Cis-Expression Quantitative Trait Loci Mapping Reveals Replicable Associations with Heroin Addiction in *OPRM1*

Supplemental Information

Supplemental Methods

Cis-expression quantitative trait loci (cis-eQTL) mapping using human prefrontal cortex in the BrainCloud cohort

Cis-eQTL mapping was conducted in the BrainCloud cohort, whose data have been made available to the scientific community. We obtained their single nucleotide polymorphism (SNP) genotype data from the Illumina Human1M-Duo (version 3) and HumanHap650Y (version 3) BeadChip arrays via the database of Genotypes and Phenotypes (dbGaP, accession number phs000417.v2.p1) and their gene expression data from the Illumina Human 49K Oligo array via the BrainCloud software (downloaded at http://braincloud.jhmi.edu/). The opioid receptor, mu 1 (*OPRM1*) gene was represented by the probe HEEBO-045-HCC45I7 (oligo ID: hHC017095, Figure S1), which is a constitutive exonic probe type that recognized all known transcripts of the gene and contained no SNPs in the probe sequence. Gene expression data are also available through the National Center for Biotechnology Information Gene Expression Omnibus (series number GSE30272). Quality control procedures taken with the BrainCloud data have been previously described (1, 2).

Our final BrainCloud analysis dataset included 110 European Americans and 114 African Americans, ranging in age from 0 to 78 years old, who had no neuropathological or neuropsychiatric diagnoses, no abuse of drugs or alcohol, and no positive toxicology result (1, 2). We did not analyze the fetal samples, given the widespread differences that can occur in gene

expression patterns between fetal development and postnatal life (1), or the few samples of Hispanic or Asian ancestry.

Discovery cohort for heroin abuse association testing: Urban Health Study (UHS) heroin addiction cases vs. population controls

Between 1986 and 2005, the UHS recruited over 12,000 subjects from the San Francisco Bay Area, by targeting communities with a high prevalence of injection drug use (3). UHS was designed to follow the incidence and prevalence of HIV-1 viral infection and other factors that affect the health of injection drug users. As previously described (4, 5), eligibility criteria for study entry included injection of an illicit drug in the past 30 days, ability to provide informed consent, and age 18 or older. UHS was a serial, cross-sectional, sero-epidemiological study designed to follow the incidence and prevalence of HIV-1 viral infection, among other factors that affect the health of injection drug users. Stored serum samples from 3,227 UHS subjects were selected for genotyping on the Illumina Omni1-Quad BeadChip, based on matching HIV-1 positive cases with highly exposed HIV-1 negative controls (6). Genotyping followed a restoration process to maximum the quality of genomic DNA using the Illumina Formalin-Fixed Paraffin-Embedded kit. The genotype concordance rate among HapMap control samples was 99.7%. Blind duplicate samples had a genotype concordance of 99.9%. The UHS genotype data are described elsewhere (6) and deposited in dbGaP (accession number phs000454.v1.p1).

UHS participants in the current study met the Office of National Drug Control Policy definition of heroin abuse (injecting 10+ times in the past 30 days) (7, 8). We used the National Survey on Drug Use and Health (NSDUH; 2008-2012) to evaluate the proportion of individuals who reported using heroin 10+ days in the past month and met Diagnostic and Statistical manual

of Mental Disorders, fourth edition (DSM-IV) (9) criteria for heroin dependence in the past year. NSDUH is an annual household-based survey of approximately 65,000 youth and adults aged 12 or older. NSDUH collects data on past month heroin and other substance use frequency and past year DSM-IV substance use disorder. To approximate the characteristics of the UHS cohort, we restricted the NSDUH analyses to adults who had used drugs in the past month (n = 245). Because NSDUH only has past year substance use disorder (unlike UHS which used past month frequency), we further restricted the analysis to individuals who used heroin at least once in the past month (n = 161). Analyses were conducted using SUDAAN version 11.0.1 to account for the complex sampling design. We found that, among individuals who reported using heroin 10 or more times in the past month, 87% met DSM-IV criteria for heroin abuse/dependence (Table S1). We expect that this positive predictive value in NSDUH would be even higher if the frequency of heroin use and DSM-IV criteria were both based on past month rather than past year assessment. UHS participants included in the current study were high-frequency abusers, averaging 80.9 times in the past month, and based on these NSDUH results, they were very likely to have met DSM-IV criteria for abuse/dependence.

For comparison with the UHS heroin addiction cases, we used other dbGaP study cohorts with genotype data available as a source of controls (Table S2). Our pipeline used to generate a data set for conducting genetic association testing with UHS cases and population controls is outlined in Figure S2. We conducted a meta-data survey to identify study cohorts that met each of the following criteria: 1) genome-wide SNP genotype data available from an Illumina platform; 2) participants aged 18 and older; 3) case/control, nested case/control, control set, longitudinal, cohort, or population-based study design; 4) availability of African American or Caucasian participants; and 5) consent group allowing for the study of substance addiction, either

specifically or under a general category such as general research or health research use. Our meta-data survey considered the inclusion of studies that were posted on dbGaP as of June 18, 2012. The six selected dbGaP study cohorts (Table S2) that passed through all steps in our selection pipeline (Figure S2) were genotyped on one of three Illumina BeadChip arrays (Omni1-Quad, 1M-Duo, or Omni2.5). Two of the selected study cohorts had data available on substance addiction ("Collaborative Genetic Study of Nicotine Dependence" and "Study of Addition: Genetics and Environment"), and participants meeting the criteria for DSM-IV (9) opioid drug dependence from these cohorts were not considered for inclusion as controls.

Quality control for the discovery cohort

Our quality control procedures, which were modeled after prior studies using population controls (10-15), were implemented on the heroin addiction cases and each of the dbGaP datasets, separately by study and by ancestral group, using PLINK software (16) unless otherwise stated. The numbers of genotyped subjects passing quality control are outlined in Table S2.

Before implementing our own quality control procedures on the dbGaP datasets, we filtered the data based on preliminary quality control criteria that were recommended by the original study investigators, where available. Then, we began our own quality control procedures by excluding genotyped participants that failed any of the following criteria: call rate <90%; sample duplication as indicated by an identity-by-state estimate >90%; first-degree relatedness as indicated by an identity-by-descent estimate >40%; gender discordance; or excessive homozygosity. In African Americans only, relatedness was corroborated using kinship coefficients (>0.177 being indicative of first degree relatedness) from the KING program, which

was designed to circumvent the inflation of IBD estimates due to population stratification (17). For both duplicates and relative pairs/clusters, we retained the subject with the highest call rate and excluded the other implicated participants. Genotyped SNPs were excluded for any of the following: call rate <90%, minor allele frequency (MAF) <1%, and Hardy Weinberg equilibrium (HWE) $p < 1 \times 10^{-4}$.

Study participants passing the aforementioned criteria were compared to HapMap phase III reference populations of YRI (West Africans), CEU (European Americans), and CHB (Chinese) to evaluate their ancestral classifications. The STRUCTURE program was applied to self-identified African Americans and European Americans, separately, with 10,000 SNPs randomly distributed across the genome. Participants analyzed as European Americans had <25% African ancestry, and participants analyzed as African Americans had >25% African ancestry.

To ensure that there were no underlying biases driven by any one of the dbGaP control datasets before combining them with heroin addiction cases from UHS, we computed the lambda, or genomic control inflation factor (λ_{gc}), for each pair-wise comparison of control datasets by using genotyped SNPs that overlap the two datasets, arbitrarily making case assignments to subjects from one dataset and control assignments to subjects from the other dataset, and running a logistic regression model to test SNP associations with the arbitrary case/control assignments. These comparisons revealed some bias between two of the African American control datasets ("Gene Environment Association Studies [GENEVA]: Genetics of Early Onset Stroke [GEOS] Study" and "A Multiethnic Genome-wide Scan of Prostate Cancer", $\lambda_{gc} = 1.10$), but this inflation was resolved by imposing a more stringent ancestral classification such that the few African American study participants with African ancestry <60% (the lower

limit of African ancestry for the ASW reference participants from HapMap) were removed. Following the additional exclusions, all of the resulting λ_{gc} values were $\leq$1.05, as shown in Table S3 for European Americans and Table S4 for African Americans.

Replication cohorts for heroin abuse association testing and their quality control

The CIDR – Gelernter Study cohort was assembled from three different genetics projects on alcohol dependence (with oversampling of African Americans), cocaine dependence, and opioid dependence, as previously described (18). All participants were administered an electronic version of the Semi-Structured Assessment for Drug Dependence and Alcoholism (19), from which DSM-IV (9) diagnoses of major psychiatric traits, including lifetime dependence on alcohol, opioid, and other drugs, were derived. Genotyping was conducted on the Illumina Omni1-Quad BeadChip. We began by applying preliminary quality control metrics, as recommended by the original study investigators and provided in dbGaP. We then applied quality control and imputation procedures mimicking those used for the discovery cohort.

