

1 GWAS including 82,707 subjects identifies functional coding variant in *OPRM1* gene associated
2 with opioid use disorder

3

4 Hang Zhou, Ph.D.,^{1,2} Christopher T. Rentsch, Ph.D.,^{2,3,4} Zhongshan Cheng, Ph.D.,^{1,2} Rachel L.
5 Kember, Ph.D.,^{5,6} Yaira Z. Nunez, B.S.,^{1,2} Janet P. Tate, M.P.H., Sc.D.,^{2,3} Cecilia Dao, Ph.D.,^{2,3}
6 Ke Xu, M.D.,^{1,2} Renato Polimanti, Ph.D.,^{1,2} Lindsay A. Farrer, Ph.D.,^{7,8} Amy C. Justice, M.D.,
7 Ph.D.,^{2,3,9} Henry R. Kranzler, M.D.,^{6,10} the VA Million Veteran Program, Joel Gelernter,
8 M.D.,^{1,2,11*}

9

10

11

12 ¹Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA

13 ²Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA

14 ³Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

15 ⁴Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical
16 Medicine, London, UK

17 ⁵Department of Genetics, University of Pennsylvania Perelman School of Medicine,
18 Philadelphia, PA, USA

19 ⁶Crescenz Veterans Affairs Medical Center, Philadelphia, PA, USA

20 ⁷Departments of Medicine (Biomedical Genetics), Neurology, and Ophthalmology, Boston
21 University School of Medicine, Boston, MA, USA

22 ⁸Departments of Biostatistics, and Epidemiology, Boston University School of Public Health,
23 Boston, MA, USA

24 ⁹Yale School of Public Health, New Haven, CT, USA

25 ¹⁰Department of Psychiatry, University of Pennsylvania Perelman School of Medicine,
26 Philadelphia, PA, USA

27 ¹¹Departments of Genetics and Neuroscience, Yale University School of Medicine, New Haven,
28 CT, USA

29

30 *Corresponding Author: Joel Gelernter, Department of Psychiatry, Yale School of Medicine,
31 Veterans Affairs Connecticut Healthcare System, 116A2, 950 Campbell Ave, West Haven, CT
32 06516, USA. Phone: +1 (203) 494-6326 x3590; Fax: +1 (203) 937-4741; Email:
33 joel.gelernter@yale.edu.

34 **Abstract**

35 One way to address the current crisis in opioid use is to improve our understanding of the
36 biological mechanisms of opioid use disorder (OUD). We completed a primary GWAS of
37 electronic health record-defined OUD in European-ancestry participants in the Million Veteran
38 Program (MVP) sample, which included 8,529 affected subjects and 71,200 opioid-exposed
39 controls. In the MVP alone, there were no genome-wide significant (GWS) associations. We
40 then subjected the MVP and additional OUD GWAS results from the Yale-Penn and SAGE
41 samples to meta-analysis (in total, 10,544 OUD cases and 72,163 opioid-exposed controls). A
42 functional coding variant (rs1799971, encoding Asn40Asp) in *OPRM1* (mu opioid receptor gene,
43 the main biological target for opioid drugs) reached GWS ($p=1.51\times 10^{-8}$); then replicated in two
44 independent samples (each at $p<0.05$). The final meta-analyzed p-value for this variant in all
45 samples was 7.81×10^{-10} . SNP-based heritability of OUD was 11.3%. OUD was genetically
46 correlated with 83 traits, including multiple substance use traits, psychiatric illnesses, cognitive
47 performance, and others. Mendelian Randomization revealed possible causal effects on OUD
48 risk from tobacco smoking, major depression, neuroticism, and cognitive performance. Despite
49 the inclusion of data from the MVP, discovery of a significant association depended on including
50 other purpose-collected samples as well. Recruitment of additional opioid dependent subjects
51 for future studies – especially of non-European ancestry – is a crucial next step.

52 Introduction

53 Opioid abuse, addiction, and overdose are at epidemic levels in the United States. Opioids are
54 the leading cause of overdose deaths, and their use has increased dramatically in recent
55 decades [1]. A multifaceted approach is needed to address the opioid crisis, including improving
56 our understanding of the biological mechanisms of opioid addiction. Opioids exert their
57 biological effects primarily by binding (mainly in brain and peripheral nervous tissues) to the
58 opioid receptors *mu* (μ), *kappa* (κ), and *delta* (δ), which are encoded by *OPRM1*, *OPRK1*, and
59 *OPRD1*, respectively [2]. Numerous candidate-gene association studies of these genes
60 (especially *OPRM1*) and those encoding related proteins have been conducted in the past two
61 decades (reviewed in ref [3, 4]), but prior studies have failed consistently to demonstrate
62 association (e.g., studies on *OPRM1**rs1799971 are reviewed in ref [5]). However, a prior
63 genome-wide association study (GWAS) of opioid dosing reported a genome-wide significant
64 (GWS) association mapping upstream of the *OPRM1* locus [6].

65 Rs1799971 (A118G, encoding Asn40Asp) [5, 7], a functional variant, is one of the most
66 studied candidate variants for substance use traits. While no consistent results were observed,
67 this could reflect limited power and population heterogeneity of previous studies (allele
68 frequency varies greatly in different populations [8, 9]). Several kinds of evidence support
69 possible functional effects of this SNP: rs1799971 reportedly alters beta-endorphin binding and
70 activity [7], may be associated with cortisol response to naloxone blockade [10], and may be
71 associated with neurobehavioral functions in a mouse model [11, 12] and human induced
72 pluripotent stem cell lines [13].

