



HHS Public Access

Author manuscript

Mol Psychiatry. Author manuscript; available in PMC 2014 December 01.

Published in final edited form as:

Mol Psychiatry. 2014 June ; 19(6): 717–723. doi:10.1038/mp.2013.99.

Genomewide association study of cocaine dependence and related traits: *FAM53B* identified as a risk gene

Joel Gelernter, MD^{1,*}, Richard Sherva, PhD², Ryan Koesterer, PhD², Laura Almasy, PhD³, Hongyu Zhao, PhD⁴, Henry R. Kranzler, MD⁵, and Lindsay Farrer, PhD^{2,6}

¹Yale University School of Medicine, Department of Psychiatry, Division of Human Genetics; and VA CT Healthcare Center; and Departments of Genetics and Neurobiology

²Boston University School of Medicine, Department of Medicine (Biomedical Genetics)

³Texas Biomedical Research Institute, Department of Genetics 4. Yale University School of Public Health, Department of Biostatistics

⁴Yale School of Public Health, Departments of Biostatistics and Genetics

⁵University of Pennsylvania Perelman School of Medicine, Department of Psychiatry and Philadelphia VAMC

⁶Boston University Schools of Medicine and Public Health, Departments of Neurology, Ophthalmology, Genetics & Genomics, Epidemiology and Biostatistics

Abstract

We report a GWAS for cocaine dependence (CD) in three sets of African- and European-American subjects (AAs and EAs, respectively), to identify pathways, genes, and alleles important in CD risk.

The discovery GWAS dataset (n=5,697 subjects) was genotyped using the Illumina OmniQuad microarray (890,000 analyzed SNPs). Additional genotypes were imputed based on the 1000 Genomes reference panel. Top-ranked findings were evaluated by incorporating information from publicly available GWAS data from 4,063 subjects. Then, the most significant GWAS SNPs were genotyped in 2,549 independent subjects.

We observed one genomewide-significant (GWS) result: rs7086629 at the *FAM53B* (“family with sequence similarity 53, member B”) locus. This was supported in both AAs and EAs; p-value (meta-analysis of all samples) = 4.28×10^{-8} . The gene maps to the same chromosomal region as the maximum peak we observed in a previous linkage study. *NCOR2* (nuclear receptor corepressor 1) SNP rs150954431 was associated with $p=1.19 \times 10^{-9}$ in the EA discovery sample. SNP rs2456778,

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Author: Joel Gelernter, MD; Yale University School of Medicine, Department of Psychiatry, Division of Human Genetics in Psychiatry; VA CT 116A2; 950 Campbell Avenue; West Haven, CT 06516; joel.gelernter@yale.edu; fax, 203-937-3897.

Conflict of Interest

Although unrelated to the current study, Dr. Kranzler has been 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’s Alcohol Clinical Trials Initiative, which is supported by Lilly, Lundbeck, Abbott, and Pfizer.

which maps to *CDKI* (“cyclin-dependent kinase 1”), was associated with cocaine-induced paranoia in AAs in the discovery sample only ($p=4.68\times 10^{-8}$).

This is the first study to identify risk variants for CD using GWAS. Our results implicate novel risk loci and provide insights into potential therapeutic and prevention strategies.

Keywords

Cocaine dependence; cocaine-induced paranoia; GWAS; population genetics; European-American and African-American populations

INTRODUCTION

Cocaine dependence (CD) is a serious form of substance dependence, with lifetime prevalence in the United States of 1.0%.¹ Cocaine use is costly to society, directly contributing to morbidity and medical costs, lost workdays, and other adverse individual, interpersonal, and societal effects.

CD is understudied, particularly in relation to the extent of the individual and societal problems it causes. Animal studies have begun to elucidate the biological substrates of CD (e.g., ref. 2), but this has not been accompanied by comparable elucidation of the sources of the genetic contribution to this trait. CD has a heritability of about 0.65 in females³ and 0.79 in males⁴, so the potential exists to identify specific genetic variants that underlie disease risk. There have been numerous candidate gene association studies of CD and related traits, and several genomewide linkage studies, the latter identifying chromosomal regions likely to harbor risk-influencing genes.⁵⁻⁶ Genomewide association studies (GWAS), when adequately powered, have generally been successful at identifying genes responsible for some of the risk for most complex traits for which they have been employed. However, no GWAS for CD has been published to date. To our knowledge, the only other GWAS for an illegal substance dependence (SD) diagnosis with genomewide-significant (GWS) results is our investigation of opioid dependence (OD).⁷ A previous GWAS of cannabis dependence⁸ did not report GWS results.

In the present study, we sought to identify genes that modify risk for CD by means of a GWAS in family-based and case-control samples of 2,379 European Americans (EAs), including 1,809 subjects with CD, and 3,318 African Americans (AAs), including 2,482 subjects with CD. Multiple independent samples of EAs and AAs (2,549 identically ascertained subjects that we collected and 4,063 subjects from the Study of Addiction: Genetics and Environment (SAGE) dataset) were used to replicate and extend our findings. We identified one novel CD risk locus at genomewide significance (GWS) and numerous others, relevant to CD and the related trait of cocaine-induced paranoia (CIP), worthy of future investigation.

MATERIALS AND METHODS

Subjects and Diagnostic Procedures

The GWAS discovery sample included a total of 5,697 subjects. A replication dataset (identically evaluated) was comprised of 2,549 subjects and was genotyped for individual markers. (Public domain GWASed samples were included in some analyses as well, as described below.) All subjects were recruited for studies of the genetics of cocaine, opioid, or alcohol dependence. The sample consisted of small nuclear families (SNFs) originally collected for linkage studies, and unrelated individuals. Subjects were recruited at five US clinical sites: Yale University School of Medicine (APT Foundation; New Haven, CT), the University of Connecticut Health Center (Farmington, CT), the University of Pennsylvania Perelman School of Medicine (Philadelphia, PA), the Medical University of South Carolina (Charleston, SC), and McLean Hospital (Belmont, MA). Details regarding the sample can be found in Table 1 and Supplementary Tables 1 and 2. Our previous CD linkage study⁵ included a subset of the SNFs included in this study.

