nominal support for *BEND4* ( $p = 0.0023$ ) association with OA, which was reported in the PGC-SUD GWAS for the opioid use phenotype (exposed vs. unexposed controls)<sup>26</sup> and *PTPRF* for POU ( $p = 0.026$ )<sup>49</sup>.

## Discussion

In this study, we demonstrated a high degree of genetic correlation between differing diagnostic and frequency-based case definitions of OA and across different types of controls (opioid exposed, unexposed, and public controls), which allowed us to conduct the GENOA meta-analysis and apply gSEM successfully to GENOA and existing summary statistics to conduct the largest GWAS among European ancestry cohort participants to date (23,367 cases and total effective sample size of 88,114 individuals) with the OA case definition. The GENOA GWAS of European ancestry identified a single GWS association (rs28386916), but the variant was not available in the MVP or PH cohorts and, consequently, was not tested for replication and was not present in the gSEM GWAS. Given the position of this variant between simple repeats and that variants in high LD with it were not significant in the gSEM GWAS, it seems likely that it represents a false positive. In the gSEM GWAS we found the strongest statistical evidence to date linking variants in intron 1 of the *OPRM1* gene to OA, extending previous candidate gene studies focused on this gene<sup>39,51,52</sup>. Gene-based analyses also identified two novel GWS genes for OA: *PPP6C* and *FURIN*. Examining the predicted differential expression of these genes in brain tissue and their colocalization with OA association signals suggest that the effect of the *PPP6C* locus on risk of OA is likely to be through effects on *PPP6C* expression, while the signal for *OPRM1* is more complex; there is limited evidence that expression differences explain the association of *FURIN* with OA.

The top finding of this gSEM GWAS for OA was centered in intron 1 of the *OPRM1* gene (lead SNP rs9478500-C). Prior candidate gene studies of this region have found nominal associations of variants in intron 1, including some of those that are GWS here (e.g., rs1381376, rs3778151, & rs3778150)<sup>39,51,52</sup>. The mu-opioid receptor gene, *OPRM1*, has long been a target of OA research and drug development. The functional coding variant rs1799971 (A118G; Asn40Asp) has been studied at length with equivocal results<sup>39,53,54</sup>. In the current GWAS era, only the MVP GWAS of OUD found rs1799971 to be GWS ( $p = 1.51 \times 10^{-8}$ )<sup>27</sup>. Adding cohorts to the MVP summary statistics in the current study reduced the variant's association with OA to  $p = 1.94 \times 10^{-6}$ .

Conditional analyses using GCTA-COJO<sup>37,38</sup> to evaluate the independence of rs9478500 and rs1799971 suggested that each association was independent of the other. However, GCTA-COJO only accounts for LD in terms of  $r^2$  which is low between these variants in EA (0.035), whereas  $D'$  is 1.0. Thus, following prior candidate gene studies<sup>39,40</sup>, we examined the associations of specific *OPRM1* haplotypes with OA. The OA association was strongest with the haplotype consisting of the major rs1799971-A allele and minor alleles at all GWS *OPRM1* variants from our study, which was associated with increased risk compared to the other haplotypes. The association with the haplotype consisting of the minor rs1799971-A allele and major alleles at all other GWS *OPRM1* variants was not significant. In the subset of cohorts used in the haplotype analysis, the variant level association for rs1799971

was also not significant, which limits the strength of our haplotype-based conclusions. However, our earlier haplotype analyses of *OPRM1* intron 1 variants and rs1799971 came to similar conclusions, albeit in more limited datasets<sup>39,40</sup>. This relationship between the underlying EA haplotype structure and risk for OA may explain the equivocal findings at the individual rs1799971 variant level, but it may be that other, as-yet-unidentified variants could be the true causal variants driving these haplotype associations. Given the GCTA-COJO conditional analysis, the haplotype analysis, and the potential for as-yet-unidentified variant the exact nature of this locus' association with OA remains to be determined.

The role of genetically driven *OPRM1* expression also appeared complex in this study. In our S-PrediXcan analysis, we observed statistically significant, predicted differential expression of *OPRM1* for OA in cerebellum, the brain tissue with the highest level of *OPRM1* expression in GTEx. Moreover, one of the two *cis*-eQTL variants in the version 8 GTEx model for *OPRM1* expression (rs478498) is in high LD with our top association variant (rs9478500;  $r^2 = 0.56$ ,  $D' = 0.98$ ), suggesting that the intron 1 locus may have its effect on OA through *OPRM1* expression. However, the colocalization analysis was more equivocal. The hypotheses with the greatest posterior probabilities were that both *OPRM1* expression and OA risk are associated with this locus, but with different causal variants (H3, *posterior probability* = 0.46) and with a single causal variant (H4, *posterior probability* = 0.38). Given the generally low level of *OPRM1* expression across bulk brain tissues, larger sample sizes and single nuclei experiments will be needed to further distinguish which of these hypotheses is most likely. Ultimately, model organism or organoid experiments are likely to be necessary to fully test gene expression as a potential mechanism for the association of this locus with OA.