73 Several GWAS of DSM-IV opioid dependence [OD] yielded significant findings [14-16];
74 one included internal replication [16] but none reported clear external replication, probably due
75 to the limited sample sizes available (the largest study so far included 2,015 OD cases [16]).
76 The risk variants identified map to *APBB2*, *PARVA*, *KCNC1*, and *KCNC2* [14] in African-

77 American (AA) samples, and *CNIH3* [15] and *RGMA* [16] in European ancestry samples. There
78 have also been GWAS of related traits including therapeutic opioid dose (noted above) [6], and
79 opioid overdose (which identified one variant near *MCOLN1* in AAs) [17]. Of these, only the
80 study on opioid dosing included an external validation. No GWAS yet has been sufficiently
81 powered to estimate the SNP-based heritability (h^2) of OD.

82 We conducted GWAS on ICD (International Classification of Diseases)–9/10-diagnosed
83 opioid use disorder (OUD) and opioid-exposed controls) in 79,729 European Americans (EAs)
84 from the Million Veteran Program (MVP). Then we meta-analyzed for OUD combining MVP,
85 Yale-Penn, and the Study of Addiction: Genetics and Environment (SAGE) samples [18]. The
86 latter two samples were included in our previous publication [16], but were reanalyzed here as a
87 binary diagnostic trait rather than a criterion count for better congruence with available MVP
88 information. Rs1799971 was the only variant that was GWS ($p=1.51\times 10^{-8}$) in this meta-analysis.
89 We then replicated the result in two independent samples. We estimated the h^2 and detected
90 genetic correlations between OUD and a variety of psychiatric traits. Causal effects of
91 substance use, psychiatric diseases, and educational attainment on liability to OUD were also
92 detected.

93

94 **Methods**

95 **MVP datasets.** The MVP is a cross-sectional mega-biobank supported by the U.S. Department
96 of Veterans Affairs (VA). Enrollment in MVP began in 2011 and is ongoing. Phenotypic data
97 were collected using the VA electronic health record (EHR), and blood samples were obtained
98 for genetic studies [19]. Two phases of genotypic data have been released according to their
99 genotyping epochs and were included in this study. MVP phase1 contains 353,948 subjects, of
100 whom 209,020 were defined previously as EAs [20]. MVP phase2 contains 108,416 subjects.

101 We used the same process as in MVP phase1 for quality control and to define EAs (see
102 Supplementary Methods) [20]; there were 67,268 EA subjects.

103 Cases were participants with at least one inpatient or two outpatient ICD-9/10 codes for
104 OUD (Supplementary Table 1) between 2000 and 2018. In MVP phase1, there were 6,367 OUD
105 EAs (3.04% prevalence), and 2,162 OUD EAs (3.21% prevalence) in MVP phase2 were
106 included in this study. Stringent criteria were applied to define incident opioid-exposed controls
107 (see details in ref [21]). In short, we started with all MVP participants and excluded subjects with
108 exposure to a prescription opioid <7 consecutive days, or with VA follow-up less than 6 months
109 after baseline, or with cancer diagnosed before or after baseline, or with baseline opioid
110 dosage >90 mg morphine equivalent daily dose (MEDD), or with OUD diagnosis or OUD
111 treatment at baseline. For the remaining participants, a latent growth mixture model was applied
112 to identify the major classes of opioid dose (measured by MEDD) trajectories that assigned
113 each individual to the trajectory with the highest probability of membership. Four resultant
114 MEDD trajectories were designated as low, moderate, escalating, and rapidly escalating. To
115 minimize the potential rate of false negatives in the control group, subjects assigned to the low-
116 dose trajectory without an incident OUD diagnosis during follow-up were defined as controls,
117 yielding 55,429 and 15,771 EA controls in MVP phase1 and MVP phase2, respectively.

118 Genotyping in MVP was performed using a customized Affymetrix Biobank Array.
119 Imputation and quality control metrics for MVP phase1 were as described previously [20].
120 Similar processes were used for MVP phase2 (see Supplementary Methods). GWAS was then
121 performed on the MVP datasets. We used logistic regression implemented in PLINK v1.90b4.4
122 [22] for the OUD GWAS correcting for age, sex, and the first 10 principal components (PCs).

123

124 Ethics statement: The Central VA Institutional Review Board (IRB) and site-specific IRBs

125 approved the MVP study. All relevant ethical regulations for work with human subjects were
126 followed in the conduct of the study, and written informed consent was obtained from all
127 participants.

128

129 **Yale-Penn and SAGE datasets.** GWAS for DSM-IV OD criterion counts were performed
130 previously, including three phases of Yale-Penn data, and the SAGE cohort (dbGaP study id
131 phs000092.v1.p1) [16]. We re-analyzed these data using OUD diagnosis. See Supplementary
132 Methods.

133

134 **Meta-analyses.** Sample-size-weighted meta-analyses were performed using METAL [23].
135 Given the unbalanced ratios of cases to controls in MVP samples, effective sample sizes were
136 calculated as follow:

137

$$n_{effective} = \frac{4}{\frac{1}{n_{case}} + \frac{1}{n_{control}}}.$$

138 The calculated effective sample sizes in MVP were used in meta-analyses and all downstream
139 analyses. Only variants present at least in MVP phase1, which is the largest sample (~75% of
140 the total), and with heterogeneity test p-value $>5 \times 10^{-8}$ were retained, leaving 5.7 M variants.

