



Published in final edited form as:

*Am J Med Genet B Neuropsychiatr Genet.* 2014 March ; 0(2): 148–156. doi:10.1002/ajmg.b.32216.

## Comparing the Utility of Homogeneous Subtypes of Cocaine Use and Related Behaviors With DSM-IV Cocaine Dependence as Traits for Genetic Association Analysis

Jinbo Bi<sup>1</sup>, Joel Gelernter<sup>2,3</sup>, Jiangwen Sun<sup>1</sup>, and Henry R. Kranzler<sup>4,5,\*</sup>

<sup>1</sup>Department of Computer Science and Engineering, School of Engineering, University of Connecticut, Storrs, Connecticut

<sup>2</sup>Departments of Psychiatry, Genetics, and Neurobiology, Yale University School of Medicine, New Haven, Connecticut

<sup>3</sup>Department of Psychiatry, VA CT Healthcare Center, West Haven, Connecticut

<sup>4</sup>Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania

<sup>5</sup>VISN 4 Mental Illness Research, Education, and Clinical Center, Philadelphia, Pennsylvania

### Abstract

Because DSM-IV cocaine dependence (CD) is heterogeneous, it is not an optimal phenotype to identify genetic variation contributing to risk for cocaine use and related behaviors (CRBs). We used a cluster analytic method to differentiate homogeneous, highly heritable subtypes of CRBs and to compare their utility with that of the DSM-IV CD as traits for genetic association analysis. Clinical features of CRBs and co-occurring disorders were obtained via a poly-diagnostic interview administered to 9,965 participants in genetic studies of substance dependence. A subsample of subjects (N = 3,443) were genotyped for 1,350 single nucleotide polymorphisms (SNPs) selected from 130 candidate genes related to addiction. Cluster analysis of clinical features of the sample yielded five subgroups, two of which were characterized by heavy cocaine use and high heritability: a *heavy cocaine use, infrequent intravenous injection group* and an *earlyonset, heavy cocaine use, high comorbidity group*. The utility of these traits was compared with the CD diagnosis through association testing of 2,320 affected subjects and 480 cocaine-exposed controls. Analyses examined both single SNP (main) and SNP–SNP interaction (epistatic) effects, separately for African-Americans and European-Americans. The two derived subtypes showed more significant *P* values for 6 of 8 main effects and 7 of 8 epistatic effects. Variants in the *CLOCK* gene were significantly associated with the *heavy cocaine use, infrequent intravenous injection group*, but not with the DSM-IV diagnosis of CD. These results support the utility of subtypes based on CRBs to detect risk variants for cocaine addiction.

© 2013 Wiley Periodicals, Inc

\*Correspondence to: Henry R. Kranzler, M.D., Department of Psychiatry, Treatment Research Center, University of Pennsylvania Perelman School of Medicine, 3900 Chestnut Street, Philadelphia, PA 19104-6178. kranzler@mail.med.upenn.edu.

Conflict of interests: Dr. Bi, Mr. Sun, and Dr. Gelernter have no disclosures.

### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

## Keywords

cocaine; cocaine dependence; candidate genes; subtyping; subtype analysis

---

## INTRODUCTION

Cocaine use is common in the United States. The National Survey on Drug Use and Health reported that in 2010 an estimated 0.6% of Americans aged 12 or older (i.e., one million people) used cocaine within the prior month (Substance Abuse and Mental Health Services Administration, 2011). Cocaine dependence (CD) is associated with serious social, legal, medical, and psychiatric problems and thus represents a major public health problem in the United States.

Results from adoption, twin, and family studies demonstrate a substantial genetic contribution to the risk for CD, including factors both specific to CD and common to different forms of substance dependence [Tsuang et al., 1996, 1998; Kendler and Prescott, 1998; Gelernter et al., 2005a]. The heritability in a female twin population was estimated to be 0.39 for cocaine use and 0.65 for CD symptoms [Kendler and Prescott, 1998]. However, the genetics of CD are complex and there is substantial phenotypic variance in CD populations [Uhl et al., 1995; Gelernter et al., 2005b; Kranzler et al., 2008]. Thus, insights into the genetic etiology of CD are limited by both genetic heterogeneity and the complex clinical manifestations of the disorder [Kranzler et al., 2008; Roncero et al., 2012]. This suggests that the trait defined by the DSM-IV diagnosis of CD [American Psychiatric Association, 1994], which is based on a syndromal model of the disorder, may not be well suited to the identification of specific genetic variants contributing to CD risk.