Participants in the Australian Heroin Dependence Study were ascertained from the following four datasets. 1) The Comorbidity and Trauma Study is a retrospective case-control study examining genetic and environmental factors contributing to heroin dependence liability, as previously described (20). The study was run in collaboration with Washington University, the Queensland Institute of Medical Research, and the National Drug and Alcohol Research Centre at the University of New South Wales. Case participants were recruited from maintenance clinics providing opioid replacement therapy in the greater Sydney region of Australia. DSM-IV (9) opioid dependence diagnoses of the cases were based on administration of a comprehensive psychiatric interview based on the Semi-Structured Assessment of the Genetics of Alcoholism—

Australia (21) augmented with sections drawn from other instruments assessing childhood trauma exposure, family history, and screening for borderline personality disorder. Control participants were recruited from employment centers and community centers, open street malls, and local press servicing the same geographical area as the opioid maintenance treatment clinics and either denied recreational use of opioids or had used these drugs recreationally fewer than 11 times in their lifetime. Their prevalence of non-opioid licit drug dependence and illicit drug dependence as well as childhood trauma exposure and other psychiatric disorders is elevated considerably versus estimates of similar measures in Australian general population samples. All participants provided blood samples as a source of DNA. 2) The Western Australia Study on Heroin Dependence focused both on genetic contributions to heroin dependence and response to naltrexone treatment of the disorder and included heroin addicted individuals from the greater Perth region. Case participants with a DSM-IV diagnosis of heroin dependence completed a clinical assessment and provided blood samples during their treatment at the Perth Naltrexone Clinic, now named as the Fresh Start Recovery Programme. 3) The Twin Study of Mole Development in Adolescence, an ongoing investigation of melanocytic naevi, was used as a source of population controls. Parents of these twins served as controls, and although they were not assessed for heroin dependence, they have largely survived the period of risk for heroin dependence, and by virtue of their participation in this research, were very likely to have a prevalence of heroin dependence lower than that in the general population. 4) The Hunter Community Study was used as an additional source of population controls. As described elsewhere (22), the Hunter Community Study is a population-based cohort study established to assess factors important in the health, well-being, social functioning, and economic consequences of ageing. Genotyping of the four assembled cohorts, altogether referred to as the

Australian Heroin Dependence Study, was conducted on either the Illumina 660W-Quad or 610-Quad BeadChip arrays. We conducted quality control using metrics, similar to the discovery cohort, as follows. Genotyped SNPs were filtered due to deviation from HWE ($p \leq 10^{-6}$), MAF ($\leq$0.01), call rate ($\leq$95%), and GenomeStudio genotype quality score ($\leq$0.7). Additional analyses were performed to detect cryptic relatedness among genotyped participants using a pi-hat cut-off of 0.1. Principal components analysis (PCA) was performed with the SmartPCA program in the Eigensoft 3.0 package (23) to identify outliers of non-European ancestry, based their relative distance from the center of the northern European group. Participants who were more than six standard deviations from the centroid of the first two eigenvectors were removed.

Genotype imputation

To capture the nominated *cis*-eQTL SNPs for association testing with heroin abuse, imputation of SNP genotypes was conducted with reference to 1000 Genomes haplotype panels (24) in UHS and each of the dbGaP datasets. We previously found that combining imputed SNPs derived from cases and controls genotyped on different arrays has the potential to create artifactual differences leading to biased genotype-phenotype associations (25). The bias can be circumvented by basing imputation on the intersection of SNPs available across all genotyping arrays (25). Therefore, for the UHS vs. population control analysis, our imputation procedure was conducted separately by study and by ancestral group, using a common set of genotyped SNPs available across all European Americans or African Americans as the input genotypes for imputation.

Among the various imputation software programs that are available, we used IMPUTE2 (26) (version 2.2.2) because we previously found that it provides the highest imputation quality

for African American studies when using the 1000 Genomes "cosmopolitan" panel consisting of all available reference participants from 14 diverse populations (denoted ALL) instead of using smaller panels consisting of more closely related reference participants (27). Others have recommended the "cosmopolitan" panel as well (26, 28, 29). More specifically, we used 1000 Genomes reference haplotype panels from 1,092 participants of European (*n* = 379), African (*n* = 246), East Asian (*n* = 286), and admixed American (*n* = 181) ancestry (February 2012 integrated variant set release version 3 available at http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html). Imputations were preceded by prephasing the study genotypes with the ShapeIT (version 2) program (30) to estimate haplotypes, using 500 conditioning states, recommended effective population sizes of 15,000 for African Americans and 11,418 for European Americans, and default settings for all other program options. The estimated study haplotypes were then input into IMPUTE2 to impute SNP genotypes based on the highly dense set of SNPs on the 1000 Genomes ALL reference haplotype panel. Imputations were conducted on 4.5 MB chunks with 1MB flanking buffers. Default options were used, except for 1) specifying "k_hap" as 468 haplotypes for African Americans and 170 haplotypes for European Americans and 2) applying the "–filt_rules_1" option to remove SNPs that are monomorphic in both the AFR and EUR panels. For each imputed variant, IMPUTE2 output included probabilities for each of the three genotypes and an "info" metric (a fractional value typically between 0 and 1 with higher values being indicative of SNPs imputed with higher certainty). The genotype probabilities were converted into a single imputed genotype dosage value (a fractional value between 0 and 2 indicating the expected number of minor allele copies) to use for association testing.

Statistical analyses: Association testing with heroin addiction

SNPs identified from the *cis*-eQTL mapping were tested for association with heroin addiction in the UHS. Significantly associated SNPs in the UHS were then tested for replication in the CIDR – Gelernter Study and the Australian Heroin Dependence Study. Observed SNP genotypes or imputed SNP genotype dosages were tested for association using logistic regression models, separately by cohort and by ancestry group, in ProbABEL (31) (UHS and CIDR – Gelernter Study) or PLINK (32) (Australian Heroin Dependence Study) with adjustment for sex and selected eigenvectors to remove any bias due to population stratification. In the UHS and the CIDR – Gelernter Study, EIGENSTRAT (33) analyses were run, separately by cohort and by ancestry group, using a pruned set of genotyped SNPs in linkage equilibrium ($r^2 < 0.5$). The top three ancestry-specific eigenvectors, which together explained >90% of the phenotype variance, were selected for each ancestry group in UHS and for the CIDR – Gelernter Study. For the Australian Heroin Dependence Study, although the participants were all of European ancestry, PCA was conducted using the SmartPCA program to determine whether additional admixture correction was needed for association analyses. The r^2 setting of 0.8 was used to remove SNPs in high linkage disequilibrium with others in the panel. PCA generated seven eigenvectors to include when testing associations with heroin addiction.

Table S1. Concordance between heroin use threshold (past 30 days) for case designation and DSM-IV heroin abuse/dependence (past 12 months) in the National Household Survey on Drug use and Health data 2008-2012.

Times used heroin in past 30 days	**DSM-IV heroin abuse/dependence No** **Weighted *n* (unweighted *n*)**	**DSM-IV heroin abuse/dependence Yes** **Weighted *n* (unweighted *n*)**
< 10	17,064 (14)	29,409 (45)
≥ 10	11,396 (9)	74,538 (93)

Positive predictive value based on weighted data = 86.74%.

Table S2. Heroin addiction cases from the Urban Health Study, population controls from dbGaP, and independent cohorts of heroin addiction cases and controls.

Study cohort	Ref.	dbGaP accession number(s)	dbGaP consent groups	Ancestry group	No. genotyped participants passing QC	
					Heroin addiction cases	Controls
Urban Health Study	(5, 6)	phs000454.v1.p1	Health research (non-commercial)	European Americans	711	-
				African Americans	1,293	-
Study of Addiction: Genetics and Environment (SAGE)[a]	(34)	phs000092.v1.p1	Health research; Alcoholism and related conditions	European Americans	-	2,386
				African Americans	-	1,189
Collaborative Genetic Study of Nicotine Dependence (COGEND)[a]	(35)	phs000092.v1.p1 phs000404.v1.p1	General research use	European Americans	-	886
				African Americans	-	446
Gene Environment Association Studies (GENEVA): Genetics of Early Onset Stroke (GEOS) Study	(36)	phs000292.v1.p1	Studies of genes for medical conditions	European Americans	-	90
				African Americans	-	119
A Multiethnic Genome-wide Scan of Prostate Cancer	(37, 38)	phs000306.v2.p1	General research use; Health research and methods	African Americans	-	615
Genome-Wide Association Study of Parkinson Disease: Genes and Environment	(39)	phs000196.v2.p1	General research use	European Americans	-	1,982

Study cohort	Ref.	dbGaP accession number(s)	dbGaP consent groups	Ancestry group	No. genotyped participants passing QC	
					Heroin addiction cases	Controls
High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation	(40, 41)	phs000187.v1.p1	General research use	European Americans	-	1,040
GWAS of Heroin Dependence (Australian Heroin Dependence Study)	(20)	phs000277.v1.p1	Substance use dependence only; Research on alcohol and tobacco related phenotypes; Substance dependence, childhood trauma and related disorders; Alcohol related conditions	Australians of European ancestry	1,976	3,144[b]
Alcohol Dependence GWAS in European- and African-Americans (CIDR –Gelernter Study)	(18)	phs000425.v1.p1	Research on alcohol and alcohol related phenotypes/disorders	African Americans	307	545

QC, quality control.

[a]SAGE is composed of three component studies, including the Collaborative Study on the Genetics of Alcoholism, the Family Study of Cocaine Dependence, and COGEND. Additional COGEND participants not included in SAGE were analyzed separately.

[b]A subset of the control group was comprised of participants from The Hunter Community Study, which are not available via dbGaP accession number phs000277.v1.p1.

Table S3. Genomic inflation factors from pairwise comparison of control study cohorts with European Americans.

