

Genome-Wide Association Study of Opioid Dependence: Multiple Associations Mapped to Calcium and Potassium Pathways

Joel Gelernter, Henry R. Kranzler, Richard Sherva, Ryan Koesterer, Laura Almasy, Hongyu Zhao, and Lindsay A. Farrer

Background: We report a genome-wide association study (GWAS) of two populations, African-American and European-American (AA, EA) for opioid dependence (OD) in three sets of subjects, to identify pathways, genes, and alleles important in OD risk.

Methods: The design employed three phases (on the basis of separate sample collections). Phase 1 included our discovery GWAS dataset consisting of 5697 subjects (58% AA) diagnosed with opioid and/or other substance dependence and control subjects. Subjects were genotyped with the Illumina OmniQuad microarray, yielding 890,000 single nucleotide polymorphisms (SNPs) suitable for analysis. Additional genotypes were imputed with the 1000 Genomes reference panel. Top-ranked findings were further evaluated in Phase 2 by incorporating information from the publicly available Study of Addiction: Genetics and Environment dataset, with GWAS data from 4063 subjects (32% AA). In Phase 3, the most significant SNPs from Phase 2 were genotyped in 2549 independent subjects (32% AA). Analyses were performed with case-control and ordinal trait designs.

Results: Most significant results emerged from the AA subgroup. Genome-wide-significant associations ($p < 5.0 \times 10^{-8}$) were observed with SNPs from multiple loci—*KCNQ2**rs62103177 was most significant after combining results from datasets in every phase of the study. The most compelling results were obtained with genes involved in potassium signaling pathways (e.g., *KCNQ1* and *KCNQ2*). Pathway analysis also implicated genes involved in calcium signaling and long-term potentiation.

Conclusions: This is the first study to identify risk variants for OD with GWAS. Our results strongly implicate risk pathways and provide insights into novel therapeutic and prevention strategies and might biologically bridge OD and other non-substance dependence psychiatric traits where similar pathways have been implicated.

Key Words: Calcium signaling, complex traits, convergence, genome-wide association, opioid dependence, potassium

Opioid dependence (OD) is associated with serious medical, legal, and social problems and co-occurring psychiatric disorders. The cost of OD to society in 2002 was approximately \$181 billion (1). Although genome-wide association study (GWAS) is a method of choice to identify risk genes for complex traits, none has been published for OD, despite an estimated heritability of $>.60$ (2). The strongest GWAS-derived noteworthy and replicable genome-wide significant (GWS) results so far for drug dependence (DD) traits identified a set of loci

From the Department of Psychiatry (JG), Division of Human Genetics, Departments of Genetics and Neurobiology, Yale University School of Medicine; VA Connecticut Healthcare Center (JG), West Haven; Departments Biostatistics and Genetics (HZ), Yale School of Public Health, Yale University, New Haven, CT; University of Pennsylvania Perelman School of Medicine (HRK); Philadelphia VA Medical Center (HRK), Philadelphia, Pennsylvania; Department of Medicine (Biomedical Genetics) (RK, RS, LAF), Boston University School of Medicine; Departments of Neurology, Ophthalmology, Genetics and Genomics (LAF), Epidemiology and Biostatistics, Boston University Schools of Medicine and Public Health, Boston, Massachusetts; and the Department of Genetics (LA), Texas Biomedical Research Institute, San Antonio, Texas.

Address correspondence to Joel Gelernter, M.D., Yale University School of Medicine, Department of Psychiatry, Division of Human Genetics, and VA CT Healthcare Center and Departments of Genetics and Neurobiology, VA CT 116A2, 950 Campbell Avenue, West Haven, CT 06516; E-mail: joel.gelernter@yale.edu.

Received May 31, 2013; revised Jul 29, 2013; accepted Aug 27, 2013.

0006-3223/\$36.00

<http://dx.doi.org/10.1016/j.biopsych.2013.08.034>

mapping to a chromosome 15 nicotinic receptor gene cluster [e.g., Thorgeirsson *et al.* (3)] for nicotine dependence (ND) and related traits; we also reported a GWS association of the *FAM53B* locus to cocaine dependence (CD) (4).

Few other DD GWAS studies have been attempted, and those that have been published are underpowered by modern standards, partly because they used dichotomous traits (i.e., DD diagnoses). Here, we used a relatively large sample and augmented power with an ordinal trait analytic design that allowed us to take into account both the presence or absence of OD and the severity of affection (including the ability to distinguish between subjects with zero and those with one or two symptoms). This increased power by enabling us to use more of the available phenotypic information than standard diagnosis-based analyses. Some of these strategies have been used previously in successful efforts to map ND risk alleles, most notably the use of large clinical samples (3). We further increased our analytic power by including, for some analyses, data from the Study of Addiction: Genetics and Environment (SAGE) sample (5,6), which includes SD trait information. This dataset is available to the scientific community through an application process and will henceforth be referred to as “public domain.”

Our GWAS discovery sample consisted of 2379 European-Americans (EA), including 1383 subjects with OD; and 3318 African-Americans (AA), including 683 subjects with OD. A second phase sample of 4603 EAs and AAs from the SAGE study and a third phase sample including 2549 EAs and AAs ascertained in a manner identical to that of the discovery sample were used to replicate and extend our findings.

Thus, our study took place in three “phases” that differed with respect to samples and genotyping. Phase 1 designates our own GWAS sample. Phase 2 designates the addition of SNP data from

Table 1. Demographic and Diagnostic Information and Subject Characteristics

Recruiting Sites	Phase 1 (SNFs)				Phase 1 (Unrelateds)				Phase 3				Site Total		
	AA		EA		AA		EA		AA		EA		AA	EA	Total
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female			
Yale (APT Foundation)	199	257	141	108	453	370	485	290	223	198	474	477	1700	1975	3675
UConn	174	227	155	161	455	355	451	296	127	93	315	299	1431	1677	3108
MUSC	42	84	52	47	53	109	33	29	21	24	47	47	333	255	588
McLean Hospital	44	36	42	30	10	6	18	11	0	2	2	3	98	106	204
UPenn	9	11	0	0	288	136	20	10	51	64	43	39	559	112	671
SAGE (Phase 2)									643	668	1222	1530	1311	2752	4063
Total Phases 1+2+3													5432	6877	12,309