Beyond *OPRM1*, we also observed a GWS association with OA for the intergenic variant rs13333582. The observations that variants in high LD with rs13333582 had much weaker associations and that rs13333582 is located in a repetitive sequence both suggest the association with OA was a false positive.

Increasing statistical power through a gene-based GWAS of the gSEM summary statistics identified two novel GWS genes for OA: *PPP6C* and *FURIN*. *PPP6C* (Protein Phosphatase 6 Catalytic subunit gene) is a component of a signaling pathway that regulates cell cycle progression known to be involved with the immune system and cancer (<https://www.uniprot.org/uniprot/O00743#function>). However, the gene is also strongly expressed in a variety of adult human brain tissues (Supplementary Fig. 17c) and is linked to abnormal locomotor behavior in mice (<http://www.informatics.jax.org/diseasePortal/genoCluster/view/20628><sup>56</sup>). Predicted biological processes for *PPP6C* include G-protein coupled purinergic nucleotide receptor signaling pathway (GO:0035589) (<https://maayanlab.cloud/archs4/gene/PPP6C>), which affects regulation of neurons, microglia and astrocytes<sup>57</sup>. Predicted genetically driven differential expression of *PPP6C* by OA was significant across several brain regions, and colocalization analysis of *PPP6C cis*-eQTLs and the OA-variant association signal at this locus also showed a high probability of being driven by a shared single variant. Because the *PPP6C*-centered association locus extends into the nearby genes *SCAI* and *RABEPK*, and significant predicted genetically driven differential expression of *SCAI* was also observed, we cannot exclude the possibility that these other genes play a role in, or are responsible for, the *PPP6C*-OA association. However, the degree of gene expression/variant association colocalization for *PPP6C* across brain tissues suggest it as the leading candidate for follow-up studies.

The GWS gene-based association of OA with *FURIN* was driven by a single variant, rs17514846. However, this signal was supported by analysis of additional *FURIN* variants in a subset of cohorts where more *FURIN* variants were available, including a GWS association with OA at the variant level for rs11372849. *FURIN* is a member of the convertase family and encodes a type 1 membrane bound protease that is expressed in neuroendocrine and brain tissues, among others (<https://www.ncbi.nlm.nih.gov/gene/5045>). Although *FURIN* shows higher expression across brain tissues than *OPRM1*, the S-PrediXcan analysis did not show significant predicted genetically driven expression differences associated with OA for this gene, and the highest posterior probabilities from the colocalization analysis favored no association of either eQTLs or the OA-variant association signal at this locus. However, this may reflect the single variant association with OA in the gSEM GWAS and an effect on OA through mechanisms other than gene expression.

LDSC analyses demonstrated moderate to strong genetic correlations between OA and a variety of substance use, psychiatric, and cognitive phenotypes in expected directions (e.g., positive correlation with cannabis use disorder, inverse correlation with age of smoking initiation). Moving from the general genomic signal to the specific OA-associated genes, we observed important differences in *OPRM1*, *PPP6C*, and *FURIN* associations with brain- and SUD-related phenotypes. Although *OPRM1* has been broadly studied, only GWAS of OUD<sup>58</sup> and methadone dose<sup>59</sup> have identified GWS associations with *OPRM1* variants (Supplementary Fig. 18; Supplementary Table 19a)<sup>60,61</sup>. The variant associated with methadone dose, rs73568641, was not associated with OA in this gSEM GWAS ( $p = 0.328$ ). In contrast with *OPRM1*, variants in *PPP6C* have been associated with numerous brain- and SUD-related phenotypes, notably opioid medication use, alcohol consumption, numerous smoking phenotypes, and depression among others (Supplementary Fig. 19; Supplementary Table 19b). The specific *PPP6C* variants nominally associated with OA in this gSEM GWAS have been associated with neuroticism, depressive symptoms and a number of smoking phenotypes at genome-wide significance (Supplementary Table 19b). Variants in *FURIN* have been associated with other brain- and SUD-related phenotypes, most predominantly schizophrenia, but also risk taking, number of sexual partners, and insomnia (Supplementary Fig. 20; Supplementary Table 19c). The variant driving the GWS gene-based result here (rs17514846) and the GWS variant in our subset meta-analysis for OA (rs11372849) are associated with each of these phenotypes at genome-wide significance (Supplementary Table 19c).