Empirical subtyping methods are based on theories that emphasize the multifaceted nature of substance dependence and related behaviors [Basu et al., 2004]. Multivariate cluster analysis has been used commonly to subtype substance dependence [Cardoso et al., 2006; Gelernter et al., 2006; Kuo et al., 2008; Chan et al., 2011], including CD [Kranzler et al., 2008]. However, these approaches have focused exclusively on the cluster homogeneity of the clinical features, rather than aiming to optimize the heritability of the subtypes and enhance the potential to identify specific genes that contribute to subgroup membership.

In this study, we differentiated homogeneous, highly heritable clinical subtypes of CD to identify subtype-specific genetic variants. To accomplish this, we designed a multivariate clustering method using clinical features and symptoms of CD to assign cocaine-using subjects to subgroups. Using traits derived from the CD subgroups, we examined a panel of genetic variants for association, including both main effects and pairwise epistatic interactions.

## MATERIALS AND METHODS

### Subjects

A total of 9,965 identically evaluated subjects were aggregated from family-based and case-control genetic studies of DSM-IV CD, opioid dependence (OD), and alcohol dependence

(AD). Subjects were recruited at five sites: Yale University School of Medicine (N = 4,450, 44.7%), University of Connecticut Health Center (N = 3,698, 37.1%), University of Pennsylvania Perelman School of Medicine (N = 968, 9.7%), Medical University of South Carolina (N = 596, 6.0%), and McLean Hospital (N = 253, 2.5%). Subjects with a clinical diagnosis of schizophrenia, bipolar disorder, or gross cognitive impairment were excluded. The institutional review board at each site approved the study protocol and informed consent forms. The National Institute on Drug Abuse and the National Institute on Alcohol Abuse and Alcoholism each provided a Certificate of Confidentiality to protect participants. Subjects were paid for their participation.

The sample included 2,379 subjects from 1,009 small nuclear families (SNFs) and 7,586 unrelated individuals. Of the families, 747 (74%) had 2 members with a lifetime DSM-IV diagnosis of CD. In addition to the proband in each of the SNFs, there were 1,160 siblings, 169 parents, and 41 other family members. Additionally, 4,575 subjects with CD and/or OD and 3,011 controls (screened to exclude those with a lifetime substance use disorder) from case-control studies were included in the analysis. Pedigree information was obtained for all families in the study.

The self-reported population distribution of the sample was 38.7% European-American (EA), 45.1% African-American (AA), 8% Hispanic, and 8.2% Native American, Pacific Islander or members of other minority groups. The majority of the sample (56.4%) was never married; 25.6% was widowed, separated, or divorced; and 18.0% was married. Few subjects (4.5%) completed grade school only; 28.8% had some high school, but no diploma; 25.5% completed high school only; and 41.2% received education beyond high school.

## Assessments

Phenotypic information was obtained through administration of the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), a computer-assisted interview comprised of 26 sections (including one for cocaine use and related behaviors) that yields lifetime DSM-IV diagnoses of substance use and Axis I psychiatric disorders, as well as antisocial personality disorder [Pierucci-Lagha et al., 2005, 2007]. The test-retest and inter-rater reliabilities of the SSADDA DSM-IV diagnosis of CD were excellent, with  $\kappa = 0.92$  and  $0.83$ , respectively [Pierucci-Lagha et al., 2005].

The four most common lifetime DSM-IV diagnoses in the sample were CD (60.6%; 3,540 men and 2,500 women), nicotine dependence (54%), AD (47.5%), and OD (32.4%). Major depressive episode (MDE) was the most common psychiatric disorder (15.5%), followed by posttraumatic stress disorder (PTSD) (13.3%), antisocial personality disorder (ASPD) (12.3%), and pathological gambling (7.9%).