	SAGE	COGEND	GENEVA Stroke	Parkinson's	Melanoma
SAGE	--	0.96	0.99	1.05	1.05
COGEND	--	--	1.00	1.04	1.04
GENEVA Stroke	--	--	--	1.01	1.00
Parkinson's	--	--	--	--	1.04

Table S4. Genomic inflation factors from pairwise comparison of control study cohorts with African Americans.

	SAGE	COGEND	GENEVA Stroke	Prostate
SAGE	--	0.97	1.00	1.05
COGEND	--	--	0.99	1.00
GENEVA Stroke	--	--	--	0.98

Table S5. *Cis*-eQTL analyses of 103 SNPs spanning *OPRM1* and its flanking region (±100kb). SNP genotypes were tested for association with *OPRM1* expression level in prefrontal cortex samples from 110 European Americans and 114 African Americans, who were aged from 0 to 78 years old. SNPs are ordered by chromosomal position. *P* values less than 0.05 are shown in bold.

SNP	Base pair position (NCBI build 36)	SNP type	Distance to *OPRM1*	1000 Genomes EUR panel		1000 Genomes AFR panel		*P*, overall	*P*, European Americans only	*P*, African Americans only
				Minor allele	Minor allele frequency	Minor allele	Minor allele frequency			
rs204426	154,305,678	intergenic	-96,457	T	0.38	C	0.47	0.68	0.87	0.16
rs6907526	154,307,441	intergenic	-94,694	T	0.0013	T	0.16	0.10	NA[a]	0.091
rs9397673	154,308,402	intergenic	-93,733	G	0.23	G	0.14	0.28	0.22	0.31
rs7745892	154,308,430	intergenic	-93,705	C	0.37	T	0.47	0.98	0.95	0.85
rs654359	154,311,095	intergenic	-91,040	C	0.38	C	0.46	0.44	0.59	0.48
rs7758009	154,311,203	intergenic	-90,932	C	0.19	C	0.09	0.89	0.66	0.31
rs7744543	154,312,928	intergenic	-89,207	A	0.04	A	0.11	0.98	0.23	0.41
rs598832	154,314,311	intergenic	-87,824	G	0.17	G	0.06	0.31	0.31	0.076
rs574350	154,316,536	intergenic	-85,599	G	0.50	A	0.45	0.69	0.95	0.22
rs12174208	154,320,077	intergenic	-82,058	T	0.19	T	0.09	0.82	0.64	0.42
rs647303	154,324,902	intergenic	-77,233	C	0.27	C	0.14	0.79	0.28	**0.0010**
rs11759340	154,325,531	intergenic	-76,604	C	0.10	C	0.06	0.56	0.72	0.31
rs17084782	154,326,487	intergenic	-75,648	A	0.11	A	0.05	0.29	0.38	0.93
rs6900805	154,329,725	intergenic	-47,271	G	0.27	G	0.13	0.79	0.28	**0.0010**

SNP	Base pair position (NCBI build 36)	SNP type	Distance to *OPRM1*	1000 Genomes EUR panel		1000 Genomes AFR panel		*P*, overall	*P*, European Americans only	*P*, African Americans only
				Minor allele	Minor allele frequency	Minor allele	Minor allele frequency			
rs10457901	154,336,610	intergenic	-65,525	G	0.09	G	0.24	0.79	0.35	0.63
rs9371769	154,349,428	intergenic	-52,707	T	0.35	T	0.21	0.95	0.96	0.93
rs1675903	154,351,287	intergenic	-50,848	A	0.14	A	0.10	0.060	0.54	0.083
rs10499276	154,351,501	intergenic	-50,634	T	0.11	T	0.06	0.44	0.42	1.00
rs12208106	154,352,661	intergenic	-49,474	C	0.04	C	0.06	0.27	0.35	0.25
rs1772945	154,353,978	intergenic	-48,157	A	0.50	A	0.47	0.68	0.60	0.98
rs9384169	154,354,864	intergenic	-47,271	C	0.35	C	0.36	0.25	0.90	**0.046**
rs6914262	154,356,609	intergenic	-45,526	A	0.39	A	0.48	0.72	0.92	0.84
rs1675902	154,357,060	intergenic	-45,075	A	0.10	A	0.01	0.29	0.65	0.46
rs1319211	154,358,275	intergenic	-43,860	G	0.27	G	0.08	0.73	0.85	0.89
rs7738859	154,359,447	intergenic	-42,688	T	0.35	T	0.29	0.95	0.93	0.71
rs4870261	154,361,495	intergenic	-40,640	T	0.15	T	0.07	0.81	0.95	0.81
rs9478495	154,365,602	intergenic	-36,533	A	0.15	A	0.24	0.22	0.89	**0.015**
rs9285541	154,372,322	intergenic	-29,813	G	0.20	G	0.04	0.63	0.84	0.86
rs1319339	154,372,877	intergenic	-29,258	C	0.14	C	0.01	0.79	0.68	0.47
rs712242	154,374,865	intronic	-27,270	C	0.49	C	0.41	0.45	0.90	0.43

SNP	Base pair position (NCBI build 36)	SNP type	Distance to *OPRM1*	1000 Genomes EUR panel		1000 Genomes AFR panel		*P*, overall	*P*, European Americans only	*P*, African Americans only
				Minor allele	Minor allele frequency	Minor allele	Minor allele frequency			
rs9478496	154,374,876	intronic	-27,259	C	0.15	C	0.10	0.99	0.97	0.93
rs712244	154,376,843	intronic	-25,292	G	0.29	G	0.03	0.62	0.86	0.43
rs1294092	154,377,366	intronic	-24,769	G	0.20	G	0.02	0.61	0.80	0.88
rs9478498	154,383,069	intronic	-19,066	T	0.22	C	0.48	0.37	0.21	0.61
rs7776341	154,390,298	intronic	-11,837	C	0.04	C	0.21	0.55	0.54	0.63
rs6912029	154,402,201	5'-UTR	0	T	0.04	T	0.09	0.60	0.83	0.43
rs1799971	154,402,490	missense	0	G	0.16	G	0.01	0.55	0.73	0.75
rs510769	154,403,712	intronic	0	T	0.23	T	0.22	0.45	0.90	0.43
rs514980	154,409,860	intronic	0	C	0.19	C	0.12	0.69	0.76	0.86
rs499796	154,414,385	intronic	0	G	0.19	G	0.46	0.55	0.72	0.72
rs3778146	154,419,618	intronic	0	C	0.15	C	0.12	0.16	0.078	0.94
rs3823010	154,420,845	intronic	0	A	0.15	A	0.12	0.12	0.12	0.98
rs3778150	154,425,351	intronic	0	C	0.15	C	0.20	0.089	**0.042**	0.38
rs6927269	154,429,999	intronic	0	G	0.15	G	0.10	0.11	0.12	0.84
rs4357147	154,431,672	intronic	0	G	0.0013	G	0.08	0.14	NA[a]	0.16
rs6923231	154,433,828	intronic	0	A	0.09	A	0.04	0.72	0.79	0.78

SNP	Base pair position (NCBI build 36)	SNP type	Distance to *OPRM1*	1000 Genomes EUR panel		1000 Genomes AFR panel		*P*, overall	*P*, European Americans only	*P*, African Americans only
				Minor allele	Minor allele frequency	Minor allele	Minor allele frequency			
rs3778153	154,435,577	intronic	0	A	0.14	A	0.11	0.29	0.39	0.96
rs563649	154,449,660	5'-UTR	0	T	0.08	T	0.09	0.50	0.26	0.83
rs9479756	154,451,807	intronic	0	A	0.10	A	0.16	0.57	0.71	0.59
rs562859	154,456,266	synonymous	0	C	0.32	C	0.41	0.054	**0.016**	0.96
rs12203621	154,459,438	intronic	0	A	0.05	A	0.08	0.38	0.83	0.26
rs3798683	154,460,107	intronic	0	A	0.16	A	0.01	0.57	0.76	0.70
rs647192	154,464,437	intronic	0	G	0.22	G	0.05	**0.0085**	**0.0018**	0.63
rs9322447	154,466,013	intronic	0	A	0.45	G	0.35	0.37	0.57	0.60
rs650825	154,470,229	intronic	0	A	0.22	A	0.17	**0.0034**	**0.0017**	0.45
rs650245	154,470,394	3'-UTR	0	A	0.10	A	0.28	0.72	0.68	0.68
rs632395	154,478,944	intronic	0	A	0.10	A	0.20	0.64	0.68	0.64
rs512053	154,481,209	intronic	0	T	0.08	T	0.01	0.78	0.58	0.69
rs671531	154,482,434	intronic	148	A	0.32	A	0.43	**0.015**	**0.010**	0.92
rs558948	154,483,556	intronic	1,270	T	0.22	T	0.16	**0.0017**	**0.00094**	0.71
rs10485058	154,486,907	intronic	4,621	G	0.17	G	0.05	0.53	0.51	0.83
rs613355	154,491,542	intronic	9,256	C	0.32	C	0.40	**0.010**	**0.012**	0.72