Phase/OD Status	AA				EA			
	n	Male (%)	Mean Age	Mean Symp. Count	n	Male (%)	Mean Age	Mean Symp. Count
Phase 1/OD Cases	683	62	44.2	6.1	1383	60	37.2	6.5
Phase 1/OD Abusers	49	76	44.8	1.5	57	72	36.9	1.6
Phase 1/Control Subjects	2550	49	40.6	.04	854	55	39.6	.1
Phase 1/Exposed Control Subjects	639	61	42.2	.14	366	64	37.1	.21
Phase 2/OD Cases	105	61	39.9	6	193	60	35.6	5.9
Phase 2/OD Abusers	2	100	40.5	0	5	100	41	0
Phase 2/Control Subjects	1204	48	40	.03	2554	43	38.8	.03
Phase 2/Exposed Control Subjects	158	67	40.8	.2	383	69	35.9	.2
Phase 3/OD Cases	142	67	45.2	5.9	407	60	35.6	6.4
Phase 3/OD Abusers	10	90	44.8	2.2	22	55	36.9	1.1
Phase 3/Control Subjects	601	47	39.9	.02	1255	46	42.3	.02
Phase 3/Exposed Control Subjects	85	78	42.3	.13	137	61	36.7	.16

Recruiting sites: Yale University School of Medicine (APT Foundation), New Haven, CT; University of Connecticut Health Center (UConn), Farmington, CT; the University of Pennsylvania School of Medicine (UPenn), Philadelphia, PA; the Medical University of South Carolina (MUSC), Charleston, SC; and McLean Hospital (Harvard Medical School; Belmont, MA).

AA, African-Americans; EA, European-Americans; OD, opioid dependence; SAGE, Study of Addiction: Genetics and Environment; SNFs, small nuclear families; Symp, symptom.

SAGE (which used a very different recruitment strategy but similarly ascertained subjects and was genotyped on a different microarray) combined with our sample by meta-analysis; this is the core of the GWAS discovery and replication strategy. Phase 3 designates our own smaller replication sample, where individual SNPs rather than GWAS arrays were genotyped. With these strategies, we identified genetic variants that increase risk for OD and related heritable traits.

Methods and Materials

Subjects and Diagnostic Procedures

The (Phase 1) GWAS discovery sample included 5697 subjects. A second identically ascertained sample comprising 2549 subjects was used for replication (Phase 3). All of these subjects were recruited for studies of the genetics of drug (opioid or cocaine) or alcohol dependence (AD). The sample consisted of small nuclear families (SNFs) originally collected for linkage studies and unrelated individuals. Subjects were recruited at five eastern US sites (Table 1). Our previous OD linkage study (7) included a subset of the small nuclear families included in this study. Subjects gave written informed consent as approved by the institutional review board at each site, and certificates of confidentiality were obtained from the National Institute on Drug Abuse and the National Institute on Alcohol Abuse and Alcoholism.

All subjects were interviewed with an electronic version of the Semi-Structured Assessment for Drug Dependence and Alcoholism (8,9) to derive DSM-IV diagnoses (10) of lifetime OD and other major psychiatric traits. The test-retest ($\kappa = .94$) and interrater ($\kappa = .91$) reliability of the OD diagnosis was excellent (9).

Genotyping and Quality Control

Samples for Phase 1 were genotyped on the Illumina HumanOmni1-Quad v1.0 microarray containing 988,306 autosomal SNPs, at the Center for Inherited Disease Research and the Yale Center for Genome Analysis. Genotypes were called with GenomeStudio software V2011.1 and genotyping module V1.8.4 (Illumina, San Diego, California).

Follow-up genotyping (Phase 3 sample) was performed with a custom Illumina GoldenGate Genotyping Universal-32, 1536-plex microarray. Most SNPs included in the custom array were selected for studies of other phenotypes. Additional SNPs were genotyped individually with the TaqMan method (11).

On the microarray, 44,644 SNPs and 135 individuals with call rates < 98% were excluded, and 62,076 additional SNPs were removed due to minor allele frequencies (MAF) < 1%. After data cleaning and quality control, 5697 individuals and 889,659 SNPs remained for imputation. We identified several instances where identical DNA marker profiles were linked to two different interview forms. When demographic information (sex, date of birth, number of reported children) was consistent across interviews, one sample was randomly removed from analysis; when demographic information was inconsistent, both were removed. Genetic relationships were examined in the family-based sample by calculating pairwise identity by descent (IBD) proportion estimates with PLINK (12). Pairs of individuals whose IBD proportions did not match their reported genetic relationship were assigned to two different families, and pairs of individuals who shared more than 25% of their alleles IBD were assigned to the same family. Self-reported male subjects with X chromosome heterozygosity > 20% and self-reported female subjects with

heterozygosity < 20% were excluded, unless their true identity could be determined.

To verify and correct the misclassification of self-reported race, we compared GWAS data from all subjects with genotypes from the HapMap 3 reference CEU, YRI, and CHB populations. Principal components (PC) analysis was conducted in the entire GWAS sample with Eigensoft (13,14) and 145,472 SNPs that were common to the GWAS dataset and HapMap panel (after pruning the GWAS SNPs for linkage disequilibrium [r^2] > 80%) to characterize the underlying genetic architecture of the samples. The first PC score distinguished AAs and EAs; these groups were subsequently analyzed separately. We then conducted PC analyses within the two groups and the first three PCs were used in all subsequent analyses to correct for residual population stratification.

Public Domain GWAS Sample: SAGE

In the Phase 2 analyses described in the following text, we included publically available GWAS data from SAGE (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1) (15). The SAGE dataset contained 1311 AA and 2750 EA unrelated individuals (Table 1). The SAGE study includes individuals from the COGA (Collaborative Study on the Genetics of Alcoholism) (5), FSCD (Family Study of Cocaine Dependence) (6), and COGEND (Collaborative Genetic Study of Nicotine Dependence) (15) studies. The COGA sample is a set of unrelated individuals recruited in Indiana, New York, St. Louis, Connecticut, Iowa, and San Diego selected for genotyping from a larger set of 8000 subjects. Cases met criteria for DSM-IV AD. The FSCD study contained cases and control subjects from the greater St. Louis metropolitan area. All cases met criteria for DSM-IV AD, and most also met criteria for DSM-IV CD. Control subjects were from the same communities and had consumed alcohol but had no lifetime history of substance dependence. A subgroup of FSCD subjects was not alcohol dependent but had a lifetime diagnosis of DSM-IV cannabis dependence or dependence on another illicit drug. The COGEND subjects were recruited in Missouri and Michigan. Cases met criteria for DSM-IV alcohol and/or ND. The SAGE samples were genotyped on the Illumina Human 1M array containing 1,069,796 total SNPs. Control subjects were selected from the nondependent population and did not meet criteria for AD, ND, or illicit DD.