## Measures

The cocaine use and related behaviors section of the SSADDA contains 25 questions on age of onset, frequency, and intensity of cocaine use; route of cocaine administration; occurrence of psychosocial and medical consequences of cocaine use; attempts to quit cocaine use; and cocaine abuse treatment sought and received, resulting in 160 variables. In a previous study, in which we used a subset of the present sample, we identified 68 key

questions (see Supplementary Tables SI and SII) in this section that performed well for subtyping cocaine use and related behaviors [Kranzler et al., 2008]. Those features were based on their clinical utility as discriminators of cocaine use behavior subtypes and were used here to generate subtypes of cocaine use and related behaviors. Demographics and other substance use and psychiatric variables obtained from the SSADDA interview and estimated heritability served to test the validity of the clusters.

The majority (53) of the 68 key variables were categorical, with four possible response categories: “yes,” “no,” “obligate no,” and “missing.” The SSADDA skipped out of the cocaine section when the respondent reported never having used cocaine more than 10 times (lifetime), and subsequent categorical items in the cocaine section were scored as “obligate no” and subsequent continuous items as “obligate missing;” these form their own categories in our analysis. About 1.1% of the data on the 68 key variables were missing for the 9,965 subjects. Thus, there were 312 possible categories for the 68 key variables.

### Genotyping

DNA was extracted from immortalized cell lines, blood, or saliva. A total of 1,350 single nucleotide polymorphisms (SNPs) selected from 130 candidate genes and 186 ancestry informative markers (AIMs) were genotyped using the Illumina GoldenGate Assay platform (Illumina, Inc., San Diego, CA) for 3,443 subjects, a subset of the aggregate sample used in the phenotypic analysis. The AIMs were used to assign ancestry coefficients and thereby identify each subject’s primary ancestry. As described previously, the 130 candidate genes were selected on the basis of their roles in functional domains important in the addictions and in the commonly co-occurring phenotypes of anxiety and depressive disorders [Hodgkinson et al., 2008].

### Phenotypic Cluster Analysis

Our phenotypic analysis consisted of three consecutive components: data reduction, cluster analysis, and heritability estimation. As described in detail in Supplementary Materials, data on the 68 key clinical variables from 9,965 subjects were analyzed to derive subtypes of cocaine use and related behaviors. Briefly, we used multiple correspondence analysis (MCA) [Abdi and Valentin, 2007; Murtagh, 2007; LeRoux and Rouanet, 2009], a non-parametric method, to reduce the large number of variables to a limited number of dimensions. Cluster analysis was used to group similar subjects together based on the retained dimensions to create clusters of subjects. To estimate the heritability of each of the clusters, logistic regression was used to compute the likelihood of each subject’s membership in the cluster. Together with pedigree information, the log likelihood values for 9,436 EAs and AAs, including 2,268 individuals from 957 multi-member families, were analyzed using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) program [Almasy and Blangero, 1998] to estimate the heritability of the cluster-derived trait, with sex, age and race as covariates.

We identified five mutually exclusive clusters. As shown in Table I, the groups differed significantly on age, sex, race, education, and marital status. Specifically, Groups 2–5 (the cocaine use clusters) included significantly more men than women and were less educated

than Group 1 (the “non-cocaine use group”). Group 5 had the lowest level of education and the fewest married members and included a significantly higher proportion of EAs than Groups 2–4. Groups 3 and 4 included a significantly higher proportion of AAs than the other three groups.

Based on the clusters’ lifetime prevalence of substance use and psychiatric disorders (Table II) and cocaine-related features (Table III), we named Group 2 the “moderate cocaine use group,” Group 3 the “late-onset heavy cocaine use group,” Group 4 the “heavy cocaine use, infrequent intravenous injection group,” and Group 5 the “early-onset, heavy cocaine use, high comorbidity group.”