SNP	Base pair position (NCBI build 36)	SNP type	Distance to *OPRM1*	1000 Genomes EUR panel		1000 Genomes AFR panel		*P*, overall	*P*, European Americans only	*P*, African Americans only
				Minor allele	Minor allele frequency	Minor allele	Minor allele frequency			
rs636433	154,494,178	intronic	11,892	A	0.10	A	0.09	0.49	0.94	0.25
rs538174	154,494,229	intronic	11,943	C	0.22	C	0.05	**0.012**	**0.0028**	1.00
rs544093	154,499,185	intronic	16,899	G	0.10	G	0.13	0.39	0.94	0.28
rs2010884	154,499,899	intronic	17,613	A	0.22	A	0.10	0.22	0.42	0.15
rs655059	154,500,548	intronic	18,262	C	0.06	C	0.01	0.86	0.67	0.47
rs13203628	154,502,815	intronic	20,529	G	0.24	G	0.35	**0.037**	**0.010**	0.82
rs13193545	154,503,514	intronic	21,228	C	0.19	C	0.01	0.063	0.064	0.81
rs518596	154,504,070	intronic	21,784	G	0.37	A	0.40	0.12	0.14	0.77
rs10223804	154,504,084	intronic	21,798	A	0.18	A	0.01	0.27	0.43	0.74
rs6557337	154,504,746	intronic	22,460	T	0.38	C	0.44	0.71	0.58	0.82
rs483481	154,505,660	intronic	23,374	A	0.43	A	0.31	0.96	0.95	0.91
rs4869817	154,508,510	intronic	26,224	A	0.47	A	0.49	0.68	0.69	0.48
rs9397178	154,511,909	intronic	29,623	G	0.07	G	0.07	0.89	0.58	0.98
rs13196610	154,512,887	intronic	30,601	G	0.07	G	0.06	0.45	0.50	0.69
rs4870268	154,514,019	intronic	31,733	C	0.47	C	0.47	0.81	0.77	0.99
rs17277929	154,519,605	3'-UTR	37,319	C	0.15	C	0.06	0.91	0.96	0.27

SNP	Base pair position (NCBI build 36)	SNP type	Distance to *OPRM1*	1000 Genomes EUR panel		1000 Genomes AFR panel		*P*, overall	*P*, European Americans only	*P*, African Americans only
				Minor allele	Minor allele frequency	Minor allele	Minor allele frequency			
rs2236256	154,520,132	3'-UTR	37,846	C	0.42	C	0.45	0.29	0.22	0.74
rs1950005	154,523,086	intronic	40,800	G	0.42	G	0.47	0.31	0.22	0.69
rs1918760	154,523,859	intronic	41,573	C	0.34	C	0.20	0.12	0.11	0.66
rs2281617	154,529,113	intronic	46,827	T	0.12	T	0.28	0.87	0.36	0.76
rs6939625	154,529,487	intronic	47,201	A	0.05	A	0.28	0.98	0.86	0.58
rs6941251	154,529,669	intronic	47,383	C	0.41	C	0.22	0.63	0.48	0.97
rs12208947	154,532,829	intronic	50,543	A	0.12	A	0.19	0.67	0.36	0.55
rs7745499	154,534,546	intronic	52,260	A	0.41	A	0.20	0.57	0.52	0.96
rs13437608	154,541,831	intronic	59,545	T	0.06	T	0.11	0.41	0.82	0.32
rs17278409	154,542,817	intronic	60,531	T	0.21	T	0.28	0.88	0.63	0.92
rs9479779	154,544,380	intronic	62,094	G	0.07	G	0.08	0.30	0.57	**0.042**
rs11965988	154,545,204	intronic	62,918	C	0.07	C	0.19	0.94	0.47	0.97
rs9371331	154,545,564	intronic	63,278	G	0.25	G	0.12	0.85	0.94	0.52
rs7452342	154,554,189	intronic	71,903	T	0.35	T	0.20	0.20	0.15	0.77
rs4626436	154,554,890	intronic	72,604	A	0.05	A	0.22	0.68	0.48	0.30
rs10485060	154,557,793	intronic	75,507	A	0.06	A	0.02	**0.039**	0.34	0.24

SNP	Base pair position (NCBI build 36)	SNP type	Distance to *OPRM1*	1000 Genomes EUR panel		1000 Genomes AFR panel		*P*, overall	*P*, European Americans only	*P*, African Americans only
				Minor allele	Minor allele frequency	Minor allele	Minor allele frequency			
rs1040822	154,560,330	intronic	78,044	G	0.41	G	0.31	0.62	0.88	0.61
rs2295676	154,562,084	intronic	79,798	A	0.31	A	0.11	0.052	0.074	0.62
rs9322451	154,564,353	intronic	82,067	A	0.20	A	0.26	0.17	0.44	0.59
rs1852629	154,564,432	intronic	82,146	T	0.39	T	0.40	**0.044**	0.090	0.38
rs6557339	154,569,846	intronic	87,560	C	0.28	C	0.23	0.30	0.060	0.95
rs9479791	154,571,206	intronic	88,920	T	0.15	T	0.24	0.43	0.37	0.49
rs6913456	154,572,672	intronic	90,386	G	0.35	G	0.34	0.55	0.21	0.39
rs790252	154,576,104	intronic	93,818	A	0.20	A	0.12	0.21	0.21	0.80
rs9478526	154,577,270	intronic	94,984	C	0.08	C	0.13	0.42	0.57	0.32

NCBI, National Center for Biotechnology Information; SNP, single nucleotide polymorphism; UTR, untranslated region.
[a]The association result was not available (NA), because the SNP was monomorphic.

Table S6. Regulatory annotation, according to HaploReg, for the 16 newly implicated *cis*-eQTL SNPs. HaploReg output includes information on 7 regulatory features: SiPhy (SIte-specific PHYlogenetic) conservation, promoter histone marks, enhancer histone markers from the Roadmap Epigenome Mapping Consortium, DNAse hypersensitivity from ENCODE, proteins bound, eQTL tissues from the Genotype-Tissue Expression eQTL browser, and motifs changed based on position weight matrix scores from ENCODE to identify transcription factor binding sites. Other HaploReg regulatory features, for which none of the *OPRM1* SNPs were implicated, are not shown. A dash ("-") indicates that the regulatory feature was not indicated for a specific SNP. SNPs are sorted by chromosomal position.

SNP	Position (NBCI build 36)	SiPhy conservation	Enhancer histone marks	DNAse	Motifs changed
rs647303	154,324,902	-	-	-	ATF2, Cdx, HLF, Hsf, lrf, Pou5f1
rs6900805	154,329,725	-	-	-	-
rs9384169	154,354,864	-	-	-	GR,T3R
rs9478495	154,365,602	-	-	-	Mrg1::Hoxa9,NF-I
rs3778150	154,425,351	Yes	-	-	NFAT, NFAT1
rs562859	154,456,266	-	-	BE2_C	Pou2f2
rs647192	154,464,437	-	-	-	Arid5b,Brachyury,Foxa
rs650825	154,470,229	-	-	-	HNF4,Smad3
rs671531	154,482,434	-	-	-	-
rs558948	154,483,556	-	-	-	Foxo, Sox, TCF11::MafG
rs613355	154,491,542	-	-	-	Foxa, Foxc1, GR, HDAC2, TCF12, p300
rs538174	154,494,229	-	-	-	Foxp1
rs13203628	154,502,815	-	H1	H1-hESC, CD34+_Mobilized, CMK, H7-hESC, Monocytes-CD14+_RO01746, NT2-D1	Hmx, Myb, RFX5
rs9479779	154,544,380	-	-	-	-

SNP	Position (NBCI build 36)	SiPhy conservation	Enhancer histone marks	DNAse	Motifs changed
rs10485060	154,557,793	-	-	-	HNF4,Pax-4,RXRA
rs1852629	154,564,4	-	-	-	SIX5,SRF

NCBI, National Center for Biotechnology Information; SNP, single nucleotide polymorphism.

Table S7. Four SNPs tested for association with heroin addiction in two independent replication cohorts, the CIDR – Gelernter Study and the Australian Heroin Dependence Study. SNPs are sorted by their meta-analysis *P* value across the replication cohorts.

CIDR – Gelernter Study, African Americans (*n* = 307 cases and 545 controls)			**Australian Heroin Dependence Study, Australians of European ancestry (*n* = 1,976 cases and 3,144 controls)**			**Meta-analysis *P* across replication cohorts**	**Meta-analysis *P* across discovery and replication cohorts**
SNP (Minor Allele)	***P***	**OR (95% CI)**	**SNP (Minor Allele)**	***P***	**OR (95% CI)**		
rs3778150 (C)	0.17	1.20 (0.93-1.59)	rs3778150 (C)	1.7×10^{-4}	1.23 (1.11-1.39)	6.3×10^{-5}	4.3×10^{-8}
rs9478495 (A)	0.76	1.04 (0.81-1.33)	rs9478495 (A)	0.087	1.10 (0.99-1.23)	0.090	3.2×10^{-5}
rs9384169 (C)	0.95	0.99 (0.79-1.25)	rs9371769 (A)[a]	0.44	1.04 (0.95-1.13)	0.49	3.8×10^{-3}
rs562859 (C)	0.55	1.07 (0.86-1.34)	rs548646 (A)[a]	0.90	1.01 (0.91-1.09)	0.73	1.3×10^{-2}

CI, confidence interval; OR, odds ratio; SNP, single nucleotide polymorphism.

[a]Rs9384169 and rs562859 were not available as genotyped SNPs in the Australian Heroin Dependence Study, so the proxy SNPs rs9371769 (r^2 = 0.97 and D' = 0.98 with rs9384169 in the 1000 Genomes EUR panel) and rs548646 (r^2 = 0.98 and D' = 0.99 with rs562859) were used.