All subjects were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)^{5,9} to derive diagnoses for lifetime CD and other major psychiatric traits according to DSM-IV¹⁰ criteria. CIP was assessed with a specific item on the SSADDA, which we have previously shown to be valid.¹¹ The inter-rater reliability of the SSADDA for the diagnosis of CD was excellent ($\kappa=0.83$),⁹ as was the reliability for the CIP trait assessment ($\kappa=0.86$).

The distribution of symptom count observations is given in Supplementary Figure S1.

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.

Genotyping and Quality Control

Samples from individuals in the discovery sample were genotyped on the Illumina HumanOmni1-Quad v1.0 microarray (988,306 autosomal SNPs). GWAS genotyping was conducted at the Yale Center for Genome Analysis (YCGA) and the Center for Inherited Disease Research (CIDR). Genotypes were called using GenomeStudio software V2011.1 and genotyping module version 1.8.4 (Illumina, San Diego, CA, USA). A total of 44,644 SNPs on the microarray and 135 individuals with call rates < 98% were excluded, and 62,076 additional SNPs were removed due to minor allele frequencies (MAF) < 1%. Additional quality control details are described in Supplementary Materials. SAGE samples (see below) were genotyped on the Illumina Human1M array.

Follow-up genotyping in the replication sample was performed using a custom Illumina GoldenGate Genotyping Universal-32, 1536-plex microarray assay. Most SNPs included in the custom array were selected for studies of other phenotypes. Additional SNPs were genotyped individually using the TaqMan method.¹²

To verify and correct misclassification of self-reported race, we compared the GWAS data from all subjects with the genotypes from the HapMap 3 reference CEU, YRI, and CHB populations. Principal components (PCs) analysis was conducted in the discovery GWAS sample using 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%) in each sample to characterize the underlying genetic architecture by deriving 10 PCs for each individual. The PCs were used to distinguish EAs from AAs by a K-means (K=2) clustering algorithm¹⁵ and the two groups were analyzed separately. Because many subjects self-identified as EA Hispanic or AA Hispanic, PC analyses were repeated within the AA and EA groups, and the first three PCs in each were used in all subsequent analyses to correct for residual population stratification within the group.⁷

The same procedures to address population classification and substructure within groups were applied to the SAGE dataset.

Additional GWAS Sample: SAGE

In Phase 2 analyses described below, we included publicly available GWAS data (obtained via an application process) from the SAGE dataset, including individuals from the Collaborative Study on the Genetics of Alcoholism (COGA),¹⁶ the Family Study of Cocaine Dependence (FSCD),¹⁷ and the Collaborative Genetic Study of Nicotine Dependence (COGEND) (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1).¹⁸ Information on these samples is provided in Supplementary Materials. The combined SAGE analysis set contained 1,311 AA and 2,752 EA individuals (Supplementary Table 1).

Analysis Overview

There were three independent sets of subjects employed in these CD-phenotype analyses, in three phases. Phases 1 and 2 evaluated GWAS data from two different sets of subjects using two different, but similarly dense, microarrays: *Phase 1* included our GWAS discovery dataset, consisting of 5,697 subjects. *Phase 2* also incorporated information from the SAGE dataset, with GWAS data from 4,063 subjects. The assessments used for our study (SSADDA) and the SAGE study are sufficiently similar for most phenotype data to be combinable directly. *Phase 3* included our replication dataset of 2,549 subjects who were (directly) genotyped for selected SNPs rather than for GWAS. Thus, analyses included up to 8,246 of our own subjects and 12,309 subjects overall. The overall analytic design is similar to that of our OD GWAS study.⁷ In our own subjects, we also analyzed the CIP trait (which is included in the SSADDA but not the SAGE assessment), limited to subjects with cocaine exposure.

Genotype Imputation

Genotypes for 37,426,733 SNPs were imputed with IMPUTE2¹⁹ using the genotyped SNPs and the 1000 Genomes reference panel released in June of 2011 (<http://www.1000genomes.org/>), which contains phased haplotypes for 1,094 individuals of various ancestries.²⁰ EA and AA samples were imputed separately. We considered for analysis imputed SNPs with r^2 greater than 0.8.

Statistical Methods for Association Analyses

Association tests in the GWAS datasets (our own dataset individually in Phase 1, then combined with the SAGE dataset in Phase 2) used linear or logistic association models with generalized estimating equations (GEE) to correct for the correlations among related individuals. We evaluated the replication sample of unrelated individuals—the part of the sample that was genotyped individually only for replication SNPs—using linear and logistic models. All models were adjusted for age, sex, and the first three PCs of ancestry.