141

142 **Replication in independent samples.** In Yale-Penn, 4,817 subjects were recently added and
143 not included in any prior analysis. We genotyped them using the Illumina Multi-Ethnic
144 Genotyping Array (San Diego, CA) which includes ~1.7 M SNPs. Subjects with mismatched
145 genotypic and phenotypic sex were removed, as were subjects with excessive heterozygosity.
146 Duplicate subjects with respect to the Yale-Penn discovery samples were removed. The

147 remaining subjects were classified into population groups as for MVP. Among the 2,041
148 genetically classified EAs, 508 were diagnosed as DSM-IV OD cases, and 206 were opioid
149 exposed controls. GEMMA was used for an association test only for rs1799971 (i.e., no other
150 markers were evaluated) and corrected for age, sex, and the first 10 PCs.

151 In the UK Biobank (UKB), we looked up the association between rs1799971 (only this
152 marker, as for the other replication sample) and buprenorphine treatment (mostly used to treat
153 OUD; treatment/medication code: 20003_1140871732) in the UKB. We examined GWAS
154 summary data released by the Neale lab (information available at [http://www.nealelab.is/uk-](http://www.nealelab.is/uk-biobank)
155 [biobank](http://www.nealelab.is/uk-biobank)) for 240 cases and 360,901 controls differentiated based on buprenorphine treatment.

156

157 **SNP-based h^2 .** LD Score Regression (LDSC) [24] was used to estimate the SNP-based h^2
158 using 1000 Genomes Project Europeans [8] as the LD reference panel. The major
159 histocompatibility complex (MHC) region (chr6: 26–34Mb) was excluded. Effective sample size
160 was used in LDSC.

161

162 **Genetic correlation.** We estimated the genetic correlation (r_g) between OUD and 715 publicly
163 available traits from LD Hub [25] or other resources using LDSC (Supplementary Table 2) [26].
164 Among the tested traits, 232 were published previously, and 483 from the UKB were
165 unpublished, but integrated in LD Hub. Bonferroni correction was applied and correlation was
166 considered significant at a p-value threshold of 6.99×10^{-5} .

167

168 **Mendelian Randomization.** We used Mendelian randomization (MR) to investigate whether
169 exposures (based on 18 published traits that were significantly correlated with OUD [r_g

170 $p < 6.99 \times 10^{-5}$) have causal effects on the liability to OUD. For instrumental variants missing in
171 the OUD summary data, we used the results of the best proxy variant in highest LD ($r^2 > 0.8$) with
172 the missing variant. If the MAF of the missing variant was < 0.01 , or none of the variants within
173 200 kb had LD $r^2 > 0.8$, we removed the instrumental variant from the analysis. Palindromic
174 SNPs (A/T or G/C alleles) with MAF [0.4, 0.5] in the OUD summary data were also removed or
175 replaced with the best proxy variant. For robust causal effect inference, we limited the traits
176 studied to those with > 30 available instruments. Accordingly, 12 exposures were analyzed. We
177 used weighted median [27], inverse-variance weighted (IVW, random-effects model) [28], and
178 MR-Egger [29], implemented in the R package “MendelianRandomization v0.3.0” [30] for MR
179 inference. Evidence of pleiotropic effects was examined by the MR-Egger intercept test, where
180 a non-zero intercept ($p < 0.05$) indicates directional pleiotropy [29]. Whereas MR analyses
181 require the beta (effect size) and standard error, we calculated these using Z-scores (z), allele
182 frequency (p) and sample size (n) from the OUD meta-analyses [31]:

$$\begin{aligned} \text{beta} &= \frac{z}{\sqrt{2p(1-p)(n+z^2)}} \\ \text{SE} &= \frac{1}{\sqrt{2p(1-p)(n+z^2)}} \end{aligned}$$

184

185 **Results**

186 **Association results for opioid use disorder (OUD)**

187 In MVP phase1, 6,367 subjects were diagnosed as OUD cases, and 55,429 subjects were
188 defined as controls. 2,162 cases and 15,771 controls were included from MVP phase2 (Table
189 1). We meta-analyzed the 8,529 OUD cases and 71,200 controls within MVP (totaling 79,729
190 individuals, Table 1), and no variant reached GWS ($p < 5 \times 10^{-8}$, Supplementary Figure 1). The
191 variant with the smallest p-value was rs1799971 in the *OPRM1* gene ($p = 5.90 \times 10^{-8}$,

192 $n_{\text{effective}}=30,443$; the minor G allele is protective with $\beta=-0.142$ and $se=0.026$).

193

194 **Table 1. Demographics – discovery sample.**

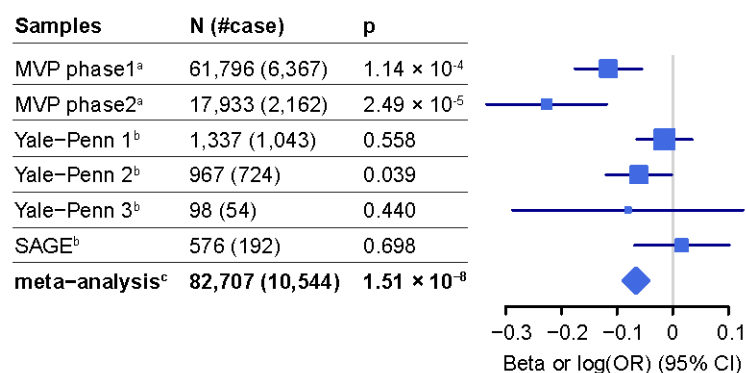
Samples	# Cases	# Controls	Age (mean+SD)	%female
MVP phase1	6,367	55,429	61.2 (13.0)	9.3
MVP phase2	2,162	15,771	61.2 (13.7)	9.7
<i>Sub-total</i>	<i>8,529</i>	<i>71,200</i>	<i>61.2 (13.2)</i>	<i>9.4</i>
Yale-Penn 1	1,043	294	36.9 (10.3)	40.5
Yale-Penn 2	724	243	36.4 (11.4)	33.0
Yale-Penn 3	54	44	33.6 (11.6)	41.8
SAGE	194	382	35.8 (9.1)	33.9
<i>Sub-total</i>	<i>2,015</i>	<i>963</i>	<i>36.4 (10.5)</i>	<i>36.8</i>
Total	10,544	72,163		