Similar variability was seen across *OPRM1*, *PPP6C*, and *FURIN* in drug repurposing analyses. While *OPRM1* is the known target of more than 100 drugs and compounds, *FURIN* is the target of one approved drug (pirlfenidone), and *PPP6C* is not a target of any known drug or compound. However, *FURIN* is the target of more than 80 compounds and the *PPP6C* protein has a 94% likelihood of being ligandable. Should these novel gene associations be validated in subsequent studies they appear to be potentially important drug development targets.

Among this study's limitations, the most notable is the focus on cohorts of European ancestry. This focus was required to maximize sample size and statistical power by combining summary statistics across GENOA, PGC-SUD, MVP and PH GWAS through application of the gSEM GWAS method. This approach allowed us to more than double the number of cases used in the EA-focused MVP-SAGE-YP analyses that yielded genome-wide significance for rs1799971 ( $n = 23,367$  vs  $n = 10,544$ ) by leveraging gSEM's ability to model multiple correlated phenotypes and account for sample overlap. These gSEM analyses will be extended to African Americans when the needed ancestry-specific LDSC reference panel becomes available.

An additional potential limitation is the variability in OA case definitions (e.g., diagnostic and frequency of use) and types of controls (e.g., exposed, unexposed, and population-based) used to define OA phenotypes across the cohorts within the GENOA and across the other contributing GWAS. However, the genetic correlations across phenotypes were uniformly high ( $r_g > 0.9$ ) and resulted in a well-fitting single latent factor gSEM model. An important caveat to the high correlations observed across cohorts with exposed, unexposed, and public controls is that the exposure to opioids was often based on prescribed medication (MVP and PH), which differs in risk of OA from exposure to illicit heroin use. Fine-grained comparison of large samples with different types of exposure to opioids will be needed to resolve this question. Because we have incorporated GENOA and previously published GWAS of OA for our discovery analyses, we do not have independent replication cohorts available in which to test the identified associations. However, OA associations with the intron 1 locus have been previously reported<sup>29,39,51,52</sup>, the chromosome 16 (rs13333582) is an eQTL for a gene previously reported as associated with OA<sup>55</sup>, and variants in both *PPP6C* and *FURIN* have previously been associated with substance use and psychiatric disorder traits that are highly associated with OA (Supplementary Figs. 12 and 13). Thus, independent replication remains to be demonstrated, but the available evidence supports the identified variants and genes as associated with OA.

In this study, we leveraged gSEM across new and existing OA GWAS that employed various opioid-related phenotypes to conduct the largest EA-focused GWAS to date. Our results show the strongest statistical evidence to date for an association between variants in intron 1 of *OPRM1* with OA. Conflicting results from variant conditional analysis and haplotype analysis of the GWS variants and previously associated rs1799971 (A118G) indicate that further functional study of intron 1 variants will be needed to determine the causal variant(s) driving this association with OA. Gene-based analyses identified two GWS associations: *PPP6C* and *FURIN*. These genes are novel for OA, however, variants within them have been associated at genome-wide significance with related phenotypes, such as cigarette smoking, alcohol consumption, general risk taking, and schizophrenia. With strong SNP-based heritability for these OA phenotypes, but only these few GWS findings, it is clear that increased sample sizes for GWAS and complementary approaches (e.g., gene regulation in postmortem brain tissues) are needed to identify much of the genetics driving risk of OA as well as to extend these studies to non-European ancestries.

## Methods

**Ethics declarations.** All contributing cohorts obtained necessary patient/participant consent and appropriate institutional IRB approval received. Meta-analyses of these genetic data is approved under RTI International IRB review (IRB ID: CR00000710 for 14107). All methods were performed in accordance with the relevant guidelines and regulations.

**Cohorts and opioid addiction phenotype.** Descriptive statistics for the GENOA studies contributing previously unpublished GWAS of opioid addiction (OA) to this investigation are provided in Supplementary Table 1, and full descriptions of the studies are provided in the Supplementary Methods. In total, this analysis provides new OA GWAS results for 304,831 individuals, including 7281 cases and 297,550 controls, with an effective sample size of  $28,428 (4 / ((1/\# \text{ Cases}) + (1/\# \text{ Controls})))$ <sup>62,63</sup>.

In the GENOA studies, OA was defined based either on the frequency of opioid use (FOU) or Diagnostic and Statistical Manual (DSM) of Mental Disorders criteria. Some studies included only opioid-exposed individuals in their control groups, while others included both exposed and unexposed individuals.

**Genotype quality control and imputation.** Sites in the GENOA consortium conducted standard genotype quality control using filters appropriate for their samples. SNPs were filtered based on call rate and deviation from Hardy-Weinberg equilibrium. Samples were filtered based on call rate, excessive homozygosity, relatedness, and sex discrepancies. Classification of European ancestry individuals was based on comparison to reference populations using STRUCTURE<sup>64</sup>. Specific filters used for each sample are provided in Supplementary Table 22.