Group 1 was the largest, consisting of 3,370 subjects (56.5% women), none of whom had a diagnosis of CD. Group 1 had a mean of 3.8 (SD = 3.1) reported lifetime episodes of cocaine use and the lowest prevalence of all other substance dependence and psychiatric disorders (except ASPD, which was slightly lower in Group 3). The estimated heritability of the non-cocaine use group was 0.41 (SE = 0.06). Group 2, comprised of 1,241 subjects, had a significantly lower rate of CD (62.1%) and was less likely to have comorbid dependence on nicotine, alcohol, sedatives, or stimulants and psychiatric disorders (except ASPD) than groups 3–5. Of the cocaine-use groups, Group 2 had the lowest percentage of individuals reporting daily or almost daily cocaine use, negative effects due to cocaine use, and a history of cocaine treatment, but a higher rate of intravenous cocaine injection (31.3%) than Groups 3 (24.4%) or 4 (3.9%). Group 2 had the highest estimated heritability (0.69, SE = 0.05).

Groups 3–5 were heavy cocaine use groups. Of these, Group 3, the late-onset group, was the smallest (N = 180) and had the lowest rate of CD (86.7%), percentage of subjects with negative cocaine-related effects, and likelihood of cocaine treatment. This group had a significantly older mean age of first cocaine use (36.5 years, SD = 11.8) and heaviest cocaine use (45.1 years, SD = 5) than Groups 4 and 5. The heritability of Group 3 was only 0.07 (SE = 0.06), likely in part due to the small sample size. Group 4 was the largest cocaine use group, comprising 3,258 subjects, 98.4% of whom were diagnosed with CD. This group had a rate of intravenous cocaine injection that was much lower than all of the other cocaine use groups. The percentages of subjects in Group 4 that experienced negative effects due to cocaine use and that received treatment were intermediate between Groups 3 and 5. The heritability of Group 4 was 0.66 (SE = 0.05). The 1,916 subjects in Group 5, the early-onset, heavy cocaine use, high comorbidity group, had the highest prevalence of substance dependence and psychiatric disorders. This group reported the earliest onset of cocaine use (17.9 years, SD = 4.3) and of heaviest use (25.8 years, SD = 8.4), and endorsed a high rate of adverse effects of cocaine use (Table III). The heritability of Group 5 was 0.64 (SE = 0.05). Because of the high prevalence of CD, Groups 4 and 5 were chosen as the most clinically informative subtypes of CD and were compared with DSM-IV CD as traits for association analysis.

### Genetic Association Analysis

Of the 3,443 subjects genotyped, 612 were never exposed to cocaine. These individuals were classified as “unknown” rather than “unaffected” and were removed from subsequent analyses. To correct possibly unreliable self-reported sex, we calculated the chromosome X

heterozygosity for each subject and removed 31 subjects: 14 self-reported males with heterozygosity of more than 20% and 17 self-reported females with heterozygosity of less than 20%. Race was classified using STRUCTURE v2.3 [Pritchard et al., 2000] and AIMs, which stratified the remaining 2,800 subjects into two population groups: AAs and EAs. Of the 1,478 AAs, 93.78% had AA as their self-reported race. Of the 1,322 EAs, 89.71% had EA as their self-reported race. The sample included 253 self-reported Hispanics, among whom 82 reported AA origin (43 were in the African ancestry and 39 in the European ancestry clusters) and 171 reported EA origin (168 were in the European ancestry and 3 in the African ancestry clusters). We excluded other population groups from the association analysis. Association tests for AA and EA population groups were performed separately. Principal components analysis (PCA) was performed on the 186 AIMs for each population. The first PCA dimension was used in the subsequent association tests as a covariate to correct for the residual population structure. SNPs for which data were available for less than 95% of the subjects, or for which the  $P$  value for Hardy–Weinberg equilibrium was less than  $10^{-7}$ , were excluded from further analysis. The minor allele frequency (MAF) of each SNP was calculated within each population in each association test for different traits. SNPs with  $MAF < 5\%$  in a population were removed from the association tests for the trait in that population.