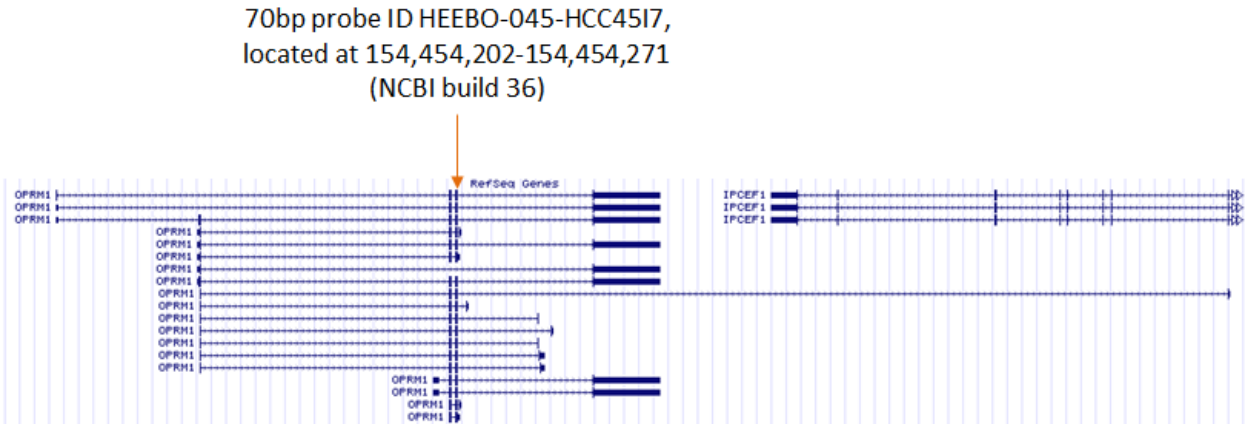


Figure S1. Location of the single *OPRM1* probe available in BrainCloud in relation to the known mRNA transcripts. NCBI, National Center for Biotechnology Information.

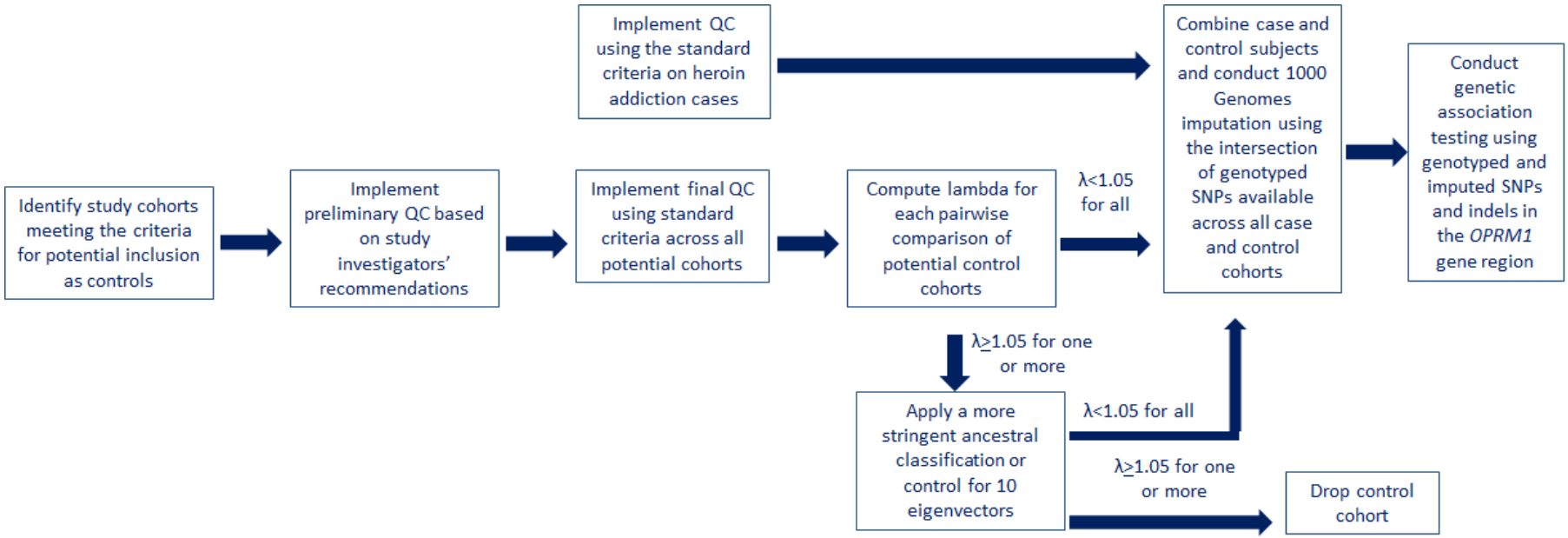


Figure S2. Pipeline for generating a control data set for comparison to heroin addiction cases from the Urban Health Study. QC, quality control; SNP, single nucleotide polymorphism.

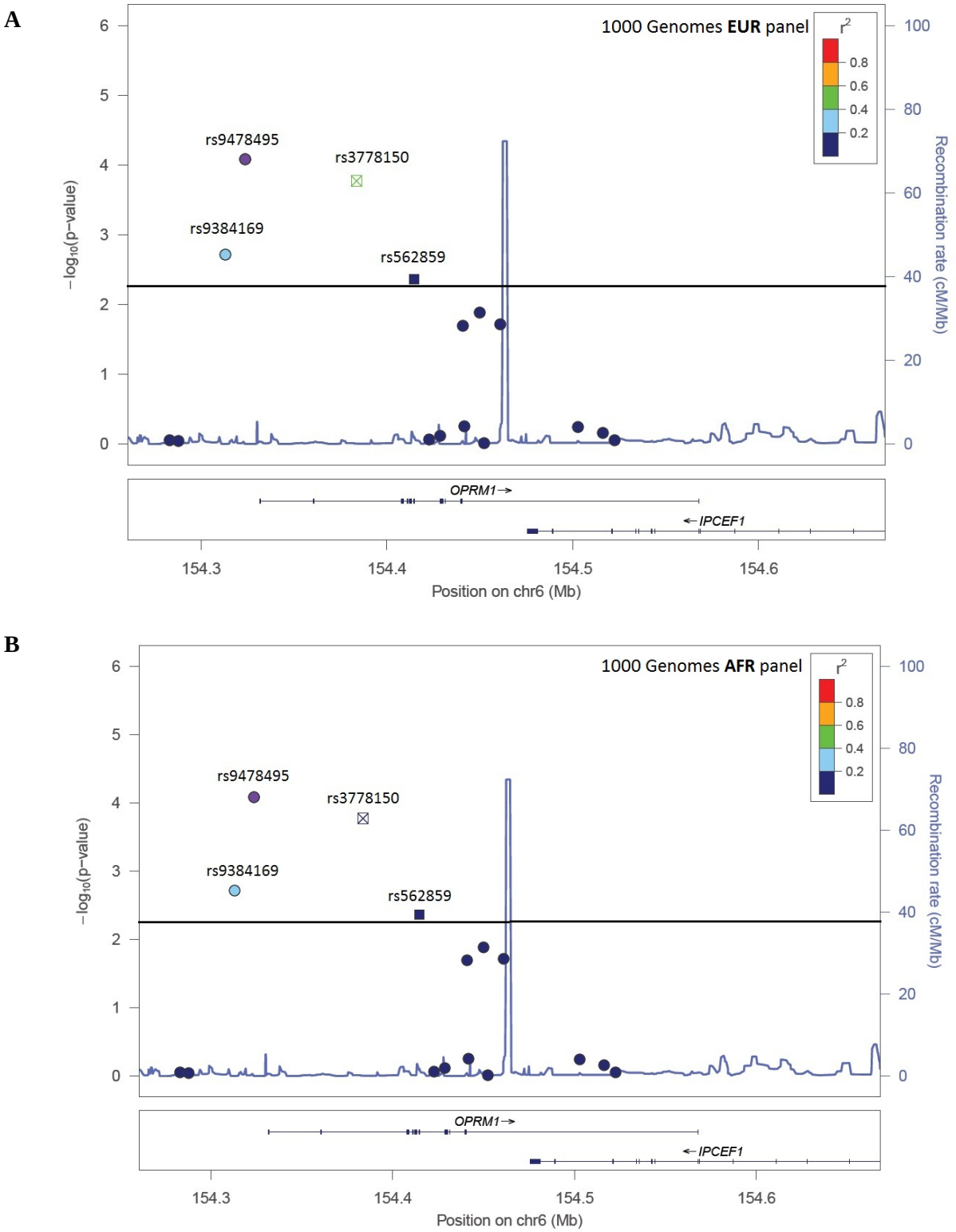
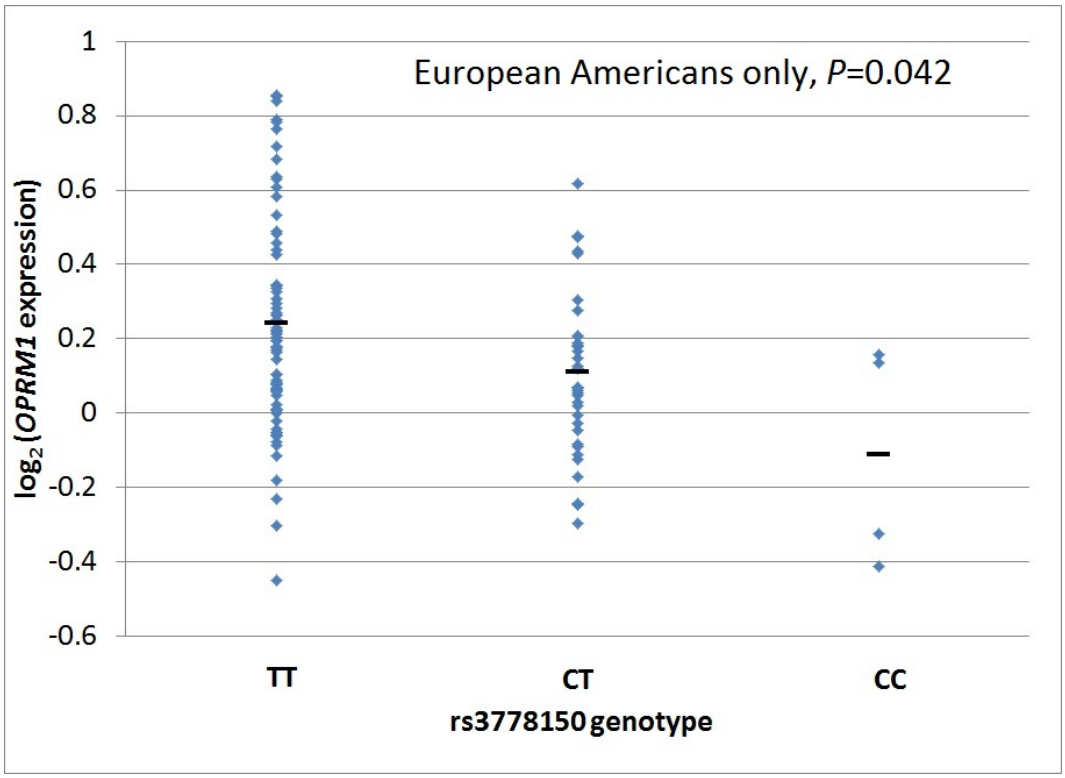