Genotypes were imputed with IMPUTE2 (16) with the genotyped SNPs and the 1000 genomes reference panel released in June 2011 (<http://www.1000genomes.org/>), which contains phased haplotypes for 1094 individuals of various ancestries, including 379 samples of European descent (CEU, FIN, GBR, IBS, and TCI), 286 of Asian descent (CHB, JPT, and CHS), 181 admixed American samples (PUR, CLM, and MXL), and 246 samples of African descent (ASW, LWK, YRI) (17). The EA and AA samples were imputed separately. Genotyped SNPs with Hardy-Weinberg equilibrium p values < 10^{-5} were set to missing, and imputed genotypes were used instead. We retained 18,564,419 SNPs with derived information content > .8 in at least one of the population groups.

Analytical Models

We employed a model with the imputed minor allele dosage as the dependent variable used DSM-IV symptom count (Symptom_{adj}) for OD and each of three other major SD diagnoses (CD, AD, and ND) as ordinal predictors (with adjustment for sex, age, and the first three PCs of ancestry [Figure S1 in Supplement 1]). The mutual adjustment of OD for measures of

dependence on other substances facilitated the identification of genetic risk factors unique to each substance and limited confounding due to comorbid dependencies. All individuals contributed to this analysis, including those meeting DSM-IV criteria for OD (who met 3–7 criteria) and those having no symptoms of OD. The ordinal trait model has greater power to detect genetic associations than a univariate model on the basis of disease status because of greater information content and improved specificity of the dependence measure. The beta coefficient and p value for the OD symptom count (adjusted for the symptom counts for CD, AD, and ND) were used to assess the magnitude and significance of the association, respectively. Ordinal trait data were derived for all samples included in the study.

Case-control status was the outcome in models that included as control subjects only individuals who had used opioids at least once without becoming dependent. This accounts for the fact that subjects who were never exposed might be better classified as phenotype “unknown” than as “unaffected.”

Association tests in the Phase 1 and Phase 2 GWAS datasets were performed with linear or logistic association models embedded in generalized estimating equations to correct for correlations among related individuals (18). The Phase 3 sample containing unrelated individuals only was evaluated with linear and logistic models. All models were adjusted for age, sex, and the first three PCs of ancestry. The data were analyzed separately within population groups, and the results were combined by meta-analysis with the inverse variance method implemented in the computer program METAL (19). In Phase 2, Phase 1 SNPs with p values $< 10^{-4}$ in either population were tested for association in the SAGE dataset in the specific population group(s) and statistical model(s) yielding the specified result in Phase 1. Tests were restricted in this manner to minimize their number and because it is unlikely that an SNP not meeting the cutoff for follow-up would attain a noteworthy significance level when combining results from the discovery and replication samples. Phase 1 and 2 results were combined within population groups by meta-analysis. In Phase 3, we selected for replication 76 SNPs (Sympcount_{adj}, $n = 44$, and case-control, $n = 28$, including 58 unique SNPs) with $p < 10^{-4}$ in either population in the Phase 2 analyses.

Multiple Testing Considerations

A p value of 5.0×10^{-8} was used as a threshold for GWS. Results were not adjusted for testing in two populations, because we tested three distinct a priori hypotheses: 1) SNPs are associated with OD and related traits in AAs; 2) SNPs are associated with OD and related traits in EAs; and 3) associations are evident

with the same SNPs in AAs and EAs (in meta-analysis). Results were not adjusted for testing two phenotypes, because the rank order correlation between the symptom count and case-control variables is 1.0 (i.e., no affected subject can have a lower symptom count than a control), and the point-biserial correlation is .95.

Pathway Analysis

Meta-analyzed GWAS results (separately by population) from the discovery and SAGE datasets were used to identify biological pathways related to OD. First, the number of independent SNP association tests for each gene in the genome was computed according to the method of Li and Ji (20). Next, the smallest p value for an individual SNP within each gene was multiplied by the number of independent tests in that gene to create a list of genes significantly associated with OD after correcting for the number of tests within that gene ($p_{adj} < .05$). The significant genes were evaluated by pathway analysis, performed with the Ingenuity Pathway Analysis software suite (<http://www.ingenuity.com>) to identify an overrepresentation of selected genes within canonical pathways that were defined with information culled from multiple sources (Kyoto Encyclopedia of Genes and Genomes, interactome studies, manual curation, etc.). A Fisher's exact p value for each pathway indicated whether the pathway contained more significantly associated genes than expected by chance. To validate the findings, we uploaded the same gene list into the Database for Annotation, Visualization and Integrated Discovery (<http://david.abcc.ncifcrf.gov/>) and performed a gene enrichment analysis with its functional annotation tools.

Results

As noted, three independent datasets were employed in these OD-phenotype analyses, in three phases. The EAs and AAs were analyzed separately in each phase. In Phase 1, GWAS was conducted in a discovery dataset containing 5697 subjects. In Phase 2, the first replication phase, SNPs with p values $< 10^{-4}$ in either or both populations combined were tested for association in the SAGE dataset, including 4063 subjects. In Phase 3, the second replication phase, SNPs with p values $< 10^{-4}$ in either population in the Phase 2 analyses were evaluated in a set of 2549 unrelated subjects ascertained in the same manner as the Phase 1 sample. Meta-analyses were conducted in a combined sample of 12,309 subjects. Analyses of symptom count accounted for comorbidity with other SD symptoms (sympcount_{adj}).

In both Phase 1 analyses, we identified population-specific variants and variants associated with risk in both EAs and AAs and (Figures S2 and S3 in Supplement 1) including GWS associations

Table 2. Genome-Wide Significant Association Results for Measures of OD African-American population

Chr	BP	SNP	SNP Type	Gene	Phase 1		Phase 2		Phase 1+2	Phase 3		Phase 1+2+3
					MAF	p	MAF	p	p	MAF	p	p
4	41056716	rs114070671 ^a	intron	APBB2	.08	4.58E-06	.08	8.27E-05	2.26E-09	.09	.55	1.04E-07
4	41056839	rs115368721	intron	APBB2	.08	3.74E-06	.08	1.16E-04	1.92E-08		not tested	
11	12454744	rs73411566	intron	PARVA	.03	5.48E-05	.03	3.06E-05	1.75E-08		not tested	
11	17796111	rs60349741	UTR_3	KCNC1	.06	1.33E-08	.07	7.79E-02	9.26E-09	.05	.45	4.81E-07
18	77624479	rs62103177	intron	KCNG2	.06	8.00E-05	.06	1.10E-07	6.68E-10	.06	.15	3.60E-10

Genome-wide significant association results for measures of opioid dependence (OD)—adjusted symptom count model. Imputation qualities ranged from .91 to .99 (Table S3 in Supplement 1).