Association testing was performed for three traits: DSM-IV CD and the two heavy cocaine-use subtypes to identify SNPs that exerted a significant main effect or that, in pair-wise combinations, exerted a joint (epistatic) effect on the traits. GEE logistic regression was used in main effects tests to correct for correlations among related individuals in families. All SNP pairs were tested for association in epistatic analyses. We compared the findings of pair-wise joint effects detected using three methods: GEE logistic regression [SAS Institute, Inc., 2008], regular logistic regression [Purcell et al., 2007] and a bioinformatics method called BOOST [Wan et al., 2010]. For each SNP pair, BOOST estimates two logistic models: one uses the two SNPs as covariates; the other uses the two SNPs and their cross-product as covariates. An upper bound is calculated to approximate the ratio of the two logistic models, which aims to assess the additional effect that the cross-product contributes. The bound has been shown to be tight, and most nonsignificant interactions can be pruned. In all three models, age, sex and the first PCA dimension calculated from the AIMs were included as covariates. Permutation tests were performed to identify empirical thresholds to correct the  $P$ -value for multiple tests of main effects. Parallelized permutation tests were developed and conducted using a high performance computing facility with 120 processors to identify empirical thresholds and avoid inflating the  $P$ -value for epistatic effects.

## RESULTS

Removing the controls never exposed to cocaine from the genotyped subsample left 480 controls (187 AAs and 293 EAs). After data cleaning, there were 2,320 cases with a lifetime DSM-IV diagnosis of CD (1,291 AAs and 1,029 EAs). Of these, 779 AAs and 390 EAs were in Group 4 and 259 AAs and 436 EAs were in Group 5. Cases with data on all three phenotypes (DSM-IV CD, Groups 4 or 5) were tested separately for association against the same control samples, separately for AAs and EAs. SNPs were processed using the quality control steps for each individual case-control comparison, resulting in different numbers



tested for association. Table IV describes the subjects and SNPs used in the association analyses.

### Main Effects

Table V lists the SNPs that were associated with any of the three traits at  $P < 5E-04$ . Two statistically significant variants were identified in AAs. *HTR2C* (rs5988072) was associated with both DSM-IV CD ( $P_{\text{adj}} = 8.89E-03$ , Bonferroni corrected) and the early-onset, heavy cocaine use, high comorbidity group (Group 5) ( $P_{\text{adj}} = 4.99E-02$ , corrected by 1,000 permutation tests). A variant in *CLOCK* (rs11939815,  $P_{\text{adj}} = 4.27E-02$ ) was significantly associated with the heavy cocaine use, infrequent intravenous injection subtype (Group 4). Several genes contained nominally significant variants ( $P < 5E-04$ ), but none were significant using permutation to adjust for multiple comparisons. In AAs, these included *ALDH1A1* (rs11143429,  $P = 3.35E-04$ ), which was associated with DSM-IV CD; *CLOCK* (3 SNPs: rs3805155,  $P = 1.11E-04$ ; rs13116194,  $P = 2.11E-04$ ; rs6850524,  $P = 2.01E-04$ ) and *GLRA1* (rs991738,  $P = 3.02E-04$ ), which were associated with Group 4. In EAs, there was also a nominal association of a variant in *SLC18A2* (rs363256,  $P = 2.27E-04$ ) with this subtype. A variant in *OXT* (rs3761248,  $P = 3.04E-04$ ) was nominally associated with Group 5 in AAs. Except for the findings related to *HTR2C* and *ALDH1A1*, the  $P$  values for all other variants listed in Table V were smaller for the derived subtypes than for DSM-IV CD.

Based on the association in AAs of one of the SNPs in the *CLOCK* gene with membership in Group 4, we examined the linkage disequilibrium (LD) structure of the 9 genotyped *CLOCK* SNPs in the 1,478 AAs for which genotypes were available. These SNPs spanned 71,151 bp on chromosome 4. Figure 1 shows the LD relationships of the nine SNPs analyzed using Haploview 4.1 software. Rs11939815, the SNP that was significantly associated with membership in Group 4, had very high  $D'$  (normalized LD) with respect to two of the nominally significant SNPs from the *CLOCK* gene: rs6850524 and rs13116194, which were also in high LD with one another.