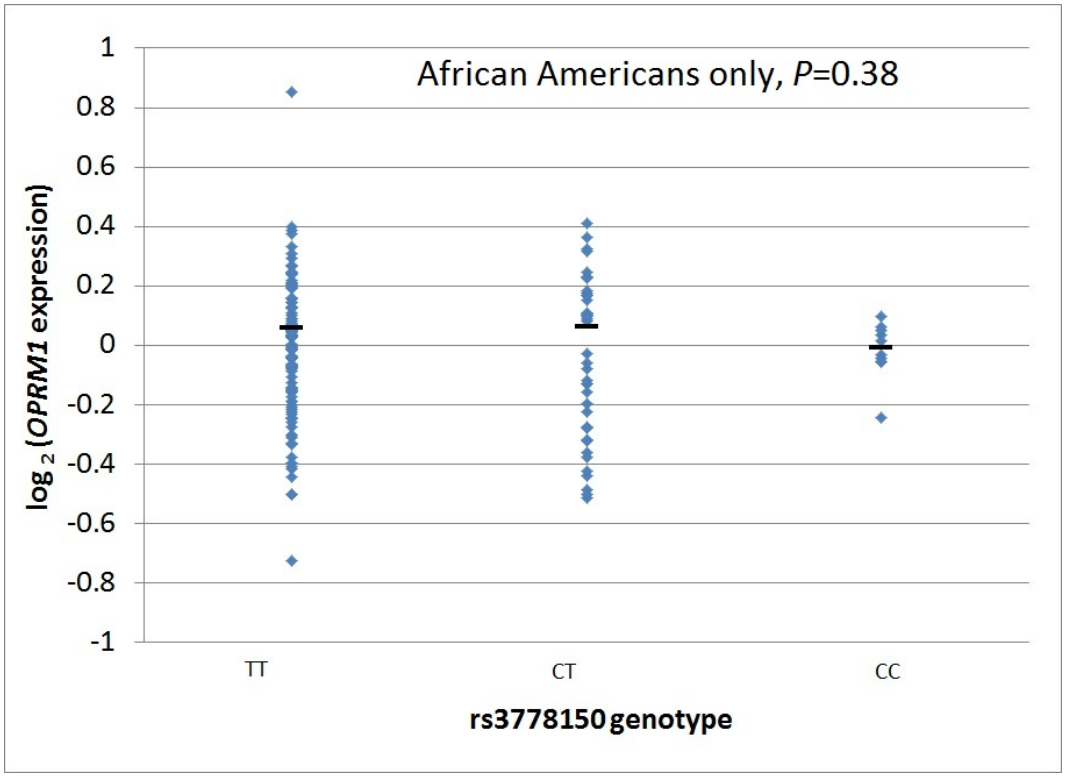
Figure S3. Linkage disequilibrium patterns among putative *cis*-eQTL single nucleotide polymorphisms (SNPs) for *OPRM1* and tested for association with heroin addiction. The $-\log_{10}$ (*P*) values from the multiancestry meta-analysis in the Urban Health Study are plotted by SNP

chromosomal position, and the four SNPs above the solid black line (meta-analysis $P < 0.005$) were tested for replication. Correlations between rs9478495 (in purple) and the 15 other putative *cis*-eQTL SNPs are shown with reference to 1000 Genomes populations of (**A**) European (denoted EUR) and (**B**) African (denoted AFR) ancestry. Annotation shows the SNPs that are located in an intergenic or intronic region (solid circles), an exon (solid square indicative of a synonymous SNP), or a region highly conserved in placental mammals (square with diagonal lines).

A


European Americans only, P=0.042
log2 (OPRM1 expression)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
TT
CT
CC
rs3778150 genotype

B


African Americans only, P=0.38
log 2 (OPRM1 expression)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
TT
CT
CC
rs3778150 genotype

C

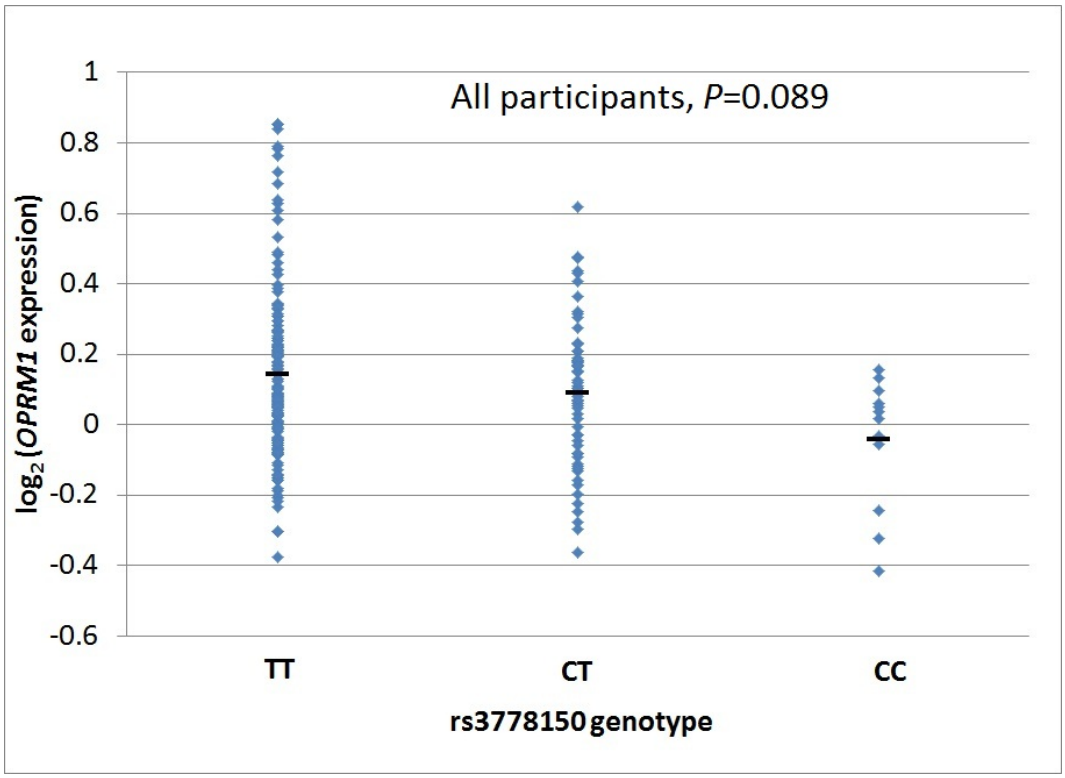
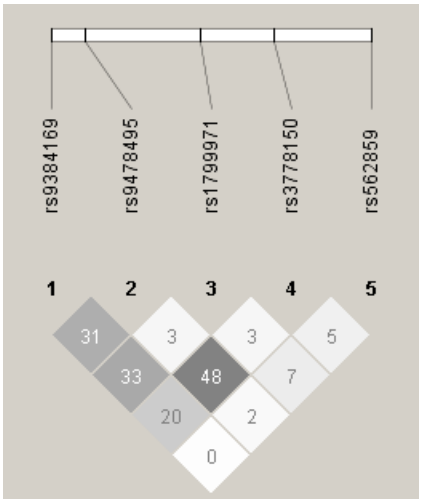


Figure S4. *OPRM1* expression as stratified by rs3778150 genotype in the BrainCloud cohort. The $\log_2$ ratio of sample to reference *OPRM1* expression is presented for (**A**) European Americans only, (**B**) African Americans only, and (**C**) all participants combined.