BP, base pair; Chr, chromosome; MAF, minor allele frequencies; SNP, single nucleotide polymorphism.

^aLinkage disequilibrium proxy for rs115368721.

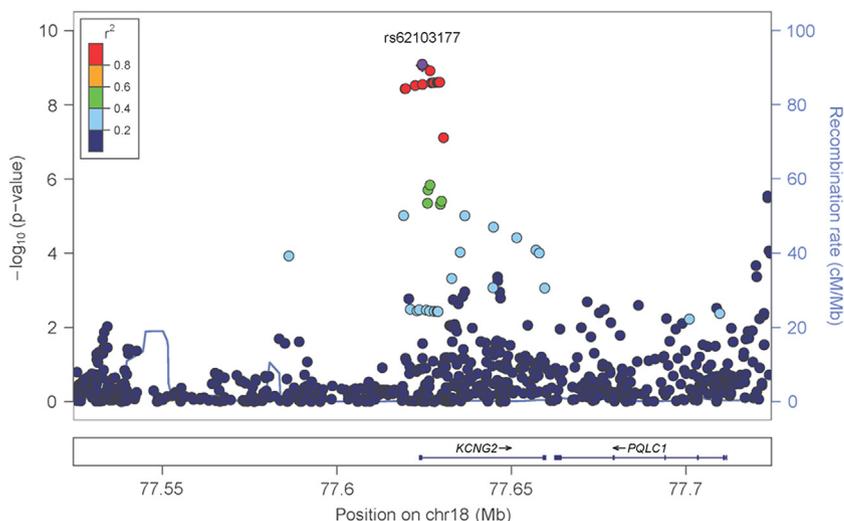


Figure 1. Regional association (Manhattan) plot showing Phase 1 + 2 meta-analyzed results for association of $\text{Sympcount}_{\text{adj}}$ with single nucleotide polymorphisms (SNPs) mapped to the *KCNG2* region on chromosome 18 in the African-American population. The SNPs are color coded according to r^2 with the most significant SNP shown in purple. The SNP with the strongest evidence for association is shown twice: once with the Phase 1 + 2 meta-analysis p value (purple with “crosshairs”); and once after the inclusion of the Phase 3 samples (purple without “crosshairs”) (they are nearly superimposed in this figure). The light blue line and right Y-axis show the observed recombination rate in the HapMap YRI samples.

in AAs with *PITPNM3* (rs9913974, $p = 8.4 \times 10^{-9}$) for the case-control model and *KCNC1* (rs60349741, $p = 1.3 \times 10^{-8}$) and *HHLA2* (rs73204138, $p = 3.7 \times 10^{-8}$) for the $\text{Sympcount}_{\text{adj}}$ model. The p values for these findings are negligibly inflated (Figures S4 and S5 in Supplement 1), and the SNPs were well-imputed (Table S1 in Supplement 1). We also identified a possible association with a *DISC1* polymorphism (rs2738888, AA $p = 8.47 \times 10^{-4}$; EA $p = 1.74 \times 10^{-3}$; meta $p = 5.22 \times 10^{-6}$ [discussed in following text]).

In Phase 2, we repeated the primary OD analyses for SNPs with $p < 10^{-4}$ (403 SNPs for the case-control model, 1522 SNPs for the $\text{Sympcount}_{\text{adj}}$ model in AAs) in the independent SAGE sample. Four distinct regions contained GWS SNP $\text{sympcount}_{\text{adj}}$ p values in AAs after meta-analysis of the results from Phases 1 and 2 (Table 2). These included the regions containing the potassium voltage-gated channel genes *KCNG2* (rs62103177, $p = 6.7 \times 10^{-10}$) (Figure 1) and *KCNC1* (rs60349741, $p = 7.5 \times 10^{-9}$) (Figure S6 in Supplement 1), *APBB2* (rs115368721, $p = 1.8 \times 10^{-9}$) (Figure S7 in Supplement 1), and *PARVA* (rs73411566, $p = 1.8 \times 10^{-8}$) (Figure S8 in Supplement 1). Several GWS results from Phase 1 did not remain so in Phase 2, including SNPs mapped to *HHLA2* and *PITPNM3*.

Although no GWS association signals identified in the meta-analysis of the results from Phases 1 + 2 were significantly associated with the corresponding trait in the much smaller Phase 3 dataset at $p < .05$, three of the results improved or remained significant after meta-analysis that included the Phase 3 data (Table 2). The associations of rs62103177 in *KCNG2* with $\text{Sympcount}_{\text{adj}}$ improved to $p = 3.6 \times 10^{-10}$ (Figure 1). Two of the GWS SNPs (rs115368721 in *APBB2*, rs73411566 in *PARVA*) in Phase 2 were imputed and assays could not be designed to genotype them in Phase 3. Linkage disequilibrium-based proxies were used for *APBB2* SNPs, but no suitable proxies (i.e., $r^2 > .4$) were available for the *PARVA* SNP.

Pathway Analysis

Pathway analysis was performed including genes with at least one SNP with a gene-based corrected p value $< .05$. Separate analyses were performed with gene sets obtained from the case-control (62 genes) and $\text{Sympcount}_{\text{adj}}$ (198 genes) models. The most remarkable pathways were identified from the case-control model. The first and third most significant pathways resulting

from this analysis were calcium signaling ($p = .002$, false discovery rate [FDR] = .15) (Figure 2) and synaptic long-term potentiation (LTP) ($p = .004$, FDR = .17) (Figure 3), respectively (the second-most significant canonical pathway [$p = .0024$] captured genes that influence cardiac hypertrophy). The Ca^{2+} signaling pathway was identified by significant associations in genes encoding calcineurin A (*PPP3CA*), calcium/calmodulin-dependent protein kinase II beta (*CAMK2B*), histone deacetyltransferase 9 (*HDAC9*), and cyclic adenosine monophosphate (cAMP) responsive element binding protein 5 (*CREB5*). The Ca^{2+} signaling is also critical to LTP; thus, this pathway emerged as significant in part due to associations with *PPP3CA*, *CAMK2B*, and *CREB5* (Table 3). Reanalysis with the Database for Annotation, Visualization and Integrated Discovery software revealed that the top gene ontology functional category was “calcium ion binding” ($p = .0089$), and the top Kyoto Encyclopedia of Genes and Genomes pathway was “LTP” ($p = .034$). These pathways (complete list, Table S2 in Supplement 2) are identical to the pathways identified by Ingenuity Pathway Analysis.