### Pairwise Epistatic Effects

As shown in Table VI, after Bonferroni correction, no SNP pair was significantly associated with any of the traits in either AAs or EAs. However, this conservative correction does not account for the correlation between SNP pairs, and likely resulted in an inflation of type II error. In permutation tests to determine an empirical significance threshold,  $>10^6$  pairs were tested for each randomly permuted phenotype based on the case-control ratios for the three traits: DSM-IV CD, Groups 4 and 5. We parallelized the standard permutation approach based on logistic regression (i.e., using PLINK) and distributed the process to 120 processors to identify empirical thresholds for significance. Table VI lists the SNP pairs with  $P_{\text{gee}} < 1E-05$  (GEE),  $P_{\text{lr}} < 1E-05$  (logistic regression using PLINK [Purcell, 2009]), or  $P_{\text{boost}} < 1E-04$  (BOOST) for at least one of the traits, which identifies candidates for possible subsequent validation. Five SNP pairs were identified in AAs, including one that was strongly associated with DSM-IV CD and Group 4 (heavy cocaine use with infrequent intravenous injection), one associated only with Group 4 and three that were associated with Group 5 (early-onset, heavy cocaine use, high comorbidity). The association of a SNP pair comprised of one variant each from *GRM1* and *OPRM1* with Group 5 was nominally

significant ( $P=7.6E-06$ ), but exceeded the empirical threshold ( $P < 7.6E-07$ ). Three pairs showed nominal significance in EAs as well, including two with Group 4 and one with Group 5. No SNP pairs were nominally associated with the DSM-IV diagnosis of CD. A SNP pair with variants from *CRHR2* and *BDNF* was nominally significant for Group 4 after multiple test correction ( $P=3.5E-06$ ), but it also exceeded the empirical threshold ( $P < 1.6E-6$ ). Despite the fact that none of the results reached statistical significance when corrected for multiple comparisons, the findings in Table VI show that the two derived subtypes provide more statistical power for association analysis than the DSM-IV diagnosis of CD.

## DISCUSSION

We compared the utility of three traits—a DSM-IV CD diagnosis and two highly heritable subtypes based on cocaine use and related behaviors—for association analysis of 130 candidate genes related to addictions in two American populations. We obtained the most significant findings in the AA population, where two SNPs exceeded the empirical significance threshold for a main effect after correction for multiple comparisons. Rs5988072 (in *HTR2C*) was significantly associated with both the CD diagnosis and the CD subtype of early-onset and heavy cocaine use, with high comorbidity. Rs11939815 (in *CLOCK*) was associated only with the CD subtype of heavy cocaine use with infrequent intravenous injection.

Because cocaine can increase the level of serotonin in the synapse by blocking reuptake via the serotonin transporter [Cami and Farre, 2003], genetic variants involved in the serotonergic system are candidates for CD risk [Saxon et al., 2005]. The 5-HT<sub>2C</sub> receptor, encoded by *HTR2C*, modulates the discriminative stimulus effects of cocaine [Callahan and Cunningham, 1995; Frankel and Cunningham, 2004]. Our finding of an association of *HTR2C* with both DSM-IV CD and one of the two subtypes we examined is, to our knowledge, the first report of such an association. The finding is consistent with the pharmacology of cocaine and, if replicated, points to differences in the pathophysiology of CD and AD, as studies of *HTR2C* have failed to show an association with AD [Herman and Balogh, 2012].

In addition, we found that *CLOCK* variation was associated with the CD subtype characterized by heavy cocaine use with infrequent intravenous injection. Because the finding was limited to that CD subtype, if replicated, it confirms the important role of phenotype refinement through multivariate cluster analysis in the identification of genetic factors for CD [Gelernter et al., 2005b; Kranzler et al., 2008; Sun et al., 2012]. The phenotypic heterogeneity in CD may partially explain why there has been no significant replicated association implicating circadian system regulation of cocaine sensitization and the reward system in animal models [Andretic et al., 1999; Abarca et al., 2002; Yuferov et al., 2005; Perreau-Lenz et al., 2007]. Of specific interest in this regard is the finding of increased cocaine reward and excitability of midbrain dopamine neurons in mice lacking a functional *CLOCK* gene [McClung et al., 2005]. In addition to the SNP in the *CLOCK* gene that reached significance (Table V), three other *CLOCK* variants were strongly associated with the CD subtype of heavy cocaine use with infrequent intravenous injection (Table V), consistent with a contribution of *CLOCK* variation to CD risk. Our previous association



study of a *CLOCK* variant with DSM-IV CD in a smaller sample was negative [Malison et al., 2006].