A



B

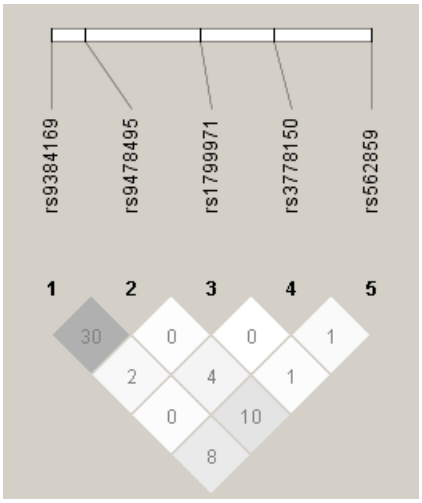


Figure S5. Linkage disequilibrium structure, based on r^2 values, among 4 putative *cis*-eQTL single nucleotide polymorphism (SNPs) significantly associated with heroin addiction in the Urban Health Study (rs9384169, rs9478495, rs3778150, and rs562859) and the previously implicated SNP rs1799971. Lighter gray to black shading indicates lower to higher r^2 values, respectively, in 1000 Genomes reference populations of (**A**) European ancestry and (**B**) African ancestry.

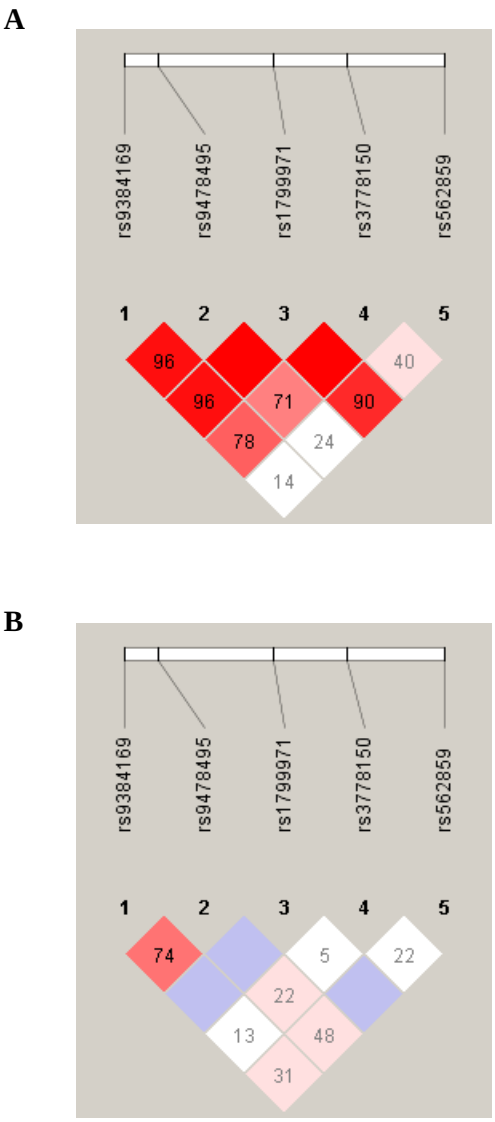


Figure S6. Linkage disequilibrium structure, based on D' values, among 4 putative *cis*-eQTL single nucleotide polymorphism (SNPs) significantly associated with heroin addiction in the Urban Health Study (rs9384169, rs9478495, rs3778150, and rs562859) and the previously implicated SNP rs1799971. Pink to red shading indicates lower to higher D' values, respectively, in 1000 Genomes reference populations of (**A**) European ancestry or (**B**) African ancestry. Purple shading indicates little statistical support for the observed correlation.

	Cohort	Ancestry Group	N, cases	N, controls
1	UHS	European Americans	711	6,384
2	UHS	African Americans	1,293	2,369
3	CIDR - Gelernter	African Americans	307	545
4	Australian Heroin Dep.	Australians	1,976	3,144

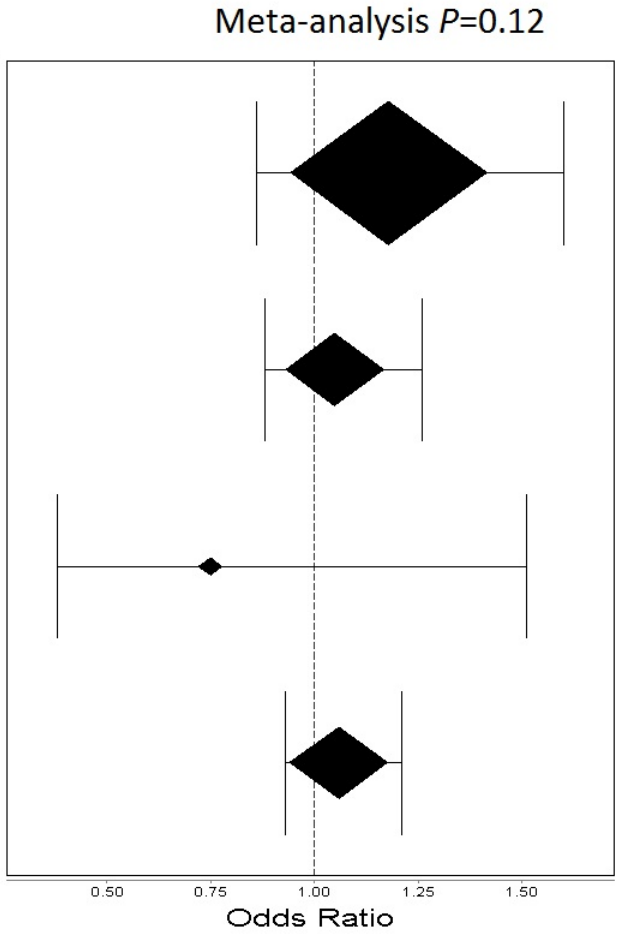
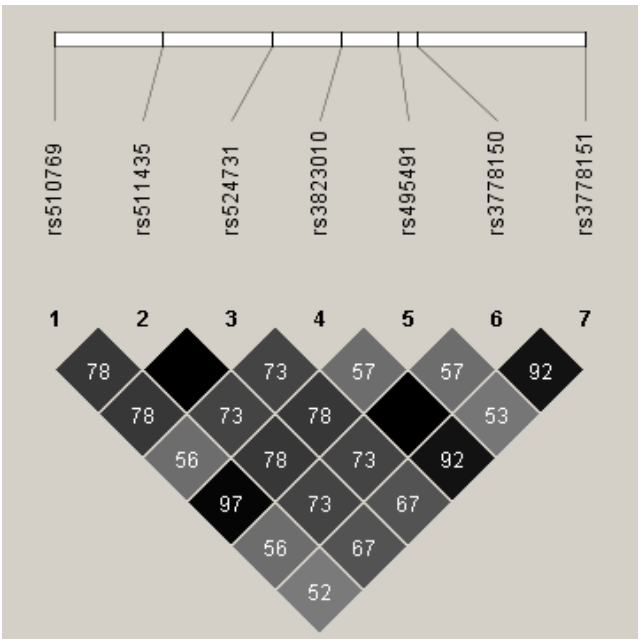


Figure S7. Forest plot of association results for the rs1799971-A allele across all cohorts (Urban Health Study [UHS], CIDR – Gelernter Study, and Australian Heroin Dependence Study) and ancestry groups (European Americans, African Americans, and Australians of European ancestry).

A



B

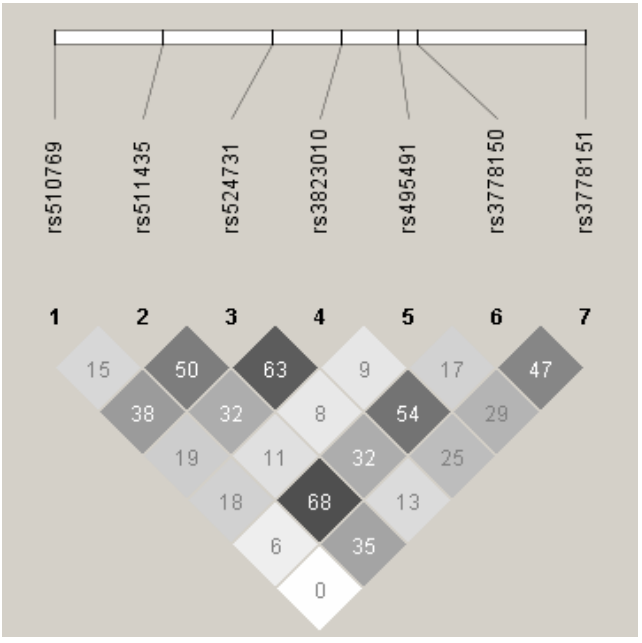


Figure S8. Linkage disequilibrium structure, based on r^2 values, for rs3778150 and six other intron 1 single nucleotide polymorphisms having prior suggestive associations with heroin addiction: rs510769, rs511435, rs524731, rs3823010, rs495491, and rs3778151. Lighter gray to black shading indicates lower to higher r^2 values, respectively, in 1000 Genomes reference populations of (**A**) European ancestry and (**B**) African ancestry.