Discussion

We report here results for a GWAS for OD in two different population groups in the United States. We made use of GWAS and replication samples collected by us; and available data from the SAGE project, which were collected with a similar psychiatric interview, were used for replication. Several of the top-ranked genes encode proteins that participate in potassium and calcium signaling pathways. Although calcium signaling genes have been studied in addiction biology, they were not previously considered key genetic candidates. These findings therefore add substantially to our knowledge of the biology of OD. Our most compelling results were obtained in the AA population.

The loci *KCNC1* and *KCNG2*, containing some of the most significantly associated SNPs, encode potassium voltage-gated channel subunits. Another locus, *PITPNM3*, encodes a protein that is involved in phosphatidylinositol transport but that also binds calcium. Variants located approximately 196 kb upstream from *PPP3CA* were associated with OD in the case-control model (although not GWS) in both populations (rs6419156, $p_{\text{AA}} = 1.4 \times 10^{-6}$, $p_{\text{EA}} = 5.6 \times 10^{-7}$); the effect directions were opposite, suggesting that population-specific causal variants or a single

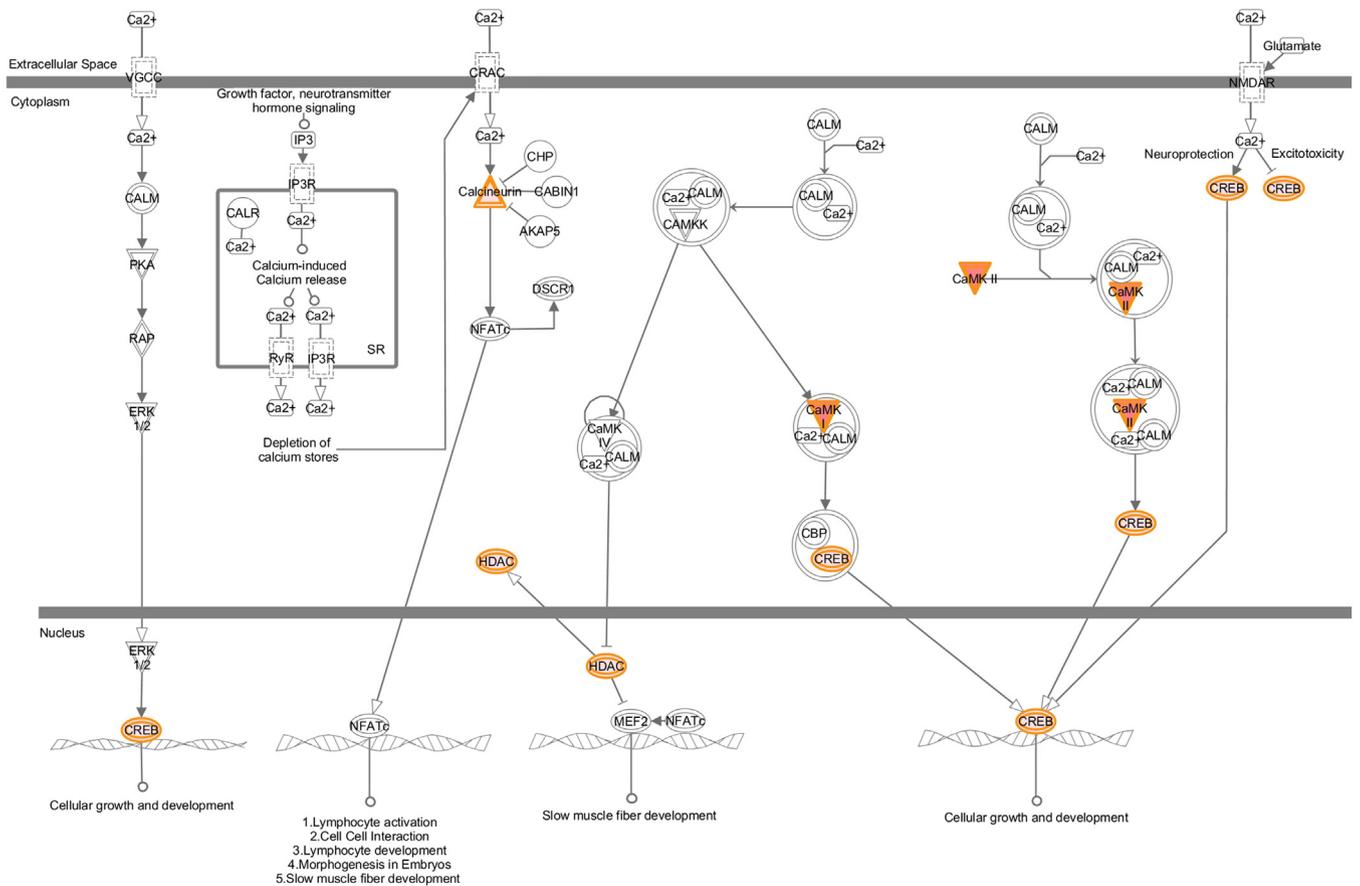


Figure 2. The role of genes identified in genome-wide association study of Phase 1 + 2 African-American subjects with a case-control model in the canonical pathway “calcium signaling.” The gene or gene families with a significant association result are shown in orange. The calcineurin family was identified by an association in *PPP3CA*, the cyclic adenosine monophosphate response element binding protein (CREB) family by *CREB5*, the histone deacetyltransferase 9 (HDAC) family by *HDAC9*, and the calmodulin-dependent protein kinase type 2 (CAMK2) family by *CAMK2B* (Table 3). For a full description of the functional relationships represented by the various lines and arrows, see: <http://www.bioreprod.org/content/suppl/2010/09/29/bioreprod.110.085910.DC1/bioreprod.110.085910-3.pdf>. AKAP5, A-kinase anchor protein 5; Ca²⁺, calcium; CABIN1, calcineurin binding protein 1; CALM, calmodulin; CBP, CREB binding protein; CHP, calcium binding protein P22; CRAC, capacitance regulated activation channels; *DSCR1*, Down syndrome candidate region 1; ERK1/2, extracellular signal-regulated kinase 1 and 2; IP3R, inositol triphosphate receptors; MEF, myocyte enhancer factors; NFATc, nuclear factor of activated T cell, cytoplasmic; NMDAR, N-methyl-D-aspartate receptor; PKA, protein kinase A; RAP, rapamycin; RyR, ryanodine receptor; VGCC, voltage-gated calcium channel.

causal variant occurring on distinct haplotypes might be responsible. The *PPP3CA* locus encodes a calcium-dependent, calmodulin-stimulated protein phosphatase involved in calcium signaling [a trinucleotide repeat at this locus was reported to be nominally associated with SD in AAs but not EAs (21)].