195

196

197 We then meta-analyzed the MVP samples with Yale-Penn (three tranches) and SAGE
198 samples, bringing the total sample size to 82,707 (10,544 cases and 72,163 opioid-exposed
199 controls, Table 1). This represents a 24% increase in the number of cases. The OD cases in
200 Yale-Penn and SAGE were diagnosed using DSM-IV, and the controls were opioid exposed
201 (this differs from our previously published GWAS [16], which defined phenotype based on DSM-
202 IV OD criterion count). From the meta-analysis, SNP-based heritability (h^2) was 0.113
203 ($se=0.018$) estimated by LD Score Regression (LDSC). The association of rs1799971 with OUD
204 was GWS ($\beta=-0.066$, $se=0.012$, $p=1.51 \times 10^{-8}$, $n_{\text{effective}}=33,421$, Figure 1, Supplementary
205 Figure 2). The effects were all in the same direction except for SAGE, which might be due to

206 limited sample size. There were no significant results from gene-based association and gene-
 207 set analyses.



208

209 **Figure 1. Associations between rs1799971*G and OUD.** ^aLogistic regression was applied on
 210 unrelated case/control samples in MVP, log(OR) is presented; ^ba linear mixed model was
 211 applied on complex family-based samples, beta is presented; ^ceffective sample size weighted
 212 meta-analysis was applied, and beta is presented.

213

214 Replication in independent samples

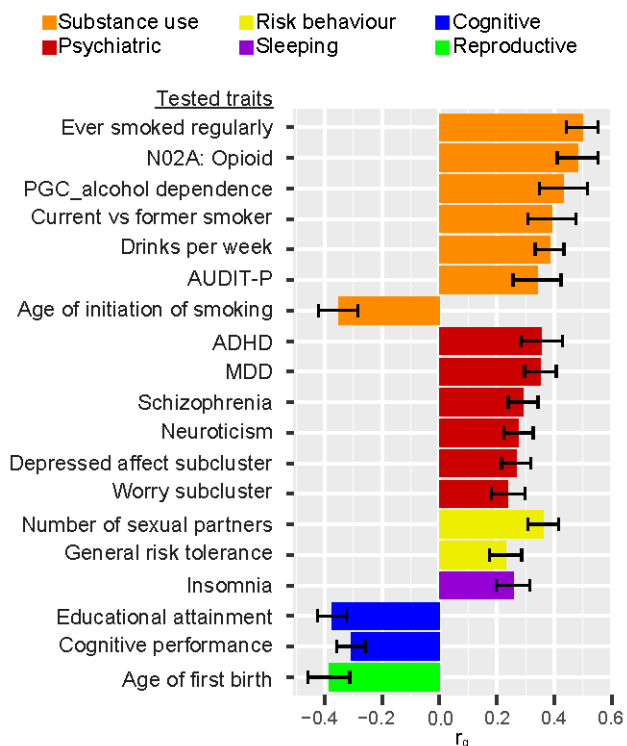
215 In total, 714 EAs (508 OD cases and 206 opioid-exposed controls) were analyzed in the new
 216 Yale-Penn samples, and rs1799971*G was associated with reduced OD risk (i.e., in the same
 217 direction as the discovery meta-analysis) (beta=-0.074, se=-0.038, p=0.049). In the UKB,
 218 rs1799971*G was negatively associated with buprenorphine treatment status (240 cases and
 219 360,901 controls, beta=- 1.90×10^{-4} , se= 9.13×10^{-5} , p=0.038), also consistent with the direction of
 220 effect in the discovery sample. Meta-analysis of discovery and replication cohorts for this variant
 221 yielded a p-value of 7.81×10^{-10} .

222

223 Genetic correlations with other traits

224 We estimated the genetic correlations (r_g) between OUD and 715 traits with publicly available
 225 summary statistics using LDSC. Among those traits, 232 were published, and 483 were

226 unpublished and derived from UKB data. OUD was significantly correlated with 83 of these traits
227 (Supplementary Table 2). Figure 1 depicts 18 correlated traits from published literature
228 (Supplementary Methods). Among the correlated substance use-related traits, “ever smoked
229 regularly” showed the highest correlation with OUD ($r_g=0.51$, $se=0.06$, $p=3.37\times 10^{-19}$), followed
230 by “opioid medication use” in UKB ($r_g=0.48$, $se=0.07$, $p=1.61\times 10^{-11}$). Both “problematic alcohol
231 use” (measured by alcohol dependence and AUDIT-P [Alcohol Use Disorders Identification
232 Test–Problems] score) and “alcohol use quantity” (measured by drinks per week) showed high
233 genetic correlations with OUD. “Unable to stop smoking” (current vs. former smoker), and
234 “earlier age of smoking initiation” were also correlated with OUD. However, correlations with
235 AUDIT-C (Alcohol Use Disorders Identification Test–Consumption), total AUDIT, cigarettes per
236 day, and lifetime cannabis use were not significant after Bonferroni correction. Several
237 psychiatric traits were correlated with OUD, including attention deficit hyperactivity disorder
238 (ADHD, $r_g=0.36$, $se=0.07$, $p=6.78\times 10^{-7}$), major depressive disorder (MDD, $r_g=0.35$, $se=0.06$,
239 $p=1.62\times 10^{-10}$), schizophrenia ($r_g=0.29$, $se=0.05$, $p=1.93\times 10^{-8}$), neuroticism ($r_g=0.27$, $se=0.05$,
240 $p=8.65\times 10^{-8}$), and neuroticism subclusters. OUD was positively correlated with risk-taking
241 behavior and insomnia, and negatively correlated with cognitive traits and age of first birth.
242 These finding are consistent with the known adverse medical, psychiatric, and social
243 consequences of OUD.