For most samples, genotype imputation was performed with the Michigan Imputation Server<sup>65</sup> using the 1000 Genomes Phase 3 v5 reference panel. For COGA, genotypes were phased with SHAPEIT2<sup>66</sup> and imputed with Minimac3<sup>65</sup> using the 1000 Genomes Phase 3 v5 reference panel. For deCODE, genotype imputation was conducted by long-range phasing and haplotype imputation of chip-genotyped individuals with methods described previously<sup>67</sup>.

**Association testing.** Imputed genotypes for GENOA studies were tested for association with opioid addiction case-control status, adjusting for sex, age, genotype principal components and in some cases recruitment site and other study-specific covariates. For all cohorts except deCODE, association testing was performed using the score test meta-analysis model of rvtests<sup>68</sup>. For COGA, which is a family study, an empirically determined kinship matrix was used with rvtests to account for relatedness. For deCODE, data were analyzed using logistic regression treating disease status as the response and imputed genotype counts as covariates. Other available individual characteristics that correlate with disease status were also included in the model as nuisance variables

(sex, age, county of origin), blood sample availability, and an indicator function for the overlap of the lifetime of the individual with the time span of phenotype collection) using previously described methods<sup>69</sup>. To account for inflation due to population stratification and relatedness, test statistics were divided by an inflation factor (1.10) estimated from linkage disequilibrium score regression (LDSC)<sup>70</sup>.

GENOA cohorts were combined via inverse variance-weighted meta-analysis of variants with MAF > 0.01 and  $R^2 > 0.8$  using METAL<sup>71</sup> with genomic control enabled. For characterization of FURIN, an additional inverse variance-weighted meta-analysis of chromosome 15 was conducted including all studies with no overlapping samples (GENOA + MVP without SAGE and Yale-Penn + Partners; Cases N = 16,849, Controls N = 379,493, Total N = 396,342, Effective Total N = 52,508).

**Cohort descriptions for GenomicSEM.** The gSEM analysis includes results on EA from the GENOA meta-analysis, Million Veteran Program (MVP), Psychiatric Genetics Consortium Substance Use Disorders Group (PGC-SUD), and the Partners Health cohorts. The MVP results are based on a meta-analysis including MVP parts 1 and 2 (total of 8529 OUD cases and 71,200 opioid-exposed controls), along with Yale-Penn and Study of Addiction: Genetics and Environment (SAGE) cohorts (total of 10,544 OUD cases and 72,163 opioid-exposed controls)<sup>27</sup>. The PGC-SUD results include 4,503 opioid dependence cases and 4173 unexposed controls<sup>26</sup>. Unexposed controls are used for the PGC-SUD because the exposed controls results have negative heritability estimates. The partners health cohort includes 1,039 OUD cases and 10,743 exposed controls<sup>28</sup>. Note that the GENOA GWAS, Yale-Penn, and SAGE parts of the MVP, and the PGC-SUD results include overlapping samples. However, accounting for this sample overlap is a feature of the gSEM approach applied in this study. A total of 2,434,903 variants were present in all cohorts and tested for association with the OA latent variable in the gSEM analysis using a total sample size of 403,915 (23,367 cases and 384,619 controls; effective sample size of 88,114).

**Genomic structural equation modeling.** The GenomicSEM package<sup>30</sup> within R was used for genomic structural equation modeling (gSEM). The gSEM implemented multivariable LDSC function within this package was used to calculate single-nucleotide polymorphism heritability on the observed and liability scale (prevalence of 10%), genetic covariance matrices, and genetic correlation. The LD scores from 1000 Genomes Project phase 3 European<sup>70</sup> were used as the reference population in this calculation. The sampling genetic covariance matrix is expanded to incorporate SNP effects by including the covariances between SNPs and each cohort. This expanded sampling genetic covariance includes the multivariate LDSC estimated genetic variances and covariances, along with the sampling covariance matrix of the SNP effects on the cohorts, which are estimated using cross-trait LDSC with the sampling correlation weighted by the sample overlap. With the gSEM implemented LDSC, the overlap of samples between GENOA, MVP meta-analysis, and PGC-SUD is not a concern.

A single latent factor gSEM was used with the residual variance of the latent factor set to 1 to normalize the loading estimates. The loadings were calculated using diagonally weighted least squares and residual variances were bound to above 0.01 to avoid negative residual variance estimates.