Calcium signaling and LTP (which relies on calcium signaling for neurotransmitter release) were two of the best-supported pathways identified in pathway analysis (although these results did not meet FDR-defined significance after correction for the number of pathways evaluated). Thus, potassium and calcium transport and signaling mechanisms seem to play essential roles in OD risk.

Possibly relevant to these findings, *CACNA1C* (which encodes a voltage-dependent calcium channel) is one of the best-supported GWAS-identified risk genes for bipolar affective disorder (22) and schizophrenia (23). Furthermore, a large mega-analysis of psychiatric traits also landed on several calcium-system genes as contributing to risk for a set of psychiatric illnesses: two of the four GWS SNPs mapped to voltage-gated calcium channel subunit genes, and calcium signaling pathways were identified as important for the five disorders studied (24). Potassium-calcium signaling might couple neuronal signaling to vasodilation in the brain (25), and

opioids can regulate calcium conductance via increasing potassium conductance in μ -opioid receptors (26). One of the genes in the calcium signaling pathway that is strongly associated with OD risk, *CAMK2B*, modulates activation of ionotropic (including alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) glutamate receptors (27). Alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid-mediated glutamatergic signaling in the hippocampus plays an important role in context-dependent sensitization to morphine. Behavioral sensitization in animal models is of interest as a neural basis of addiction (28,29). LTP (which underlies learning and memory) was also implicated directly in pathway analysis. These pathways might interact to produce biologically important effects both on behavior after an initial exposure to opioids and on impaired control over use after chronic exposure (i.e., relapse risk, a key issue in the natural history and treatment of OD).

There were additional GWS associations (albeit not-yet-replicated) of biological interest. Encoding alpha parvin, *PARVA*—which is reported to play roles in cell adhesion and cytoskeletal organization—was associated in the GWAS for OD (Phase 2 meta $p = 1.75 \times 10^{-8}$). The *APBB2* (Phase 2 meta $p = 2.26 \times 10^{-9}$) encodes a member of a family of proteins that bind the amyloid precursor

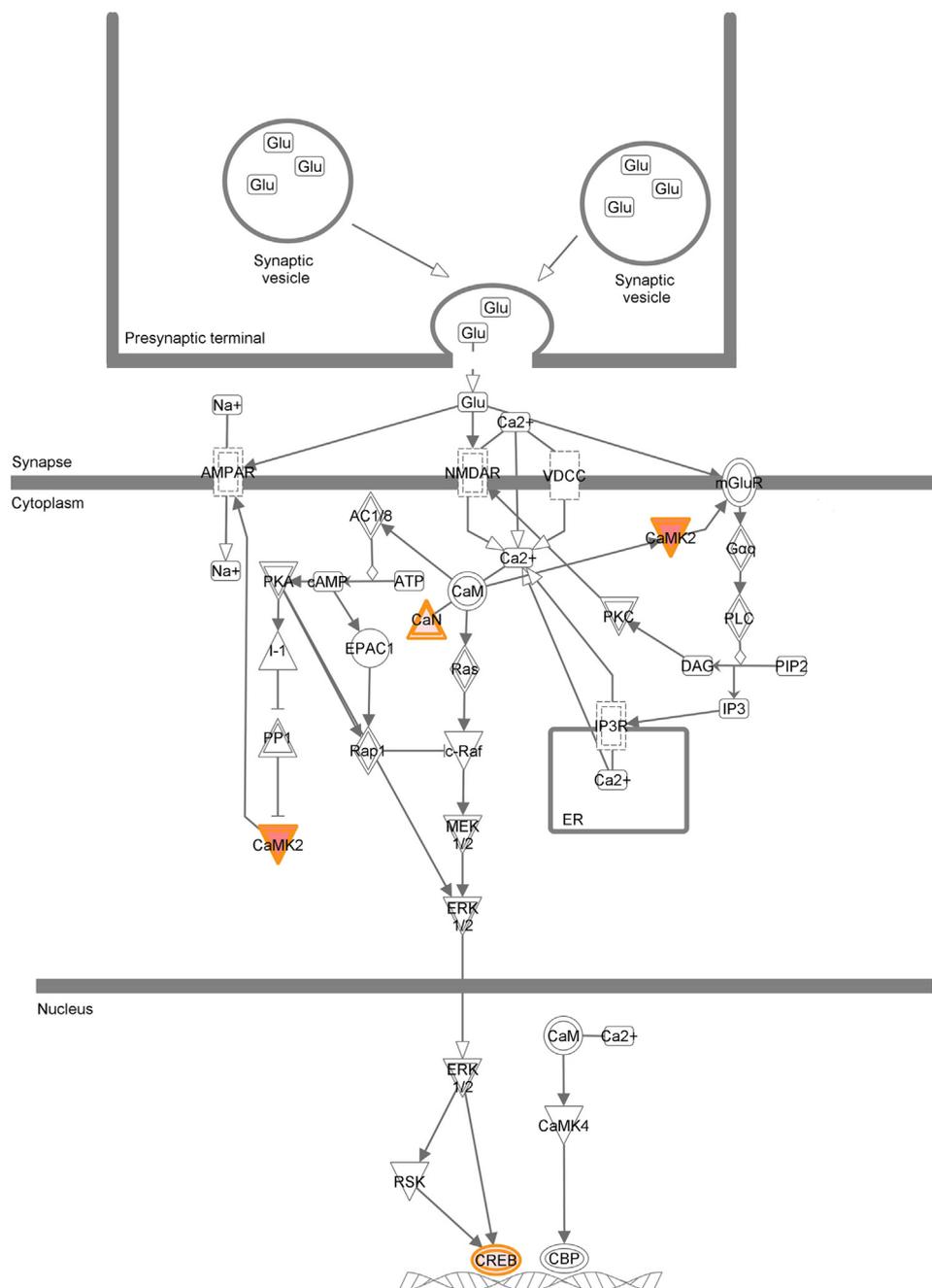


Figure 3. The role of genes identified in genome-wide association study of Phase 1 + 2 AA subjects with a case-control model in the canonical pathway “synaptic long-term potentiation.” The gene or gene families with a significant association result are shown in orange. The calcineurin family was identified by an association in PPP3CA, the CREB family by *CREB5*, and the calcium/calmodulin-dependent protein kinase II beta (*CAMK2*) family by *CAMK2B* (Table 3). For a full description of the functional relationships represented by the various lines and arrows, see: <http://www.bioreprod.org/content/suppl/2010/09/29/bioreprod.110.085910.DC1/bioreprod.110.085910-3.pdf>. AC1/8, adenylyl cyclase 1 and 8; AMPAR, alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor; ATP, adenosine triphosphate; DAG, diacyl-glycerol; EPAC, exchange protein directly activated by Camp; ER, endoplasmic reticulum; Glu, glutamate; mGluR, metabotropic glutamate receptor; MEK1/2, mitogen-activated protein kinase kinase; Na⁺, sodium ion; PIP2, phosphatidylinositol 4,5-bisphosphate; PKC, protein kinase C; PLC, phospholipase C; other abbreviations as in Figure 2.

protein that might be important for signal transduction and is associated with Alzheimer’s disease risk (30). A closely related gene, *APBB1*, was shown previously to be associated with ND (31).