244

245 **Figure 2. Genetic correlations between OUD and published traits.** Listed are the 18
 246 published traits significantly correlated with OUD. *N02A: Opioid*: self-reported medication-use of
 247 opioid drugs (Anatomical Therapeutic Chemical [ATC] Classification code: N02A) in UK
 248 Biobank; *PGC*: Psychiatric Genomics Consortium; *AUDIT-P*: the Alcohol Use Disorders
 249 Identification Test–Problems; *ADHD*: attention deficit hyperactivity disorder, *MDD*: major
 250 depressive disorder; *Depressed affect subcluster*: depressed affect neuroticism subcluster;
 251 *Worry subcluster*: worry neuroticism subcluster.

252

253 Mendelian Randomization

254 Using MR, we explored possible causal effects of exposures on OUD (Table 2). Among the 12
 255 tested exposures, five supported a possible causal effect on liability to OUD by at least one
 256 method and without evidence of horizontal pleiotropy (MR-Egger intercept $p > 0.05$): positively
 257 with ever smoked regularly, MDD, neuroticism and worry neuroticism subcluster, and negatively
 258 with educational attainment. There was weak evidence of a causal effect of drinks per week on
 259 OUD risk by the IVW method, but the estimate could be biased due to horizontal pleiotropy.

260

261 **Table 2. Causal effects on OUD by MR.**

Exposure (#instruments)	IVW		Weighted median		MR-Egger		MR-Egger intercept p
	β (se)	p	β (se)	p	β (se)	p	
Ever smoked regularly (197)	0.33 (0.04)	1.88×10⁻¹⁹	0.28 (0.05)	4.79×10⁻⁸	0.18 (0.14)	0.206	0.264
Drinks per week (63)	0.31 (0.10)	1.31×10⁻³	0.23 (0.13)	0.078	-0.04 (0.16)	0.820	8.77×10 ⁻³
MDD (78)	0.18 (0.05)	6.21×10⁻⁴	0.17 (0.06)	4.28×10 ⁻³	-0.25 (0.34)	0.468	0.202
Schizophrenia (108)	0.03 (0.01)	0.020	0.03 (0.02)	0.097	-0.01 (0.07)	0.825	0.439
Neuroticism (131)	0.22 (0.06)	4.07×10⁻⁴	0.16 (0.08)	0.050	0.19 (0.27)	0.468	0.934
Depressed affect subcluster (55)	0.32 (0.10)	2.12×10 ⁻³	0.34 (0.12)	4.01×10 ⁻³	-0.35 (0.50)	0.482	0.173
Worry subcluster (61)	0.28 (0.09)	1.32×10⁻³	0.15 (0.11)	0.157	0.24 (0.40)	0.545	0.919
Number of sexual partners (64)	0.27 (0.09)	1.77×10 ⁻³	0.18 (0.10)	0.083	0.43 (0.40)	0.283	0.689
General risk tolerance (64)	0.16 (0.13)	0.212	0.18 (0.16)	0.257	0.45 (0.58)	0.438	0.608
Insomnia (158)	0.06 (0.02)	3.45×10 ⁻³	0.03 (0.03)	0.226	-0.03 (0.09)	0.731	0.275
Educational attainment (564)	-0.32 (0.04)	2.34×10⁻¹⁶	-0.33 (0.05)	4.09×10⁻¹⁰	-0.18 (0.15)	0.222	0.311
Cognitive performance (133)	-0.12 (0.05)	9.72×10 ⁻³	-0.11 (0.06)	0.061	0.03 (0.22)	0.894	0.477

262 P-values in bold are significant after multiple testing correction (significant threshold, $0.05/36=1.39\times 10^{-3}$). Traits labeled in bold are
 263 those having a causal effect on OUD by at least one method without evidencing horizontal pleiotropy (MR-Egger intercept $p>0.05$).
 264 IVW: inverse-variance weighted (IVW) linear regression. MDD: major depressive disorder. Depressed affect subcluster: depressed
 265 affect subcluster of neuroticism. Worry subcluster: worry subcluster of neuroticism.

266 Discussion

267 Opioid use is at epidemic levels in the United States and is a major cause of death and disability
268 worldwide. Understanding the genetic architecture of OUD might provide clinically useful clues
269 about its biology. However, only a few risk variants have been identified by GWAS so far, and
270 none has had clear external replication. Several factors contribute to this situation: 1). OUD is a
271 complex psychiatric disease with relatively low heritability, and there is no single variant with
272 large effect size that can be detected in small cohorts (contrary to, for example, alcohol
273 dependence [32] with *ADH1B*, and nicotine dependence with the chromosome 15 nicotine
274 receptor cluster [33]); 2). Previous OUD GWAS were relatively small compared to those for
275 legal substance use disorders (e.g., the number of alcohol use disorder cases reached 57,564
276 in a large meta-analysis [34]); 3). In published work relevant to opioid use, there was
277 considerable phenotypic heterogeneity across samples. The ascertainment of OUD cases (e.g.,
278 ICD-diagnosed OUD in the EHR, DSM-IV-assessed OD, patients receiving opioid substitution
279 therapy, and daily injectors of illicit opioids) and controls (e.g., opioid exposed, or random
280 population with unknown opioid exposure status) differ by study. One way to reach a better
281 understanding of OUD genetics is increasing the sample size in a homogeneous cohort.