Three primary analyses to identify genetic factors contributing to risk for CD

1. A model (Sympcount_{adj}) used imputed minor allele dosage as the dependent variable and DSM-IV symptom count for CD and each of three other major SD diagnoses (opioid, alcohol, and nicotine dependence—OD, AD, and ND, respectively) as ordinal predictors of genotype. This allowed us to remove the effect attributable to substances other than CD, thereby facilitating the identification of genetic risk factors unique to that trait by limiting confounding due to comorbid dependence symptoms. All individuals contributed to this analysis, including those meeting DSM-IV criteria for CD and individuals with 0–2 CD symptoms (who did not receive a diagnosis of CD). The ordinal trait model has greater power to detect genetic associations than a univariate model based on disease status because it contains more information and is more specific. The β coefficient and p-value for the CD symptom count (adjusted for the symptom counts for OD, AD, and ND) were used to assess the magnitude and significance of the association, respectively. To ensure that modeling minor allele dose as the dependent variable did not produce unreliable results and to assess the effects of comorbid dependence, we tested *post hoc* the top SNPs identified from this model in a model (Sympcount) using CD symptom count as the dependent variable and SNP (not adjusted for OD, AD, or ND) as the independent variable.
2. We used case-control status as the outcome in logistic regression models but included as controls only individuals who had used cocaine at least once in their lives without becoming dependent. This excludes subjects who have genetic liability but were never exposed to cocaine (i.e., “false negative” cases).
3. We used logistic regression to examine association with cocaine-induced paranoia (CIP). A majority of chronic cocaine users experience transient paranoid symptoms that typically resolve with abstinence.^{21–23} CIP represents a genetically distinct phenotype reflecting inter-individual differences in cocaine response.^{23,24} Subjects who answered the question, “Have you ever had a paranoid experience when you were using cocaine?” affirmatively were diagnosed as being affected with CIP. Subjects with CIP must be cocaine-exposed and most meet CD criteria.

In each model, the data were analyzed separately by population group, and the results from the two groups were combined by meta-analysis using the inverse variance method implemented in the computer program METAL.²⁵

Replication of Top Findings

In Phase 2, SNPs with $p < 1.0 \times 10^{-3}$ in either EAs or AAs, or in the EA-AA meta-analysis, were tested for association in the SAGE dataset using identical statistical models (3,381 SNPs from the $\text{Symcount}_{\text{adj}}$ model, 3,323 from the case-control model). Results from Phases 1 and 2 were combined within population groups by meta-analysis. The threshold for evaluating a SNP in Phase 2 was chosen to minimize false negatives, assuming that an equally strong effect observed in the Phase 2 sample would result in a GWS meta-analysis result. Based on the combined Phase 1+2 meta-analysis, we selected 153 SNPs ($\text{Symcount}_{\text{adj}}$, $N=34$; case-control, $N=84$; CIP, $N=39$) for further replication in Phase 3 based on a cutoff of $p < 1.0 \times 10^{-4}$ (four SNPs met this criterion for more than one trait).

Pathway Analysis

We used the association results from the discovery + SAGE meta-analysis (i.e. Phase 1 \pm Phase 2) for each of the primary traits (except CIP) to conduct a pathway analysis with the Ingenuity Pathway Analysis (IPA) software suite (<http://www.ingenuity.com>). First, the number of independent SNP association tests for each gene in the genome (only including SNPs within the transcribed portion of each gene) was computed according to the method of Li and Ji.²⁶ We multiplied the smallest P-value within a gene by the number of independent SNPs in that gene and created a list of genes containing a SNP with gene-based multiple-test-corrected significance ($P_{\text{adj}} < 0.05$). The genes in the list were evaluated by pathway analysis to identify an overrepresentation of genes within defined canonical pathways based on information culled from multiple sources. A Fisher's exact P-value was computed for each pathway indicating whether, after accounting for the total number of pathways and the number of genes in a given pathway, there were more significantly associated SNPs than would be expected by chance. Separate gene lists were created for the $\text{Symcount}_{\text{adj}}$ and case-control results within each population.

RESULTS

We observed GWS association of CD with rs2629540 at the *FAM53B* ("family with sequence similarity 53, member B") locus in the Symcount model after removing OD, AD, and ND symptom counts as covariates (Figure 1). This was supported by evidence in both AAs and EAs. The p-value for all samples combined was 4.28×10^{-8} (Table 2). Under the same model, *NCOR2* (nuclear receptor corepressor 1) SNP rs150954431 was associated in the EA discovery sample at the GWS level ($p=1.19 \times 10^{-9}$), but there were no consistent observations of this association in any other sample. Numerous additional SNPs were associated at the 10^{-7} level. We also observed GWS association of CIP with rs2456778, which maps to *CDKI* ("cyclin-dependent kinase 1"), in the AA discovery sample ($p=4.68 \times 10^{-8}$). This association nearly reached nominal significance in the EA discovery sample ($p=0.0502$) and was slightly improved in the two populations combined ($p=4.26 \times 10^{-8}$), but was not well supported in the Phase 3 replication sample. Phenotypic data for CIP were not available for SAGE. Additional associations ($p < 1 \times 10^{-6}$) were observed with numerous other SNPs in AAs, EAs or in both groups combined (Table 2). (Manhattan plot and associated Q/Q plot, Figures 2 and 3; Complete results, Supplementary Table 3).

The pathway analyses identified several pathways significantly associated with CD. The most significant canonical pathway was calcium transport in the AA case-control analysis ($p = 0.002$) (Supplementary Figure S2). The pathway was identified via associations in two genes encoding Ca^{2+} -transporting ATPases, which are important for Ca^{2+} homeostasis: ATPase, Ca^{2+} -transporting, plasma membrane (*ATP2B2*) and ATPase, Ca^{2+} -transporting, type 2C, member 2 (*ATP2C2*). The highest ranked networks from both the EA Sympcount_{adj} analysis and the AA case control analysis, and the second highest ranked network from the AA Sympcount_{adj} model, showed associated genes (*SNAP25*, *KCNQ4*, *KCNN2*, and *ATP2B2*) with direct interactions with *CALM1*, which encodes calmodulin, a key calcium binding protein (Supplementary Figures S3, S4).