An interesting possible OD risk gene identified in Phase 1 was *DISC1*, which (among other functions) regulates development of the cerebral cortex. The *DISC1* gene was originally identified as a schizophrenia risk gene in a cytogenetic study (32), and considerable

evidence since has tied it to cognitive phenotypes in human (33) and animal (34) models. We observed nominally significant associations with *DISC1* in both EAs and AAs (meta $p = 5.2 \times 10^{-6}$). Subsequent to this common-variant association finding, we used deep sequencing at this locus to identify a set of rare variants in *DISC1* that contribute to risk for OD (35). Thus, this study ties OD to other cognitive phenotypes via a common gene and, because

Table 3. Pathway SNPs

Chr	SNP	Gene	<i>p</i>	BP	Total SNPs	Independent SNPs	Adjusted <i>p</i>
7	rs2075069	CAMK2B	1.27E-06	44284564	61	20	2.55E-05
4	4-102128801	PPP3CA	3.15E-04	1.02E+08	67	18	5.66E-03
7	rs9639575	CREB5	1.19E-04	28417214	135	52	6.16E-03
7	7-18189274	HDAC9	2.40E-05	18189274	219	73	1.75E-03

Abbreviations as in Table 2.

schizophrenia was one of the five disorders identified as associated with calcium pathways (24), also through gene pathways. Previously, we reported that another schizophrenia-associated locus, *NRG1*, is associated with cannabis dependence in AAs (36).

Strengths of this study are a relatively large sample size, including SAGE data to increase power. All subjects were assessed with comprehensive, standardized instruments. Our primary sample has unusually good representation of AA subjects. Most of the strongest findings emerged from this understudied population; some of these were supported by findings from the EAs. We employed several novel analytic strategies, including models defining OD symptoms as a quantitative trait adjusted for addiction to other substances. Such approaches minimize concern about the misclassification of control subjects, because SD requires exposure to the drug. An overall picture emerges of genes and pathways involved in regulating calcium and potassium signaling as major contributors to OD risk.

Our study has several limitations. The GWS and other top-ranked findings were identified with imputed SNPs; however, empirically, our results show that the top-ranked SNPs in our study were imputed with excellent quality. Although our strongest results were with imputed SNPs, they are not rare (two had MAF = .08; two had MAF = .06; and one had MAF = .03.) The primary concern about rare (MAF <1%) SNPs is the greater rate of false positive results possible, due to large effect sizes (i.e., greatly increased risk) derived from small allele frequency differences between cases and control subjects or genotyping errors. However, studies examining this issue have concluded that nominally significant results occur less frequently than expected, even for low MAF SNPs, provided there is no genotyping or imputation quality bias between cases and control subjects (37,38). Moreover, the concern about false positives is substantially less when the results are replicated or supported by evidence from multiple samples, which is the case for the associations that we reported. Also, although our sample size is reasonable, experience has shown that many true associations are detected only in samples or meta-analyses that are much larger. The relatively small size of the sample used in Phase 3 for the primary trait analyses and as replication for the OD subtype findings is in our view the most likely cause of few replication observations in that sample. Specifically, there were proportionally fewer AAs available for Phase 3 ($n = 805$, approximately 24% as large as the discovery sample), which is the part of the sample in which most positive results were initially observed. Ideally, the replication sample would be larger than the discovery sample. It is also noteworthy that there was little evidence of association in EAs for any of our top findings in AAs; however, all but one of the SNPs reported in Table 2 were monomorphic in EAs. Finally, our findings are not adjusted for testing association in two populations and with two trait models (case-control and ordinal). Bonferroni correction is too conservative, given the high correlation between the traits and distinct hypotheses for EAs and AAs, populations that often have distinctive risk loci and/or distinctive risk alleles at the same

locus. Future studies in large independent populations are necessary to address some of these concerns.

In conclusion, we identified numerous novel GWS associations with OD. We found compelling evidence for association in several brain systems, most consistently for calcium and potassium signaling pathways, which should provide the impetus for further research on these brain systems in OD. Furthermore, these findings are largely consonant with a set of SNP and pathway-based findings that are emerging for a broader range of neuropsychiatric traits and, as such, might bridge OD and other neuropsychiatric phenotypes.

This study was supported by National Institutes of Health Grants RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535 and the VA Connecticut and Philadelphia VA Mental Illness Research, Education and Clinical Centers.

We appreciate the work in recruitment and assessment provided at Yale University School of Medicine and the APT Foundation by James Poling, Ph.D.; at McLean Hospital by Roger Weiss, M.D.; at the Medical University of South Carolina by Kathleen Brady, M.D., Ph.D., and Raymond Anton, M.D.; and at the University of Pennsylvania by David Oslin, M.D. Genotyping services for a part of our genome-wide association study were provided by the Center for Inherited Disease Research and Yale University (Center for Genome Analysis) CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University (contract number N01-HG-65403). We are grateful to Ann Marie Lacobelle, Michelle Cucinelli, Christa Robinson, and Greg Dalton-Kay for their excellent technical assistance; to the Semi-Structured Assessment for Drug Dependence and Alcoholism interviewers, led by Yari Nuñez and Michelle Slivinsky, who devoted substantial time and effort to phenotype the study sample; and to John Farrell and Alexan Mardigan for database management assistance.

Dr. Kranzler has served as a consultant or advisory board member for Alkermes, Lilly, Lundbeck, Pfizer, and Roche. He is also a member of the American Society of Clinical Psychopharmacology Alcohol Clinical Trials Initiative, which is supported by Lilly, Lundbeck, Abbott, and Pfizer. The other authors declare no biomedical financial interests or potential conflicts of interest.

The publicly available datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1 through dbGaP accession number phs000092.v1.p.

Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative (GEI) (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446).

Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEN; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract “High throughput genotyping for studying the genetic contributions to human disease” (HHSN268200782096C).