**Post-GWAS QC.** Standard post-GWAS QC was performed on the association results for individual cohorts, the GENOA meta-analysis results, and the gSEM results. For the association results for individual cohorts, the following QC was performed on the summary statistics: (1) variants with MAF < 0.01, imputation quality < 0.8, missing values, or values outside expected ranges were removed; (2) variant IDs were standardized; (3) QQ plots were generated to confirm lack of excessive bias and genomic control was applied to remove residual bias (see Supplementary Table 23 for genomic control factors); and (4) LDSC was performed to confirm correlation among the different OA measures used by cohorts (Supplementary Table 2). For the meta-analysis and gSEM results, the following QC steps were performed: (1) QQ plots were generated to confirm lack of excessive bias and (2) GWS signals were examined for consistency of effect size and direction of effect across cohorts.

**Conditional analyses.** To assess the interdependence of OPRM1 signals, GCTA-COJO<sup>37,38</sup> was used to conduct conditional analyses (using the --cojo-cond option to specify the variant to condition on) based on the summary statistics from the gSEM analysis. Separate analyses conditioning on rs1799971 and rs9478500 were performed. UHS was used as the genotype reference for these analyses.

**Haplotype analyses.** To conduct haplotype analyses raw genotype data is needed. Chromosome 6 genotypes were available from UHS, VIDUS, ODB, Yale-Penn, CATS and Kreek (Supplementary Table 1). These cohorts' data were phased with Eagle v2.4 via the Michigan Imputation Server<sup>65</sup>. Haplotypes for samples from each study were constructed by extracting OPRM1 variants that were GWS in the gSEM analysis and concatenating ordered by genomic position. Supplementary Table 24 provides counts for the various haplotypes observed across the studies. The 3 most common haplotypes, which accounted for 98% of observed haplotypes, were tested for association with OA in R adjusting for sex and genotype principal components. Only individuals carrying exclusively these haplotypes were included in the analysis. Two models were run, one in which the haplotype containing all major alleles served as the reference haplotype and one in which the haplotype containing the minor rs1799971-G allele served as the reference. This approach provided three effective comparisons: (a) rs1799971-G haplotype versus the major allele haplotype; (b) minor allele+rs1799971-A haplotype versus the major allele haplotype; (c) minor allele+rs1799971-A haplotype versus rs1799971-G haplotype. Individual cohort results were combined in an inverse variance-weighted meta-analysis using METAL (N = 21,037).

**Gene-based analyses.** Gene-based associations with OA were calculated from the gSEM summary statistics with MAGMA v1.08<sup>41</sup> with a 10 kb gene window via the Functional Mapping and Annotation (FUMA) of GWAS web tool v1.3.6a<sup>72</sup>. The gSEM summary statistics were mapped to 15,977 protein coding genes, resulting in a Bonferroni-corrected threshold of  $p = 3.129e-6$  for declaring genome-wide significance.

**Cross-trait genetic correlations with OA.** Summary statistics from the gSEM were used as input into LD score regression (LDSC) with reference to the 1000 Genomes EUR panel to estimate genetic correlations between OA and 38 other complex phenotypes. These phenotypes were categorized into the following groups: drug and alcohol use, cigarette smoking, psychiatric, personality, neurological, cognitive/education, and brain volume. The full list of phenotypes and GWAS datasets, as obtained from LD Hub or shared by the original investigators, are provided in the Supplementary Table 25.

**PrediXcan.** To investigate the transcriptome-wide associations between predicted gene expression and OA, we employed the MetaXcan v0.6.6<sup>42</sup> method. Briefly, MetaXcan uses association summary statistics to predict associations between gene expression and a phenotype of interest association. Gene expression models were predicted from tissue-specific eQTL datasets. To increase the performance of our prediction models, we used the MASHR-M<sup>73</sup> models built on fine-mapped variables from DAP-G<sup>74</sup>. The specific models we used were pre-computed MetaXcan models available through PredictDB (<http://predictdb.org/>) for 12 brain regions (Supplementary Table 14) that were generated using the GTEx<sup>75</sup> version 8 datasets.

Summary level statistics from the Genomic SEM analysis were used as input to MetaXcan. Prior to input, summary statistics were harmonized according to the best practices guide outlined on the MetaXcan wiki. As part of this process, the `gwas_parsing.py` utility (<https://github.com/hakyimlab/summary-gwas-imputation>) was used to lift summary statistics over to the human genome build version 38 and provide harmonized variant identifiers compatible with those used by GTEx v8. To increase the number of overlapping markers between our summary statistics and the fine-mapped pre-built MASHR models, we imputed missing summary associations as suggested by the best practices workflow. Imputation was performed separately for each chromosome using the `gwas_summary_imputation.py` utility (<https://github.com/hakyimlab/summary-gwas-imputation>) and the pre-computed parquet genotype, genotype metadata files, and European LD block files available through the MetaXcan zenodo repository<sup>76</sup>. Imputed summary statistics were finally re-combined using the `gwas_summary_imputation_postprocess.py` utility.