DISCUSSION

This is, to our knowledge, the first GWAS reported for CD. To obtain these findings, we made use of our own SSADDA-assessed GWAS sample, an additional SSADDA-assessed replication sample, and publicly available data from the SAGE project. Our strongest finding statistically, and the only one that meets genomewide significance in the entire sample (i.e., $p < 5 \times 10^{-8}$), is at *FAM53B* (Figure 2, Table 2). Although both the AA and EA parts of the sample contributed to this association signal, it was stronger in AAs, where the MAF was 0.07, vs. 0.25 in the EAs. *FAM53B* falls within the 1-lod support interval of the most significant genome-wide linkage peak for CD (lod score 2.7) that we identified previously.⁵ As in the present association result, both the EA and AA parts of the sample contributed to the linkage finding. It is relatively uncommon for association and linkage findings to coincide in this way. The previous positional information from linkage increases the probability that the association finding is valid.

FAM53B seems to play a role in regulating cell proliferation,²⁷ but additional work is needed to determine the relationship of this function to CD risk or whether the gene has additional biological functions. The effect was strongest for the Sympcount measure unadjusted for comorbid dependence, indicating that the gene may influence susceptibility to CD with a co-occurring SD disorder, or SD more generally.

Several other SNPs attained genomewide significance in some analysis phases or in specific subgroups or were nearly GWS. We observed association of the DSM-IV diagnosis of CD with two SNPs near *RANP6* (rs4129566 and rs11944332) in the AA discovery sample (4.26×10^{-7} and 4.11×10^{-7} , respectively) and the Phase 3 EA sample ($p = 1.92 \times 10^{-4}$ and 2.17×10^{-4} , respectively). In the EA discovery sample, CD was also associated with rs6677435 (2.18×10^{-7}), located approximately 400 kb from its nearest gene, *KCNT2*, which encodes a potassium voltage-gated channel that we previously reported to be associated ($p = 2.1 \times 10^{-7}$) with OD in AAs.⁷ There was also evidence of association with rs1757939 ($p = 7.88 \times 10^{-7}$) in the AA discovery sample. This SNP is approximately 132kb 5' of *SCLT1*, which encodes a protein that links the voltage-gated sodium channel Na(v)1.8 with clathrin. Other notable associations include that of *CDK1* (cyclin-dependent kinase 1, a serine/threonine protein kinase) and cocaine-induced paranoia (4.86×10^{-8} in AAs and 0.0502 in EAs), and *NCOR2* (nuclear receptor corepressor 2) and CD symptom count ($p = 1.19 \times 10^{-9}$ in EAs). These Phase 1 (discovery sample) results, some reaching GWS, were not replicated.

Although these may be false positive findings, which is more likely when the MAF is <5% (as for rs72840936 and *NCOR1*), our replication samples were smaller than the discovery sample, so that lack of replication in later study phases could reflect inadequate statistical power.

Similar to *FAM53B*, the meta-analysis result for rs2005290—a SNP located in a cluster of olfactory receptor genes, between *OR3A1* and *OR3A2* and about 8 kb from each—is supported by evidence in both populations ($p=4.47\times 10^{-7}$). It may be relevant to this finding implicating olfaction that variation in a taste receptor gene was previously associated with alcohol dependence.²⁸ Further, these genes have a structure similar to that of neurotransmitter and hormone receptors.

The pathway analysis results are noteworthy primarily because they implicate variation in systems regulating calcium signaling. Although there were no individually GWS findings in calcium-system genes, calcium signaling was one of two primary domains implicated in our recent opioid dependence GWAS (the other was potassium signaling). There is prior association evidence of calcium system genes in cocaine dependence (e.g., neuronal calcium sensor 1, *NCSI*).²⁹ Our results approaching GWS in the discovery Phase 1 sample obtained with SNPs near *KCNT2* and *SCLT1* suggest a possible link between CD and potassium signaling. Although the evidence of overlap in risk loci for opioid and cocaine dependence is limited, it is consistent with the high rate of comorbidity of these two disorders.

This study generated numerous remarkable findings, including those at or near GWS. Several design factors may have contributed to the results. First, we studied two distinct populations of reasonable sample size. Second, one of the analytic models that we employed defined cocaine-related effects as an ordinal trait. This approach increased the average phenotypic information for subjects and increased power; similar approaches have been used successfully in previous SD GWAS, e.g., for alcohol dependence.³⁰ This was especially important in light of our case-control design, which used exposed controls, reducing sample size for those analyses (but excluding from the control group individuals who were unexposed to cocaine and who therefore can reasonably be considered diagnosis-unknown in this context).

Our findings should be viewed in the context of several limitations. In Phases 1 and 2 (but not Phase 3), many associated loci were imputed, albeit with excellent quality (Table 2). Although in absolute terms our sample was reasonably large (over 12,000 subjects considered overall), in the context of complex trait GWAS, it is still modest. This factor may have led to false negative findings at all phases of the study. In addition, 11 of the 27 top-ranked associations in Phase 1 were observed with infrequent alleles (<5%); as none of these results replicated in subsequent phases, they may be false positives. Finally, our findings are not adjusted for testing association in two populations and with three (albeit highly correlated) traits. However, a Bonferroni correction is too conservative given the high correlation among the traits and distinct hypotheses for EAs and AAs (different populations frequently have different common risk alleles). Future studies in large independent samples are necessary to address these concerns.

In summary, we identified one locus with GWS support for association to CD, and others with more limited support. Although there have been prior GWAS for related traits, such as methamphetamine response³¹ and opioid sensitivity,³² to our knowledge, there are no studies published for stimulant dependence *per se*. The risk loci we identified did not conform to what might have been regarded as the most likely candidate gene predictions, and therefore will lead to novel directions in research that aims to increase our understanding of the genetics and pathophysiology of cocaine dependence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We appreciate the work in recruitment and assessment provided 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 GWAS study were provided by the Center for Inherited Disease Research (CIDR) 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 SSADDA interviewers, led by Yari Nuñez and Michelle Slivinsky, who devoted substantial time and effort to phenotype the study sample and to John Farrell for database management assistance. This study was supported by National Institutes of Health grants RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, R01 AA017535, and the VA Connecticut and Philadelphia VA MIRECCs.