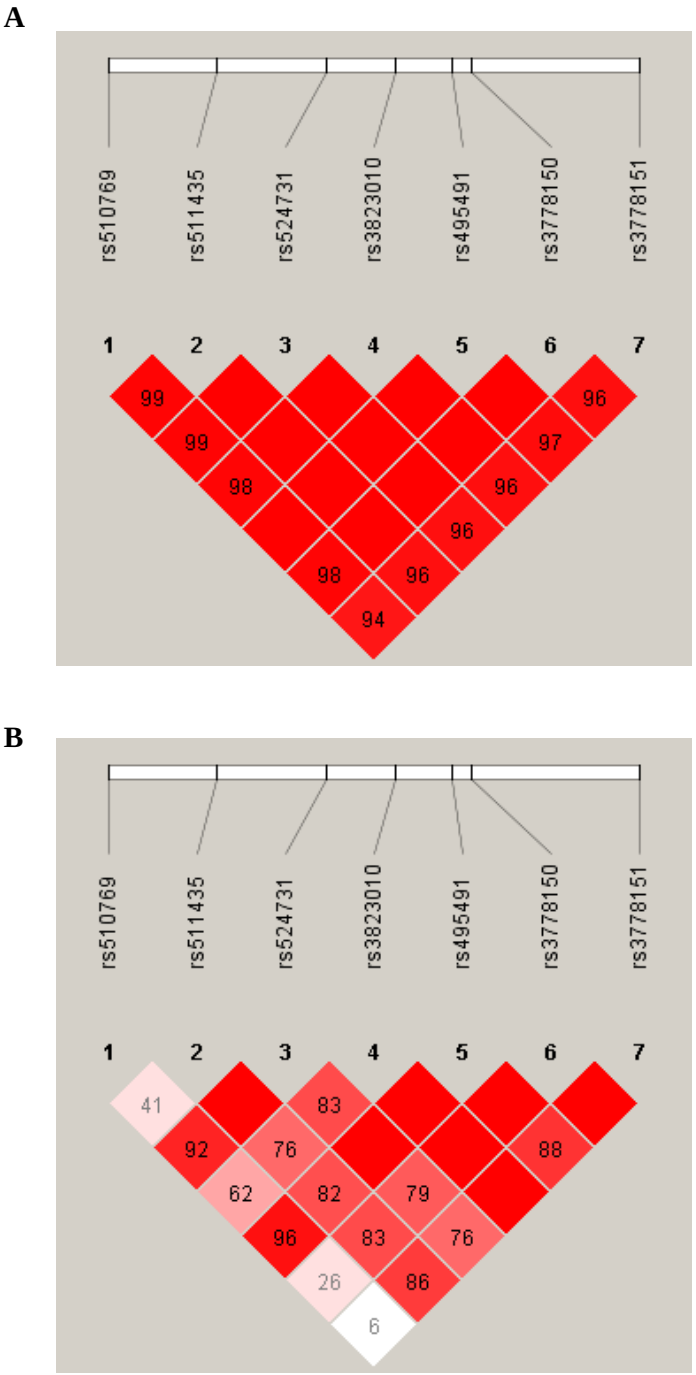


Figure S9. Linkage disequilibrium structure, based on D' values, for rs3778150 and six other intron 1 single nucleotide polymorphisms (SNPs) having prior suggestive associations with heroin addiction: rs510769, rs511435, rs524731, rs3823010, rs495491, and rs3778151. Pink to red shading indicates lower to higher D' values, respectively, in 1000 Genomes reference populations of (**A**) European ancestry or (**B**) African ancestry.

Supplemental References

1. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, *et al.* (2011): Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 478:519-523.

2. Lipska BK, Deep-Soboslay A, Weickert CS, Hyde TM, Martin CE, Herman MM, *et al.* (2006): Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia. *Biological Psychiatry*. 60:650-658.

3. Watters JK, Biernacki P (1989): Targeted sampling: options for the study of hidden populations. *Soc Problems*. 36:416-430.

4. Kral AH, Bluthenthal RN, Lorvick J, Gee L, Bacchetti P, Edlin BR (2001): Sexual transmission of HIV-1 among injection drug users in San Francisco, USA: risk-factor analysis. *Lancet*. 357:1397-1401.

5. Kral AH, Lorvick J, Gee L, Bacchetti P, Rawal B, Busch M, *et al.* (2003): Trends in human immunodeficiency virus seroincidence among street-recruited injection drug users in San Francisco, 1987-1998. *Am J Epidemiol*. 157:915-922.

6. Johnson EO, Hancock DB, Gaddis NC, Levy JL, Page G, Novak SP, *et al.* (in press): Novel genetic locus implicated for HIV-1 acquisition with putative regulatory links to HIV replication and infectivity: a genome-wide association study. *PLoS ONE*.

7. Morral AR, McCaffrey D, Iguchi MY (2000): Hardcore drug users claim to be occasional users: drug use frequency underreporting. *Drug and Alcohol Dependence*. 57:193-202.

8. Rhodes W., Layne M., Johnston P., Hozik L. (2000): *What America's Users Spend on Illegal Drugs 1988-1998*. Office of National Drug Control Policy.

9. Association AP (1994): *Diagnostic and statistical manual of mental disorders: DSM-IV, 4th ed.* Washington DC: American Psychiatric Association.

10. Luca D, Ringquist S, Klei L, Lee AB, Gieger C, Wichmann HE, *et al.* (2008): On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *Am J Hum Genet*. 82:453-463.

11. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, *et al.* (2009): Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet*. 41:216-220.

12. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, *et al.* (2009): Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet*. 84:581-593.

13. Gibson J, Griffiths H, De Salvo G, Cole M, Jacob A, Macleod A, *et al.* (2012): Genome-wide association study of primary open angle glaucoma risk and quantitative traits. *Molecular Vision*. 18:1083-1092.

14. Teerlink C, Farnham J, Allen-Brady K, Camp NJ, Thomas A, Leachman S, *et al.* (2012): A unique genome-wide association analysis in extended Utah high-risk pedigrees identifies a novel melanoma risk variant on chromosome arm 10q. *Hum Genet*. 131:77-85.

15. Genovese G, Tonna SJ, Knob AU, Appel GB, Katz A, Bernhardy AJ, *et al.* (2010): A risk allele for focal segmental glomerulosclerosis in African Americans is located within a region containing APOL1 and MYH9. *Kidney Int*. 78:698-704.

16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007): PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559-575.

17. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010): Robust relationship inference in genome-wide association studies. *Bioinformatics*. 26:2867-2873.

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

19. Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, *et al.* (2005): Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug and Alcohol Dependence*. 80:303-312.

20. Nelson EC, Lynskey MT, Heath AC, Wray N, Agrawal A, Shand FL, *et al.* (2014): Association of OPRD1 polymorphisms with heroin dependence in a large case-control series. *Addiction Biology*. 19:111-121.

21. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, Jr., *et al.* (1994): A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol*. 55:149-158.

22. McEvoy M, Smith W, D'Este C, Duke J, Peel R, Schofield P, *et al.* (2010): Cohort profile: The Hunter Community Study. *International Journal of Epidemiology*. 39:1452-1463.

23. Patterson N, Price AL, Reich D (2006): Population structure and eigenanalysis. *PLoS Genet*. 2:e190.

24. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, *et al.* (2010): A map of human genome variation from population-scale sequencing. *Nature*. 467:1061-1073.

25. Johnson EO, Hancock DB, Levy JL, Gaddis NC, Saccone NL, Bierut LJ, *et al.* (2013): Imputation across genotyping arrays for genome-wide association studies: assessment of bias and a correction strategy. *Hum Genet*. 132:509-522.

26. Howie B, Marchini J, Stephens M (2011): Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 1:457-470.

27. Hancock DB, Levy JL, Gaddis NC, Bierut LJ, Saccone NL, Page GP, *et al.* (2012): Assessment of Genotype Imputation Performance using 1000 Genomes in African American Studies. *PloS One*. 7:e50610.

28. Marchini J, Howie B (2010): Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 11:499-511.

29. Jostins L, Morley KI, Barrett JC (2011): Imputation of low-frequency variants using the HapMap3 benefits from large, diverse reference sets. *Eur J Hum Genet*. 19:662-666.

30. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012): Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*. 44:955-959.

31. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007): GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 23:1294-1296.

32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007): PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559-575.

33. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006): Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 38:904-909.

34. Rice JP, Hartz SM, Agrawal A, Almasy L, Bennett S, Breslau N, *et al.* (2012): CHRNB3 is more strongly associated with Fagerstrom test for cigarette dependence-based nicotine dependence than cigarettes per day: phenotype definition changes genome-wide association studies results. *Addiction*. 107:2019-2028.

35. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, *et al.* (2007): Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 16:24-35.

36. Mez JB, Cole JW, Howard TD, Macclellan LR, Stine OC, O'Connell JR, *et al.* (2009): Evaluation of self-reported ethnicity in a case-control population: the stroke prevention in young women study. *BMC Research Notes*. 2:260.

37. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, *et al.* (2007): Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet*. 39:638-644.

38. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, *et al.* (2000): A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol*. 151:346-357.

39. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, *et al.* (2010): Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat Genet*. 42:781-785.

40. Li C, Liu Z, Wang LE, Gershenwald JE, Lee JE, Prieto VG, *et al.* (2008): Haplotype and genotypes of the VDR gene and cutaneous melanoma risk in non-Hispanic whites in Texas: a case-control study. *International Journal of Cancer*. 122:2077-2084.

41. Li C, Zhao H, Hu Z, Liu Z, Wang LE, Gershenwald JE, *et al.* (2008): Genetic variants and haplotypes of the caspase-8 and caspase-10 genes contribute to susceptibility to cutaneous melanoma. *Human Mutation*. 29:1443-1451.