282 We conducted GWAS of OUD in a large cohort, the MVP, comprising 8,529 cases and
283 71,200 opioid-exposed controls of European ancestry. Most previously reported variants
284 associated with a wide range of opioid-related traits were not significant in MVP (cf. summary
285 statistics). For some, this reflects lack of marker information or LD proxies in the MVP, or the
286 lack of non-European populations in this analysis; some associations were previously reported
287 in African-ancestry populations only [14]; others were reported in EAs, but relevant variants are
288 missing in the MVP data (e.g., rs12442183 near *RGMA* reported by Cheng et al. [16] was
289 filtered by low genotype call rate in imputation). No variant reached GWS in this largest-ever
290 cohort individually; *OPRM1**rs1799971 was nominally significantly associated with OUD

291 ($p=5.90\times 10^{-8}$). We meta-analyzed MVP samples with Yale-Penn and SAGE (re-analyzed here
292 as case/control rather than the quantitative criterion count traits reported originally, to match the
293 available phenotype from the MVP more closely), increasing total sample size to 82,707 (10,544
294 cases and 72,163 opioid-exposed controls). By adding four samples from Yale-Penn and
295 SAGE, rs1799971 reached GWS. The final meta-analyzed p-value for this marker is 1.51×10^{-8}
296 (excluding independent replications).

297 Rs1799971 (A118G) maps to exon 1 of the mu opioid receptor (*OPRM1*) gene, causing
298 an amino acid change from asparagine to aspartic acid. Extensive candidate studies of this
299 variant with a wide range of addictive and other behavioral traits have been conducted over two
300 decades. Associations between rs1799971 and opioid-related traits have not been consistent
301 [5]. We conducted hypothesis-free, genome-wide analyses for OUD and detected association at
302 rs1799971 by almost quintupling the number of cases compared to any previous studies [5, 16].
303 Our increment in exposed controls, which have often been even more limiting than affected
304 OUD subjects in previous studies, is even greater. Since many individuals exposed to opioids
305 become dependent, an unassessed control group is not an ideal alternative to an opioid-
306 exposed control group even if greater numbers of subjects can be achieved – because the
307 former group is more correctly “diagnosis unknown” and inevitably contains many subjects
308 genetically predisposed to OUD who would express that phenotype had they been exposed. We
309 sought replication in two independent samples. One included newly genotyped Yale-Penn
310 subjects, and the other, a proxy-phenotype buprenorphine treatment sample from the UKB. The
311 association was replicated in each of these samples.

312 Although we still encounter a relative lack of power overall as indicated by no additional
313 variant detected beyond rs1799971, a lack of any significant gene associated in gene-based
314 analysis, and no gene-set enrichment, there is reasonable power for genetic correlation.
315 Multiple substance use-related traits including smoking, alcohol, and opioid use and psychiatric

316 traits were among the top correlates. Several smoking traits were positively correlated with
317 OUD, consistent with the strong correlation between nicotine use and opioid use disorder [35,
318 36]. MR analysis provided evidence (considered weak, as it is not supported by all the three
319 tested methods) that the genetic liability to substance use related traits has causal effects on
320 susceptibility to OUD. Medical opioid use is correlated with OUD, as expected. And problematic
321 alcohol use (measured by AD and AUDIT-P score) and drinking quantity are also genetically
322 correlated with OUD. Thus, it may be feasible for prevention or treatment efforts directed at
323 legal substance use to reduce the burden of consequent opioid dependence. Psychiatric traits
324 including ADHD, MDD, schizophrenia, and neuroticism are genetically correlated with OUD,
325 consistent with phenotypic evidence [37, 38]. Weak evidence from MR analyses also indicated
326 possible causal effects on OUD risk of MDD and neuroticism.

327 We note limitations of this study. First, the sample size, though a major improvement, is
328 still not as large as what can be obtained for legal substance use-related traits, and this limited
329 power to detect more GWS signals and to obtain insight into OUD biological mechanisms. Legal
330 substance use traits are more common, and data pertinent to these traits is collected more
331 commonly than for illegal traits in biobanks and EHRs. Second, the phenotypes in the samples
332 we studied were not identical. The MVP used ICD 9/10-diagnosed OUD. We expect false
333 negatives in a sample like the MVP, owing to stigma and insufficient vigilance on the part of
334 treatment teams concentrating mostly on medical illness, but few false positives. Overall, we
335 believe these diagnoses to be quite useful, as false positives would be more of a problem than
336 false negatives. Third, the only GWS variant, rs1799971, has a very small effect size (beta=
337 0.066). The association was close to GWS in MVP alone, and improved by meta-analysis. High
338 quality genotype data and replication in an independent sample were required. Rs1799971 was
339 genotyped directly (was not imputed) in all samples, discovery and replication. Fourth, the
340 replication samples are small (508 OD cases in the new Yale-Penn sample and 240

341 buprenorphine treatment cases in UKB), and the associations only nominally significant
342 ($p \sim 0.05$). We examined only this one marker in the replication samples, so nominal significance
343 should be a sufficient test of replication (and arguably a 1-tailed test could have been applied
344 rather than the more stringent 2-tailed test applied here). The phenotype in UKB is a proxy
345 phenotype—buprenorphine treatment. Although buprenorphine is a first-line drug for OUD
346 treatment, it could have been used for other purposes in the UKB population, including pain
347 management; but if this is true to any considerable extent, it should reduce our power to detect
348 an association due to added noise, rather than lead to a false positive finding. Fifth, we only
349 studied samples with European ancestry. There has been a lack of recruitment for non-
350 European populations globally, e.g., only a few GWASs have been conducted in AAs [6, 14] in
351 smaller cohorts.