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 (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support 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).

References

1. Compton WM, Thomas YF, Stinson FS, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry*. 2007; 64(5):566–576. [PubMed: 17485608]
2. Lobo MK, Covington HE III, Chaudhury D, Friedman AK, Sun HS, Domez-Werno D, et al. Cell type specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science*. 2010; 330:385–390. [PubMed: 20947769]
3. Kendler KS, Prescott CA. Cocaine use, abuse and dependence in a population-based sample of female twins. *Br J Psychiatry*. 1998; 173:345–350. [PubMed: 9926041]
4. Kendler KS, Karkowski LM, Neale MC, Prescott CA. Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. *Arch Gen Psychiatry*. 2000; 57:261–269. [PubMed: 10711912]

5. Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, et al. 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.* 2005; 136(1):45–52.
6. Yang BZ, Han S, Kranzler HR, Farrer LA, Elston RC, Gelernter J. Autosomal linkage scan for loci predisposing to comorbid dependence on multiple substances. *Am J Med Genet B Neuropsychiatr Genet.* 2012; 159B(4):361–9. [PubMed: 22354695]
7. Gelernter, J.; Kranzler, HR.; Sherva, R.; Koesterer, R.; Sun, J.; Bi, J. Genomewide association study of opioid dependence and related traits: multiple associations mapped to calcium and potassium pathways. in review
8. Agrawal A, Lynskey MT, Hinrichs A, Grucza R, Saccone SF, Krueger R, et al. A genome-wide association study of DSM-IV cannabis dependence. *Addict Biol.* 2011; 16(3):514–8. [PubMed: 21668797]
9. Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, et al. Diagnostic Reliability of the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug Alcohol Depend.* 2005; 80(3):303–12. [PubMed: 15896927]
10. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 4. Washington, DC: American Psychiatric Press; 1994.
11. Cubells JF, Feinn R, Pearson D, Burda J, Tang Y, Farrer LA, et al. Rating the severity and character of transient cocaine-induced delusions and hallucinations with a new instrument, the Scale for Assessment of Positive Symptoms for Cocaine-Induced Psychosis (SAPS-CIP). *Drug Alcohol Depend.* 2005; 80:23–33. [PubMed: 15894433]
12. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5' 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci USA.* 1991; 88:7276–7280. [PubMed: 1871133]
13. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38(8): 904–9. [PubMed: 16862161]
14. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006; 2(12):e190. [PubMed: 17194218]
15. Hartigan JA, Wong MA. A K-means clustering algorithm. *Applied Statistics.* 1979; 28:100–108.
16. Edenberg HJ. The collaborative study on the genetics of alcoholism: an update. *Alcohol Res Health.* 2002; 26:214–218. [PubMed: 12875050]
17. Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB. Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend.* 2008; 95:14–22. [PubMed: 18243582]
18. Bierut LJ. Genetic variation that contributes to nicotine dependence. *Pharmacogenomics.* 2007; 8:881–883. [PubMed: 17716221]
19. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet.* 2009; 5(6):e1000529.10.1371/journal.pgen.1000529 [PubMed: 19543373]
20. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467:1061–1073. [PubMed: 20981092]
21. Brady KT, Lydiard RB, Malcolm R, Ballenger JC. Cocaine-induced psychosis. *J Clin Psychiatry.* 1991; 52:509–512. [PubMed: 1752853]
22. Satel SL, Southwick SM, Gawin FH. Clinical features of cocaine-induced paranoia. *Am J Psychiatry.* 1991; 148:495–498. [PubMed: 2006696]
23. Cubells JF, Feinn R, Pearson D, Burda J, Tang Y, Farrer LA, Gelernter J, Kranzler HR. Rating the severity and character of transient cocaine-induced delusions and hallucinations with a new instrument, the Scale for Assessment of Positive Symptoms for Cocaine-Induced Psychosis (SAPS-CIP). *Drug Alcohol Depend.* 2005; 80:23–33. [PubMed: 15894433]
24. Farrer LA, Kranzler HR, Yu Y, Weiss RD, Brady KT, Cubells JF, Gelernter J. Association of variants in MANEA with cocaine-related behaviors. *Arch Gen Psychiat.* 2009; 3:267–74. [PubMed: 19255376]

25. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–2191. [PubMed: 20616382]
26. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)*. 2005; 95(3):221–7. [PubMed: 16077740]
27. Thermes V, Candal E, Alunni A, Serin G, Bourrat F, Joly JS. Medaka simplet (FAM53B) belongs to a family of novel vertebrate genes controlling cell proliferation. *Development*. 2006; 133:1881–90. [PubMed: 16611694]
28. Hinrichs AL, Wang JC, Bufo B, Kwon JM, Budde J, Allen R, et al. Functional Variant in a Bitter-Taste Receptor (hTAS2R16) Influences Risk of Alcohol Dependence. *Am J Hum Genet*. 2006; 78:103–111. [PubMed: 16385453]
29. Multani PK, Clarke TK, Narasimhan S, Ambrose-Lanci L, Kampman KM, Pettinati HM, et al. Neuronal calcium sensor-1 and cocaine addiction: a genetic association study in African-Americans and European Americans. *Neurosci Lett*. 2012; 531(1):46–51. [PubMed: 22999924]
30. Wang JC, Foroud T, Hinrichs AL, Le NX, Bertelsen S, Budde JP, et al. A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53. *Mol Psychiatry*. 2012 [Epub ahead of print]. 10.1038/mp.2012.143
31. Hart AB, Engelhardt BE, Wardle MC, Sokoloff G, Stephens M, et al. Genome-Wide Association Study of d-Amphetamine Response in Healthy Volunteers Identifies Putative Associations, Including Cadherin 13 (CDH13). *PLoS ONE*. 2012; 7(8):e42646.10.1371/journal.pone.0042646 [PubMed: 22952603]
32. Nishizawa D, Fukuda K, Kasai S, Hasegawa J, Aoki Y, Nishi, et al. Genome-wide association study identifies a potent locus associated with human opioid sensitivity. *Mol Psychiatry*. 2012 [Epub ahead of print]. 10.1038/mp.2012.164
33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet*. 2007; 81(3):559–75. Epub 2007 Jul 25. [PubMed: 17701901]