The resulting imputed, harmonized association summary statistics were then used as input to MetaXcan. The number of genes tested for each tissue is found in Supplementary Table 14. FDR correction was applied to account for the number of genes tested across all tissues (156,215 total tests). A gene's predicted expression was considered significantly associated with OA if its FDR-adjusted p-value fell below a threshold of 0.05.

**Colocalization.** Co-localization analysis was performed using the `coloc` package in R<sup>77</sup>. Cis-eQTL data for individuals of European ancestry from the GTEx v8 eQTL Tissue-Specific All SNP Gene Associations dataset (dbGaP Accession phs000424.v8.p2) were input as a quantitative trait into `coloc` (sample sizes for each tissue type indicated in Supplementary Table 15). Summary statistics from the gSEM analysis were input as a quantitative trait, with a sample size of 403,915. Summary statistics from the standard meta-analysis of OA were input as a case-control trait with 16,849 cases and 379,493 controls. All SNP positions were lifted over to build 38. The cis-eQTL data was partitioned into blocks based on the gene in the SNP-gene pair. For each gene block, only SNPs in the gSEM or meta-analysis summary statistics overlapping with the cis-eQTL data were input into the `coloc` function for (approximate) Bayes Factor colocalization analysis.

## Data availability

The GWAS summary statistics generated and/or analyzed during the current study will be made available via dbGaP; the dbGaP accession assigned to the UHS is phs000454.v1.p1. The website is [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000454.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000454.v1.p1).

## Code availability

Docker images and WDL workflows for analyses of primary GENOA data and gSEM meta-analysis are available at the following Github repositories: [https://github.com/RTIInternational/biocloud\\_docker\\_tools](https://github.com/RTIInternational/biocloud_docker_tools); [https://github.com/RTIInternational/biocloud\\_gwas\\_workflows](https://github.com/RTIInternational/biocloud_gwas_workflows); [https://github.com/RTIInternational/biocloud\\_wdl\\_tools](https://github.com/RTIInternational/biocloud_wdl_tools).

Received: 27 July 2022; Accepted: 21 September 2022

Published online: 07 October 2022

## References

- Ahmad, F.B., Rossen, L.M. & Sutton, P. Provisional drug overdose death counts. (National Center for Health Statistics, 2021).
- Rudd, R. A., Aleshire, N., Zibbell, J. E. & Gladden, R. M. Increases in drug and opioid overdose deaths - United States, 2000–2014. *MMWR Morb. Mortal Wkly. Rep.* **64**, 1378–1382 (2016).
- National Safety Council. In *Motor vehicle deaths estimated to have dropped 2% in 2019* (2020).
- Substance Abuse and Mental Health Services Administration. In *Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health* (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55) (Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration., Rockville, MD, 2020).
- Florence, C. S., Zhou, C., Luo, F. & Xu, L. The economic burden of prescription opioid overdose, abuse, and dependence in the United States, 2013. *Med. Care* **54**, 901–906 (2016).