352 In summary, we report here the largest GWAS and the largest meta-analysis for OUD so
353 far. This finding may not have direct implications for personalized medicine – because the
354 relevant gene is already the main physiological target of all opioids, illegal and therapeutic;
355 providing, at least, a “proof of principle” of relevance of the finding. OUD is genetically correlated
356 with substance use traits, other psychiatric traits, insomnia, and cognitive performance. Among
357 these, ever smoked regularly, MDD, neuroticism, and cognitive performance have potential
358 causal or protective effects on the liability of OUD, which provides clues for future prevention
359 efforts. Recruitment of additional OUD subjects – especially of non-European ancestry – is a
360 crucial next step. Considering the general lack of private foundation funding for study of
361 substance use disorders, it is likely that government-supported funding agencies will be required
362 to accomplish this goal.

363

364 **Acknowledgements**

365 This research used data from the Million Veteran Program, Office of Research and
366 Development, Veterans Health Administration, and was supported by award #1I01BX003341.
367 This publication does not represent the views of the Department of Veterans Affairs or the
368 United States Government. Supported also by a NARSAD Young Investigator Grant from the
369 Brain & Behavior Research Foundation (HZ). Dr. Gelernter was supported by National Institutes
370 of Health grants RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330,
371 and R01 AA017535, the New England MIRECC, and by the Department of Veterans Affairs
372 Medical Research Program. Dr. Kranzler was supported by National Institutes of Health grants
373 R21 DA10242 and R01 DA18432 and both he and Dr. Kember were supported by the VISN 4
374 MIRECC. Genotyping services for a part of the Yale-Penn GWAS study were provided by the
375 Center for Inherited Disease Research (CIDR) and Yale University (Center for Genome
376 Analysis). CIDR is fully funded through a federal contract from the National Institutes of Health
377 to The Johns Hopkins University (contract number N01-HG-65403).

378

379 The publicly available dataset of Study of Addiction: Genetics and Environment (SAGE) used for
380 the analysis was obtained from dbGaP at [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1)
381 [bin/study.cgi?study_id=phs000092.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1). Funding support for SAGE was provided through the
382 NIH (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of
383 the Gene Environment Association Studies (GENEVA) under Genes, Environment and Health
384 Initiative (GEI). Assistance with phenotype harmonization and genotype cleaning, as well as
385 with general study coordination, was provided by the GENEVA Coordinating Center (U01
386 HG004446).

387

388 We also thank Ann Marie Lacobelle, and Christa Robinson who provided technical assistance
389 for the genotyping.

390

391 Disclosures: Dr. Kranzler is a member of the American Society of Clinical
392 Psychopharmacology's Alcohol Clinical Trials Initiative, which in the past three years was
393 supported by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and
394 Amygdala Neurosciences. Drs. Kranzler and Gelernter are named as inventors on PCT patent
395 application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24,
396 2018.