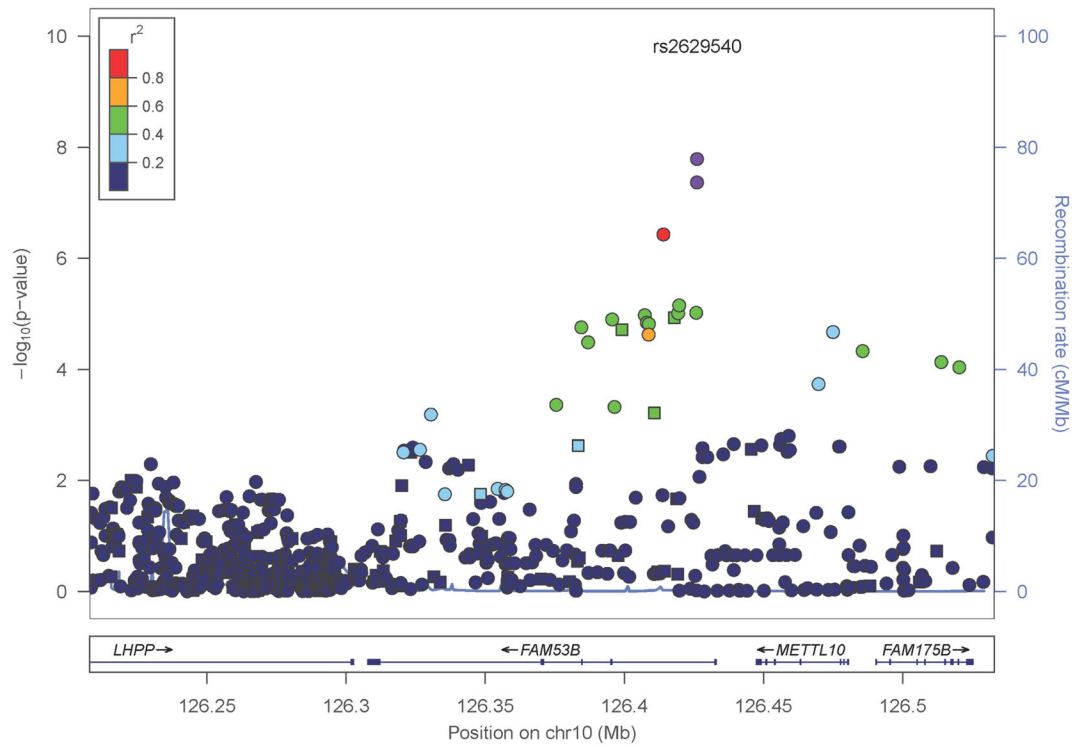


Figure 1.

Regional Manhattan plot for *FAM53B*, showing the meta-analysis P-value from EAs and AAs in the discovery and SAGE, as well as a single point for the phase 1–3 meta-analysis result (the highest purple dot on the graph). Since the result is driven primarily by AAs, the LD heat map is based on AAs also. Imputed SNPs are shown as circles, and genotyped SNPs as squares.

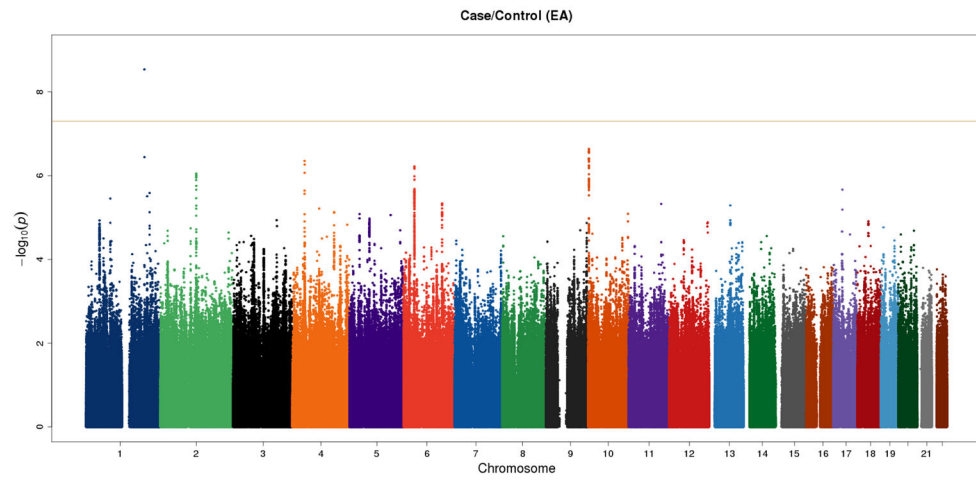


Figure 2.
Manhattan plot, CD case/control analysis, EA population (discovery sample)