6. Leslie, D. L., Ba, D. M., Agbese, E., Xing, X. & Liu, G. The economic burden of the opioid epidemic on states: The case of Medicaid. *Am. J. Manag. Care* **25**, S243–S249 (2019).
7. National Academies of Sciences, Engineering, and Medicine. In *Medications for opioid use disorder save lives* (eds Leshner, A. & Manchler, M.) (Washington D.C., 2019).
8. Koob, G. F. & Volkow, N. D. Neurocircuitry of addiction. *Neuropsychopharmacology* **35**, 217–238 (2010).
9. Kreek, M. J. *et al.* Opiate addiction and cocaine addiction: Underlying molecular neurobiology and genetics. *J. Clin. Invest.* **122**, 3387–3393 (2012).
10. Santiago-Rivera, O. J., Havens, J. R., Parker, M. A. & Anthony, J. C. Risk of heroin dependence in newly incident heroin users. *JAMA Psychiat.* **75**, 863–864 (2018).
11. 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).
12. 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).
13. Goldman, D., Oroszi, G. & Ducci, F. The genetics of addictions: Uncovering the genes. *Nat. Rev. Genet.* **6**, 521–532 (2005).
14. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
15. Klaver, C. C. *et al.* Genetic risk of age-related maculopathy. Population-based familial aggregation study. *Arch. Ophthalmol.* **116**, 1646–1651 (1998).
16. Zaitlen, N. *et al.* Leveraging population admixture to characterize the heritability of complex traits. *Nat. Genet.* **46**, 1356–1362 (2014).
17. Levran, O., Yuferov, V. & Kreek, M. J. The genetics of the opioid system and specific drug addictions. *Hum. Genet.* **131**, 823–842 (2012).
18. Hancock, D. B., Markunas, C. A., Bierut, L. J. & Johnson, E. O. Human genetics of addiction: New insights and future directions. *Curr. Psychiatry Rep.* **20**, 8 (2018).
19. Crist, R. C., Reiner, B. C. & Berrettini, W. H. A review of opioid addiction genetics. *Curr. Opin. Psychol.* **27**, 31–35 (2019).
20. Gelernter, J. & Polimanti, R. Genetics of substance use disorders in the era of big data. *Nat. Rev. Genet.* **22**, 712–729 (2021).
21. Nielsen, D. A. *et al.* Genotype patterns that contribute to increased risk for or protection from developing heroin addiction. *Mol. Psychiatry* **13**, 417–428 (2008).
22. Nielsen, D. A. *et al.* Genome-wide association study identifies genes that may contribute to risk for developing heroin addiction. *Psychiatr. Genet.* **20**, 207–214 (2010).
23. 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).
24. Nelson, E. C. *et al.* Evidence of CNH3 involvement in opioid dependence. *Mol. Psychiatry* **21**, 608–614 (2016).
25. 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).
26. 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**, 1–15 (2020).
27. Zhou, H. *et al.* Association of OPRM1 functional coding variant with opioid use disorder: A genome-wide association study. *JAMA Psychiat.* **77**, 1020–1080 (2020).
28. Song, W. *et al.* Genome-wide association analysis of opioid use disorder: A novel approach using clinical data. *Drug Alcohol Depend.* **217**, 108276 (2020).
29. Nelson, E. C. *et al.* Association of OPRD1 polymorphisms with heroin dependence in a large case-control series. *Addict Biol.* **19**, 111–121 (2014).
30. 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).
31. 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).
32. Wootton, R. E. *et al.* Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: A Mendelian randomisation study. *Psychol. Med.* **50**, 2435–2443 (2020).
33. Baselmans, B. M. L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* **51**, 445–451 (2019).
34. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–7 (2014).
35. Periyasamy, S. *et al.* Association of schizophrenia risk with disordered niacin metabolism in an Indian genome-wide association study. *JAMA Psychiat.* **76**, 1026–1034 (2019).
36. Karlsson-Linner, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).
37. 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).
38. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**(369–75), S1–3 (2012).
39. Hancock, D. B. *et al.* Cis-expression quantitative trait loci mapping reveals replicable associations with heroin addiction in OPRM1. *Biol. Psychiatry* **78**, 474–484 (2015).
40. Levran, O., Awolesi, O., Linzy, S., Adelson, M. & Kreek, M. J. Haplotype block structure of the genomic region of the mu opioid receptor gene. *J. Hum. Genet.* **56**, 147–155 (2011).
41. 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).
42. 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).
43. Giambartolomei, C. *et al.* A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018).
44. Cotto, K. C. *et al.* DGIdb 3.0: A redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* **46**, 1068–1073 (2018).
45. Subramanian, A. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452. e17 (2017).
46. Sheils, T. K. *et al.* TCRD and Pharos 2021: Mining the human proteome for disease biology. *Nucleic Acids Res.* **49**, D1334–D1346 (2021).
47. Mitsopoulos, C. *et al.* canSAR: Update to the cancer translational research and drug discovery knowledgebase. *Nucleic Acids Res.* **49**, D1074–D1082 (2021).
48. 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).