397

398 References

- 399 1. Jalal, H., et al., *Changing dynamics of the drug overdose epidemic in the United States from 1979*
400 *through 2016*. Science, 2018. **361**(6408).
- 401 2. Darq, E. and B.L. Kieffer, *Opioid receptors: drivers to addiction?* Nat Rev Neurosci, 2018. **19**(8):
402 p. 499-514.
- 403 3. Reed, B., et al., *Genetics of opiate addiction*. Curr Psychiatry Rep, 2014. **16**(11): p. 504.
- 404 4. Mistry, C.J., et al., *Genetics of Opioid Dependence: A Review of the Genetic Contribution to*
405 *Opioid Dependence*. Curr Psychiatry Rev, 2014. **10**(2): p. 156-167.
- 406 5. Schwantes-An, T.H., et al., *Association of the OPRM1 Variant rs1799971 (A118G) with Non-*
407 *Specific Liability to Substance Dependence in a Collaborative de novo Meta-Analysis of*
408 *European-Ancestry Cohorts*. Behav Genet, 2016. **46**(2): p. 151-69.
- 409 6. Smith, A.H., et al., *Genome-wide association study of therapeutic opioid dosing identifies a novel*
410 *locus upstream of OPRM1*. Mol Psychiatry, 2017. **22**(3): p. 346-352.
- 411 7. Bond, C., et al., *Single-nucleotide polymorphism in the human mu opioid receptor gene alters*
412 *beta-endorphin binding and activity: possible implications for opiate addiction*. Proc Natl Acad
413 Sci U S A, 1998. **95**(16): p. 9608-13.
- 414 8. 1000 Genomes Project Consortium, et al., *A global reference for human genetic variation*.
415 Nature, 2015. **526**(7571): p. 68-74.
- 416 9. Gelernter, J., H. Kranzler, and J. Cubells, *Genetics of two mu opioid receptor gene (OPRM1) exon*
417 *I polymorphisms: population studies, and allele frequencies in alcohol- and drug-dependent*
418 *subjects*. Mol Psychiatry, 1999. **4**(5): p. 476-83.
- 419 10. Chong, R.Y., et al., *The mu-opioid receptor polymorphism A118G predicts cortisol responses to*
420 *naloxone and stress*. Neuropsychopharmacology, 2006. **31**(1): p. 204-11.
- 421 11. Zhang, Y., et al., *Mouse model of the OPRM1 (A118G) polymorphism: differential heroin self-*
422 *administration behavior compared with wild-type mice*. Neuropsychopharmacology, 2015. **40**(5):
423 p. 1091-100.
- 424 12. Robinson, J.E., et al., *Receptor Reserve Moderates Mesolimbic Responses to Opioids in a*
425 *Humanized Mouse Model of the OPRM1 A118G Polymorphism*. Neuropsychopharmacology,
426 2015. **40**(11): p. 2614-22.
- 427 13. Halikere, A., et al., *Addiction associated N40D mu-opioid receptor variant modulates synaptic*
428 *function in human neurons*. BioRxiv, 2018. <http://dx.doi.org/10.1101/328898>.
- 429 14. Gelernter, J., et al., *Genome-wide association study of opioid dependence: multiple associations*
430 *mapped to calcium and potassium pathways*. Biol Psychiatry, 2014. **76**(1): p. 66-74.
- 431 15. Nelson, E.C., et al., *Evidence of CNH3 involvement in opioid dependence*. Mol Psychiatry, 2016.
432 **21**(5): p. 608-14.
- 433 16. Cheng, Z., et al., *Genome-wide Association Study Identifies a Regulatory Variant of RGMA*
434 *Associated With Opioid Dependence in European Americans*. Biol Psychiatry, 2018. **84**(10): p.
435 762-770.
- 436 17. Cheng, Z., et al., *Genome-wide scan identifies opioid overdose risk locus close to MCOLN1*. Addict
437 Biol, 2019: p. e12811.
- 438 18. Edenberg, H.J., *The collaborative study on the genetics of alcoholism: an update*. Alcohol Res
439 Health, 2002. **26**(3): p. 214-8.
- 440 19. Gaziano, J.M., et al., *Million Veteran Program: A mega-biobank to study genetic influences on*
441 *health and disease*. J Clin Epidemiol, 2016. **70**: p. 214-23.
- 442 20. Kranzler, H.R., et al., *Genome-wide association study of alcohol consumption and use disorder in*
443 *274,424 individuals from multiple populations*. Nat Commun, 2019. **10**(1): p. 1499.

- 444 21. Rentsch, C.T., et al., *Patterns and Correlates of Prescription Opioid Receipt Among US Veterans: A National, 18-Year Observational Cohort Study*. *AIDS Behav*, 2019.
- 445
- 446 22. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets*. *Gigascience*, 2015. **4**: p. 7.
- 447
- 448 23. Willer, C.J., Y. Li, and G.R. Abecasis, *METAL: fast and efficient meta-analysis of genomewide association scans*. *Bioinformatics*, 2010. **26**(17): p. 2190-1.
- 449
- 450 24. Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies*. *Nat Genet*, 2015. **47**(3): p. 291-5.
- 451
- 452 25. Zheng, J., et al., *LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis*. *Bioinformatics*, 2017. **33**(2): p. 272-279.
- 453
- 454
- 455 26. Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits*. *Nat Genet*, 2015. **47**(11): p. 1236-41.
- 456
- 457 27. Bowden, J., et al., *Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator*. *Genet Epidemiol*, 2016. **40**(4): p. 304-14.
- 458
- 459 28. Bowden, J., et al., *A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization*. *Stat Med*, 2017. **36**(11): p. 1783-1802.
- 460
- 461 29. Bowden, J., G. Davey Smith, and S. Burgess, *Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression*. *Int J Epidemiol*, 2015. **44**(2): p. 512-25.
- 462
- 463
- 464 30. Yavorska, O.O. and S. Burgess, *MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data*. *Int J Epidemiol*, 2017. **46**(6): p. 1734-1739.
- 465
- 466
- 467 31. Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets*. *Nat Genet*, 2016. **48**(5): p. 481-7.
- 468
- 469 32. Gelernter, J., et al., *Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci*. *Mol Psychiatry*, 2014. **19**(1): p. 41-9.
- 470
- 471 33. Tobacco Genetics Consortium, *Genome-wide meta-analyses identify multiple loci associated with smoking behavior*. *Nat Genet*, 2010. **42**(5): p. 441-7.
- 472
- 473 34. Zhou, H., et al., *Meta-analysis of problematic alcohol use in 435,563 individuals identifies 29 risk variants and yields biological insights, pervasive pleiotropy and evidence of causality*. *bioRxiv*, 2019.
- 474
- 475
- 476 35. Clemmey, P., et al., *Smoking habits and attitudes in a methadone maintenance treatment population*. *Drug Alcohol Depend*, 1997. **44**(2-3): p. 123-32.
- 477
- 478 36. Richter, K.P., et al., *Tobacco use and quit attempts among methadone maintenance clients*. *Am J Public Health*, 2001. **91**(2): p. 296-9.
- 479
- 480 37. Sullivan, M.D., et al., *Association between mental health disorders, problem drug use, and regular prescription opioid use*. *Arch Intern Med*, 2006. **166**(19): p. 2087-93.
- 481
- 482 38. Seal, K.H., et al., *Association of mental health disorders with prescription opioids and high-risk opioid use in US veterans of Iraq and Afghanistan*. *JAMA*, 2012. **307**(9): p. 940-7.
- 483

484