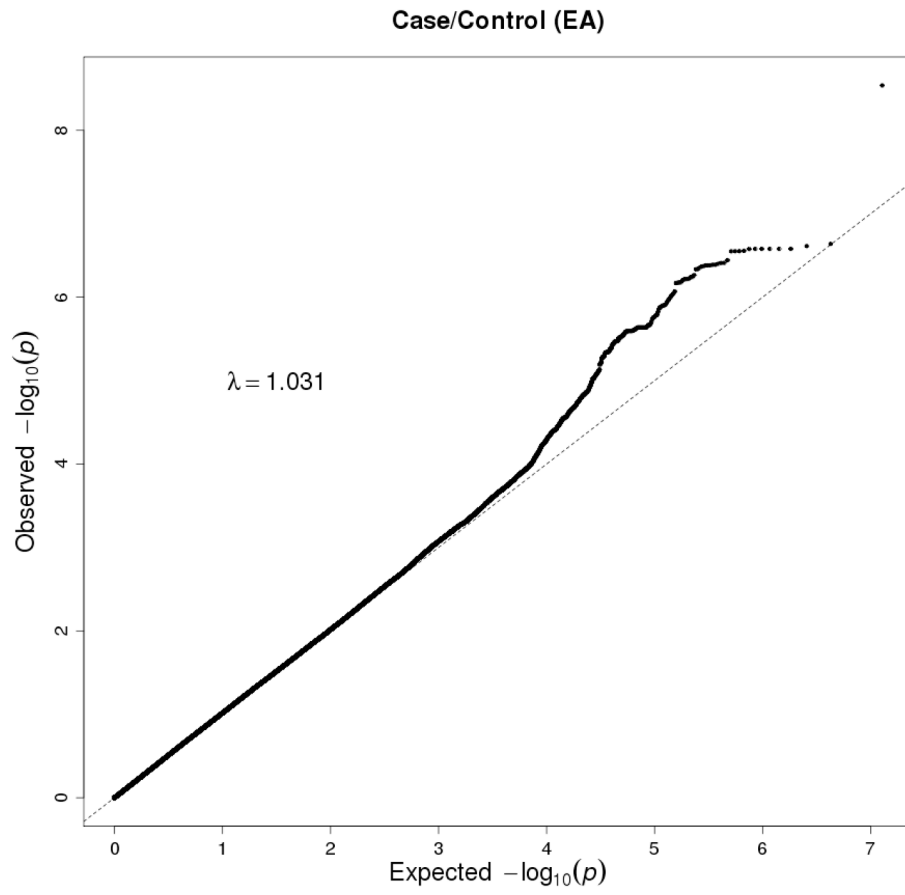


Figure 3.
Q/Q plot for same. Other Q/Q plots similarly showed negligible inflation.

Table 1

Sample characteristics.

<u>Sample Description</u>	GWAS Sample (SNFs)				GWAS Sample (Unrelateds)				Replication Sample (Unrelateds)				<i>Total</i>	
	AA	Male	Female	EA	AA	Male	Female	EA	AA	Male	Female	EA	AA	EA
Yale (APT Foundation)	199	257	141	108	453	370	485	290	223	198	474	477	1700	1975
University of CT	174	227	155	161	455	355	451	296	127	93	315	299	1431	1677
MUSC	42	84	52	47	53	109	33	29	21	24	47	47	333	255
McLean Hospital	44	36	42	30	10	6	18	11	0	2	2	3	98	106
Univ. Pennsylvania	9	11	0	0	288	136	20	10	51	64	43	39	559	112
PD Sample: SAGE	52	71	23	53	591	597	1199	1477					4121	4125

<u>Sample, by diagnosis</u>	AA			EA			<i>Total</i>	
	Male (%)	Average age	Average symptom count	Male (%)	Average age	Average symptom count	AA	EA
GWAS Affecteds	56	42	5.9	59	37.7	6	2482	1809
GWAS Unaffecteds	38	38.3	0.1	58	39.5	0.27	800	485
CD with CIP	59	42.3	6.2	62	37.7	6.2	1703	1242
CD without CIP	51	42.7	5.5	50	37.8	5.5	779	567
GWAS Exposed Unaffecteds	52	42.5	0.6	63	37.2	0.5	186	292
Replication Affecteds	65	44.4	6.1	62	38.3	6.2	315	415
Replication Unaffecteds	41	38.5	0.2	46	41.4	0.06	438	1269
Replication Exposed Unaffecteds	54	43.1	0.7	56	37.2	0.3	113	263
PD Sample: SAGE Affecteds	55	40.4	6.2	62	35.4	6	564	540
PD Sample: SAGE Unaffected	44	39.6	0.08	40	39.4	0.06	744	2207
PD Sample: SAGE Exposed Unaffecteds	61	40.7	0.4	54	38.2	0.3	138	535

Table 2

Results from each GWAS phase where at least one result generated $p < 10^{-6}$. AA=African American; EA=European American; RAF=reference allele frequency; RSO=imputation quality; genome wide significant results shown in bold, italicized font and underlined; p-values $< 1 \times 10^{-6}$, underlined; SNPs in bold were genotyped in phase 1; Meta=meta-analysis results.