49. Sanchez-Roige, S. *et al.* Genome-wide association study of problematic opioid prescription use in 132,113 23 and Me research participants of European ancestry. *Mol. Psychiatry* **26**, 6209–6217 (2021).
50. Kalsi, G. *et al.* Genome-wide association of heroin dependence in Han Chinese. *PLoS ONE* **11**, e0167388 (2016).
51. Zhang, H. *et al.* Association between two mu-opioid receptor gene (OPRM1) haplotype blocks and drug or alcohol dependence. *Hum. Mol. Genet.* **15**, 807–819 (2006).
52. Levran, O. *et al.* Genetic susceptibility to heroin addiction: A candidate gene association study. *Genes Brain Behav.* **7**, 720–729 (2008).
53. Schwantes-An, T. H. *et al.* Association of the OPRM1 variant rs1799971 (A118G) with non-specific liability to substance dependence in a collaborative de novo meta-analysis of european-ancestry cohorts. *Behav. Genet.* **46**, 151–169 (2016).
54. Levran, O. & Kreek, M. J. Population-specific genetic background for the OPRM1 variant rs1799971 (118A>G): implications for genomic medicine and functional analysis. *Mol. Psychiatry* **26**, 3169 (2020).
55. Polimanti, R. *et al.* Multi-environment gene interactions linked to the interplay between polysubstance dependence and suicidality. *Transl. Psychiatry* **11**, 34 (2021).
56. Bult, C. J. *et al.* Mouse genome database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).
57. Zarrinmayeh, H.A.-O. & Territo, P. R. Purinergic receptors of the central nervous system: Biology, PET ligands, and their applications. *Mol. Imaging* **19**, 15 (2020).
58. Zhou, H. *et al.* GWAS including 82,707 subjects identifies functional coding variant in OPRM1 gene associated with opioid use disorder. *medRxiv* **10**, 156 (2019).
59. Smith, A. H. *et al.* Genome-wide association study of therapeutic opioid dosing identifies a novel locus upstream of OPRM1. *Mol. Psychiatry* **22**, 346–352 (2017).
60. Ghousaini, M. *et al.* Open targets genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
61. Ochoa, D. *et al.* Open targets platform: Supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
62. Marquez-Luna, C., Loh, P.R., South Asian Type 2 Diabetes, C., Consortium, S.T.D. & Price, A.L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).
63. Holland, D. *et al.* Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genet.* **16**, e1008612 (2020).
64. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
65. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
66. Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F. & Marchini, J. Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
67. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
68. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016).
69. Price, A. L. *et al.* The impact of divergence time on the nature of population structure: An example from Iceland. *PLoS Genet.* **5**, e1000505 (2009).
70. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
71. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
72. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
73. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).
74. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient integrative Multi-SNP association analysis via deterministic approximation of posteriors. *Am. J. Hum. Genet.* **98**, 1114–1129 (2016).
75. Consortium, T. G. The genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
76. Barbeira, A.N. & Im, H.K. GWAS summary statistics imputation support data and integration with PrediXcan MASHR. (ed. Zenodo) (2019).
77. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).

## Acknowledgements

The late Dr. Mary Jeanne Kreek from the Laboratory of the Biology of Addictive Diseases at the Rockefeller University, New York, NY, was a key contributor to the research presented in this manuscript. The authors thank Nancy Saccone, Weimin Duan, Yi-Ling Chou Duan and Congcong Zhu for assisting with protocol development and testing.

## Author contributions

E.O.J., N.G., and R.M. led the meta-analysis and manuscript writing. D.B.H., A.W., J.M., and L.Z. contributed to the manuscript writing. N.G., D.B.H., L.B., J.G., R.S., E.C.N., and E.O.J. developed and tested the protocol. A.A., D.F.G., G.E., H.J.E., H.R.K., J.G., K.S., L.B., N.M., O.L., and T.T. provided substantive revisions to the manuscript. C.S., N.I., W.E., G.K., S.M., and B.M. contributed the ALIVE cohort to the GENOA analysis. R.G. contributed the ADAA cohort to the GENOA analysis. O.L., M.R., M.A., and M.J.K. contributed the Addictions cohorts to the GENOA analysis. P.W.J., E.C.J., N.G.M., L.D., G.W.M., and ECN contributed the CATS-1 cohort to the GENOA analysis. LW, DL, KB, TF, AA, BP, and HE contributed the COGA cohort to the GENOA analysis. L.B. contributed the COGEND cohort to the GENOA analysis. V.R., T.T., G.E., D.F.G., K.S., and T.T. contributed the deCODE cohort to the GENOA analysis. RCC and WDB contributed the START cohort to the GENOA analysis. N.G., D.B.H., L.B., and E.O.J. contributed the UHS, VIDUS, and SAGE cohorts to the GENOA analysis. JG, HRK, and HZ contributed the Yale-Penn cohort to the GENOA analysis. G.H., D.W., E.K., J.A., E.G.H., M.M., R.J.S., and S.G.S. contributed the CATS-2 cohort to the GENOA analysis. All authors contributed to analysis plan development and reviewed and approved the manuscript.

## Funding

Funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript. The National Institute on Drug Abuse: R01DA044014, R01DA038632, R33DA027486 (EO), NG, DBH); K01DA036751 (RCC), U10DA13043 (WHB), R01DA04401 (RCC, WHB). National Institute on Alcoholism and Alcohol Abuse U10AA008401 (LW, DL, KB, TF, AA, BP, HE). The Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (OL, MR, MJK), The Clinical and Translational Science Award UL1RR024143 from the National Center for Advancing Translational Sciences of the NIH (MJK). Pennsylvania State Department of Health Tobacco Settlement non-formulary grant Pharmacogenetics of Opioid Use Disorder (RCC, WHB). deCODE authors (GE, DFG, TT, KS) acknowledge financial support from the European Commission to the painFACT project (H2020-2020–848099). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: GTEx Analysis v8 of the GTEx Portal on 05/18/21 and/or dbGaP accession number phs000424.v8.p2 on 02/17/2021.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-21003-y>.

**Correspondence** and requests for materials should be addressed to E.O.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022