Supplementary material cited in this article is available online at <http://dx.doi.org/10.1016/j.biopsych.2013.08.034>.

- Office of National Drug Control Policy (2004): *The Economic Costs of Drug Abuse in the United States, 1992–2002*. Washington, DC: Executive Office of the President (Publication No. 207303).
- Goldman D, Oroszi G, Ducci F (2005): The genetics of addictions: Uncovering the genes. *Nat Rev Genet* 6:521–532.
- Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, *et al.* (2008): A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452:638–642.
- Gelernter J, Sherva R, Koesterer R, Almsy L, Zhao H, Kranzler HR, Farrer LA (2013): Genomewide association study of cocaine dependence and related traits: *FAM53B* identified as a risk gene [published online ahead of print August 20]. *Mol Psychiatry*. doi:10.1038/mp.2013.99.
- Edenberg HJ (2002): The collaborative study on the genetics of alcoholism: An update. *Alcohol Res Health* 26:214–218.
- Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB (2008): Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend* 95:14–22.
- Gelernter J, Panhuysen C, Wilcox M, Hesselbrock V, Rounsaville B, Poling J, *et al.* (2006): Genomewide linkage scan for opioid dependence and related traits. *Am J Hum Genet* 78:759–769.
- Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, *et al.* (2005): Genomewide linkage scan for cocaine dependence and related traits: Linkages for a cocaine-related trait and cocaine-induced paranoia. *Am J Med Genet Neuropsych Genet* 136:45–52.
- Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastris A, *et al.* (2005): Diagnostic Reliability of the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug Alcohol Depend* 80:303–312.
- American Psychiatric Association (1994): *Diagnostic and Statistical Manual of Mental Disorders, 4th ed.* Washington, DC: American Psychiatric Press.
- Holland PM, Abramson RD, Watson R, Gelfand DH (1991): Detection of specific polymerase chain reaction product by utilizing the 5′→3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci U S A* 88:7276–7280.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007): PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–575.
- 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.
- Patterson N, Price AL, Reich D (2006): Population structure and eigenanalysis. *PLoS Genet* 2:e190.
- 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.
- Howie BN, Donnelly P, Marchini J (2009): A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5:e1000529.
- 1000 Genomes Project Consortium (2010): A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.
- Zeger SL, Liang KY (1986): Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42:121–130.
- Willer CJ, Li Y, Abecasis GR (2010): METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190–2191.
- Li J, Ji L (2005): Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 95:221–227.
- Chiocco MJ, Zhu X, Walther D, Pletnikova O, Troncoso JC, Uhl GR, *et al.* (2010): Fine mapping of calcineurin (PPP3CA) gene reveals novel alternative splicing patterns, association of 5′UTR trinucleotide repeat with addiction vulnerability, and differential isoform expression in Alzheimer’s disease. *Subst Use Misuse* 45:1809–1826.
- Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, *et al.* (2008): Whole-genome association study of bipolar disorder. *Mol Psychiatry* 13:558–569.
- Nyegaard M, Demontis D, Foldager L, Hedemand A, Flint TJ, Sørensen KM, *et al.* (2010): CACNA1C (rs1006737) is associated with schizophrenia. *Mol Psychiatry* 15:119–121.
- Cross-Disorder Group of the Psychiatric Genomics Consortium (2013): Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* 381:1371–1379.
- Filosa JA, Bonev AD, Straub SV, Meredith AL, Wilkerson MK, Aldrich RW, *et al.* (2006): Local potassium signaling couples neuronal activity to vasodilation in the brain. *Nat Neurosci* 9:1397–1403.
- North RA, Williams JT (1983): Opiate activation of potassium conductance inhibits calcium action potentials in rat locus coeruleus neurons. *Br J Pharmacol* 80:225–228.
- Jiao Y, Jalan-Sakrikar N, Robison AJ, Baucum AJ 2nd, Bass MA, Colbran RJ (2010): Characterization of a central Ca²⁺/calmodulin-dependent protein kinase II α /beta binding domain in densin that selectively modulates glutamate receptor subunit phosphorylation. *J Biol Chem* 286:24806–24818.
- Steketee JD, Kalivas PW (2011): Drug wanting: Behavioral sensitization and relapse to drug-seeking behavior. *Pharmacol Rev* 63:A–R.
- Xia Y, Portugal GS, Fakra AK, Melyan Z, Neve R, Lee HT, *et al.* (2011): Hippocampal GluA1-containing AMPA receptors mediate context-dependent sensitization to morphine. *J Neurosci* 31:16279–16291.
- Li Y, Hollingworth P, Moore P, Foy C, Archer N, Powell J, *et al.* (2005): Genetic association of the APP binding protein 2 gene (APBB2) with late onset Alzheimer disease. *Hum Mutat* 25:270–277.
- Chen GB, Payne TJ, Lou XY, Ma JZ, Zhu J, Li MD (2008): Association of amyloid precursor protein-binding protein, family B, member 1 with nicotine dependence in African and European American smokers. *Hum Genet* 124:393–398.
- St. Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, *et al.* (1990): Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* 336:13–16.
- Thomson PA, Parla JS, McRae AF, Kramer M, Ramakrishnan K, Yao J, *et al.* (2013): 708 Common and 2010 rare DISC1 locus variants identified in 1542 subjects: Analysis for association with psychiatric disorder and cognitive traits [published online ahead of print June 4]. *Mol Psychiatry*. doi:10.1038/mp.2013.68.
- Kvajo M, McKellar H, Arguello PA, Drew LJ, Moore H, MacDermott AB, *et al.* (2008): A mutation in mouse Disc1 that models a schizophrenia risk allele leads to specific alterations in neuronal architecture and cognition. *Proc Natl Acad Sci U S A* 105:7076–7081.
- Xie P, Kranzler HR, Krystal JH, Farrer LA, Zhao H, Gelernter J (2013): Deep resequencing of 17 glutamate system genes identifies rare variants in DISC1 and GRIN2B affecting risk of opioid dependence [published online ahead of print July 16]. *Addict Biol*. doi:10.1111/adb.12072.
- Han S, Yang B-Z, Kranzler HR, Oslin D, Anton R, Farrer LA, Gelernter J (2012): Linkage analysis followed by association show NRG1 associated with cannabis dependence in African-Americans. *Biol Psychiatry* 72:637–644.
- Moskvina V, Craddock N, Holmans P, Owen MJ, O’Donovan MC (2006): Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered* 61:55–64.
- Tabangin ME, Woo JG, Martin LJ (2009): The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proc* 3(suppl 7):S41.