Chr	SNP	Gene	Phase 1			Phase 2						Phase 3						Meta.EA.P	Meta.AA.P	Meta.AA.P
			Meta Phase 1	RAF.AA	RAF.EA	RSO.AA	RSO.EA	AA.P	EA.P	RAF.AA	RAF.EA	RSO.AA	RSO.EA	AA.P	EA.P	RAF.AA	RAF.EA			
1	rs200085570	NA		0.97	0.96	0.71	0.83	7.32E-01	3.28E-01	0.96	0.96	0.76	0.88	NA	NA	NA	NA	NA		
1	rs6677435	NA		0.99	0.96	0.77	0.84	8.36E-01	1.41E-01	0.98	0.96	0.980	0.98	5.05E-01	9.29E-01	0.02	0.05	8.61E-01	8.82E-06	
1	rs116439821	LGR6		0.96	1.00	0.91	0.91	5.67E-01	NA	0.96	1.00	0.94	0.00	4.53E-01	NA	0.04	1.00	1.55E-05	1.55E-05	
2	rs72840936	SITEP3		0.98	0.96	0.87	0.96	6.29E-01	1.71E-01	0.98	0.96	0.90	0.99	9.84E-01	3.41E-01	0.01	0.04	4.61E-01	3.19E-06	
3	rs111325002	NA		0.97	1.00	0.93	0.17	4.41E-01	NA	0.98	1.00	0.96	0.82	NA	NA	NA	NA	3.70E-07	3.70E-07	
4	rs4861386	UCHL1		0.51	0.41	0.96	0.99	4.18E-01	4.51E-01	0.54	0.43	0.96	0.96	8.35E-01	7.70E-01	0.51	0.41	3.32E-01	1.61E-04	
4	rs1757939	SCLT1		0.41	0.43	0.95	0.93	4.02E-01	9.64E-01	0.40	0.42	0.99	0.99	3.09E-01	9.45E-01	0.38	0.42	2.90E-05	8.05E-01	
4	rs4129566	NA		0.90	0.86	0.98	0.99	2.05E-01	8.16E-01	0.89	0.86	0.99	0.94	7.13E-01	1.92E-04	0.90	0.86	2.89E-06	5.49E-02	
4	rs11944332	RANP6		0.90	0.86	0.98	0.99	2.08E-01	8.57E-01	0.89	0.86	0.99	0.94	9.19E-01	2.17E-04	0.91	0.86	1.86E-06	5.95E-02	
6	rs6912117	PXT1		0.69	0.79	0.97	0.99	9.47E-01	3.96E-01	0.71	0.80	0.94	0.98	6.52E-01	8.54E-01	0.69	0.79	7.16E-02	1.60E-03	
6	rs5955083	PXT1		0.69	0.79	0.97	0.99	9.29E-01	3.59E-01	0.71	0.80	0.94	0.99	NA	NA	NA	NA	8.09E-02	7.95E-04	
8	rs75686122	RIMS2		0.96	0.91	0.97	0.97	7.68E-01	4.70E-01	0.97	0.91	0.95	1.00	3.03E-01	5.78E-01	0.02	0.08	2.83E-06	5.21E-01	
10	rs34831910	NA		0.98	0.87	0.96	0.98	8.57E-01	3.40E-01	0.98	0.85	0.95	1.00	3.47E-01	3.93E-01	0.04	0.13	6.86E-01	9.13E-03	
10	rs7899919	NA		0.98	0.87	0.97	1.00	6.67E-01	2.36E-01	0.98	0.86	0.96	0.98	8.78E-01	3.71E-01	0.04	0.13	4.25E-01	8.52E-04	
10	rs9664175	NA		0.98	0.87	0.96	1.00	6.84E-01	2.51E-01	0.98	0.86	0.96	0.98	3.94E-01	2.86E-01	0.03	0.13	2.93E-01	4.72E-04	
10	rs7086629	CHST3		0.96	1.00	0.99	NA	6.04E-01	NA	0.96	1.00	0.99	NA	7.32E-01	9.74E-01	0.96	1.00	5.11E-05	NA	
17	rs2005290	OR3A2/OR3A1		0.15	0.06	0.79	0.66	4.43E-01	5.98E-02	0.14	0.04	0.80	0.65	6.16E-01	1.42E-01	0.35	0.14	1.49E-02	1.96E-06	
17	rs114903983	HSF5		0.94	1.00	0.99	NA	7.31E-01	NA	0.96	1.00	0.94	NA	3.00E-01	7.90E-01	0.95	1.00	1.56E-04	NA	
17	rs116087723	MTMR4		0.94	1.00	1.00	NA	6.16E-01	NA	0.96	1.00	0.96	NA	1.34E-01	7.90E-01	0.06	0.00	2.92E-04	NA	
18	rs79794368	MYL12A		0.96	1.00	0.93	NA	4.16E-01	NA	0.96	1.00	0.88	NA	1.20E-01	NA	0.05	0.00	1.48E-05	NA	
18	rs13381416	MYL12A		0.96	1.00	0.93	NA	4.13E-01	NA	0.96	1.00	0.88	NA	6.49E-02	NA	0.95	1.00	1.06E-05	NA	
18	rs61751192	MYL12A		0.96	1.00	0.93	NA	4.12E-01	NA	0.96	1.00	0.89	NA	1.08E-01	NA	0.05	0.00	1.73E-05	NA	
18	rs12956327	FAM69C		0.93	0.81	0.86	0.93	8.50E-01	1.65E-01	0.93	0.81	0.80	0.93	9.26E-01	9.27E-01	0.06	0.17	1.57E-05	3.52E-01	
Symptom count																				
10	rs2629540	FAM53B		0.93	0.75	0.94	0.95	6.02E-02	9.44E-02	0.94	0.74	0.95	0.96	4.57E-01	4.87E-01	0.94	0.77	1.38E-06	2.62E-03	
12	rs150954431	NCOR2		0.99	0.97	0.63	0.65	5.84E-01	1.47E-01	0.99	0.97	0.61	0.59	3.32E-01	1.83E-01	1.00	0.98	5.66E-01	5.35E-07	
16	rs4782559	CDH13		0.22	0.46	0.90	0.97	7.90E-01	9.62E-01	0.25	0.47	0.83	0.93	NA	NA	1.00	1.00	1.82E-05	2.77E-01	
Cocaine induced Paranoia																				
10	rs2456778	CDK1		0.25	0.26	0.98	0.98	NA	NA	NA	NA	NA	NA	1.58E-01	6.36E-01	0.23	0.23	4.77E-06	5.53E-02	