As a complex disorder, CD, like other forms of substance dependence, is the product of both environmental and genetic risk factors. Its genetic determinants likely include interactions among multiple neurochemical systems, including those involving dopamine, GABA, endogenous opioid, endogenous cannabinoid, and serotonin [Saxon et al., 2005]. Variability in these interactions due to genetic effects is a possible source of vulnerability to CD. We found several pairs of variants that approached significance for an association with CD and the two subtypes. An interaction between the genes encoding the beta 1 and beta 3 subunits of the GABA-A receptor was present in both AAs and EAs. Because GABA interneurons synapse on dopamine neurons to inhibit dopamine release, variation in the expression or function of GABA-A receptor subunits could alter the rewarding effects of cocaine [Spanagel and Weiss, 1999]. In AAs, the interaction of variants in *DDC* and *GABRG3* also approached significance for an association with the CD subtype of heavy cocaine use with infrequent intravenous injection, which could affect the dopamine-mediated reward produced by cocaine, both directly (through dopamine metabolism) and indirectly (through GABA interneurons).

In the present study, we employed a cluster analytic approach to derive subtypes of cocaine use and related behaviors that were more homogeneous than the group defined by DSM diagnosis in terms of their clinical features, including frequency and intensity of cocaine use, cocaine withdrawal symptoms, cocaine-related adverse effects and treatment history. We hypothesized that the subtypes would be superior to DSM-IV CD as traits for variant detection. Our association analysis provided support for the hypothesis. The association results with the two derived subtypes showed more significant *P* values for six of eight SNPs in the analysis of main effects (Table V) and seven of eight SNP pairs of epistatic effects (Table VI). Moreover, one variant, rs11939815 in *CLOCK*, was significantly associated only with Group 4 in AAs. An advantage of the subtyping approach to trait definition was also evident in the pair-wise epistatic effect tests, in which the most significant results were for variants associated with either of the two derived subtypes for all SNP pairs except rs4591574 and rs8041610.

The cluster analysis approach employed in the current study was refined from the one used in Kranzler et al. [2008], from which it differs in three major respects. First, in contrast to the approach used in Kranzler et al., the clustering process in the current analysis consisted of two layers of a cascaded process, as depicted in the supplemental material. This allowed us to identify subtypes that are both homogeneous and highly heritable. Second, in Kranzler et al., k-means cluster analysis was used to generate intermediate clusters. This method is an iterative procedure that is initialized with randomly chosen cluster centers. It is sensitive to outliers, with different initialization known to yield different clusters. In the present study, we used smart k-medoids cluster analysis, which is more robust to variation in the sample distribution. The supplementary material contains additional information about the cluster approach that we used in the present study. Third, the sample used here is much larger than the 1,393 subjects in Kranzler et al. and it yielded larger and more clearly differentiated clusters. Kranzler et al. found 6 clusters, each of which contained 42–350 subjects. The first

three were heavy cocaine user groups, but all six clusters (including the low-cocaine-use group) contained subjects with a diagnosis of CD. In contrast, the present analysis yielded a non-cocaine-user group in which no one had a diagnosis of CD. Subjects in Cluster 1 of Kranzler et al. (N = 336) were included in our Cluster 4 (N = 3,258 subjects), which was the largest user group. Subjects in Cluster 2 (N = 303) of Kranzler et al. were included in Cluster 5 in the current study, which contained 1,916 subjects. Cluster 3 in Kranzler et al. was split in to Clusters 3,4, and 5 in the current study. Thus, we believe that the larger sample and more sophisticated methodology in the present study yielded more homogeneous clusters, which are more useful for association analysis.

The study is limited by the lack of suitable samples for independent replication to validate these findings. The scoring functions for each of the two subtypes, which use cocaine use and related behaviors to yield a probabilistic membership score for each subject in each cluster, are available from the first author upon request to allow other investigators to use the Supplementary Material to attempt to replicate our findings for Groups 4 and 5. The permutation tests required to estimate empirically the significance threshold for type I error for the tests of epistatic effects were very computationally intensive. Advances in statistical methods are required to address this problem, which is often encountered in the examination of higher-order interactions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by NIH grants DA12849, DA12690, AA03510, AA11330, and AA13736 and the VISN 1 and VISN 4 Mental Illness Research, Education, and Clinical Centers (MIRECCs).

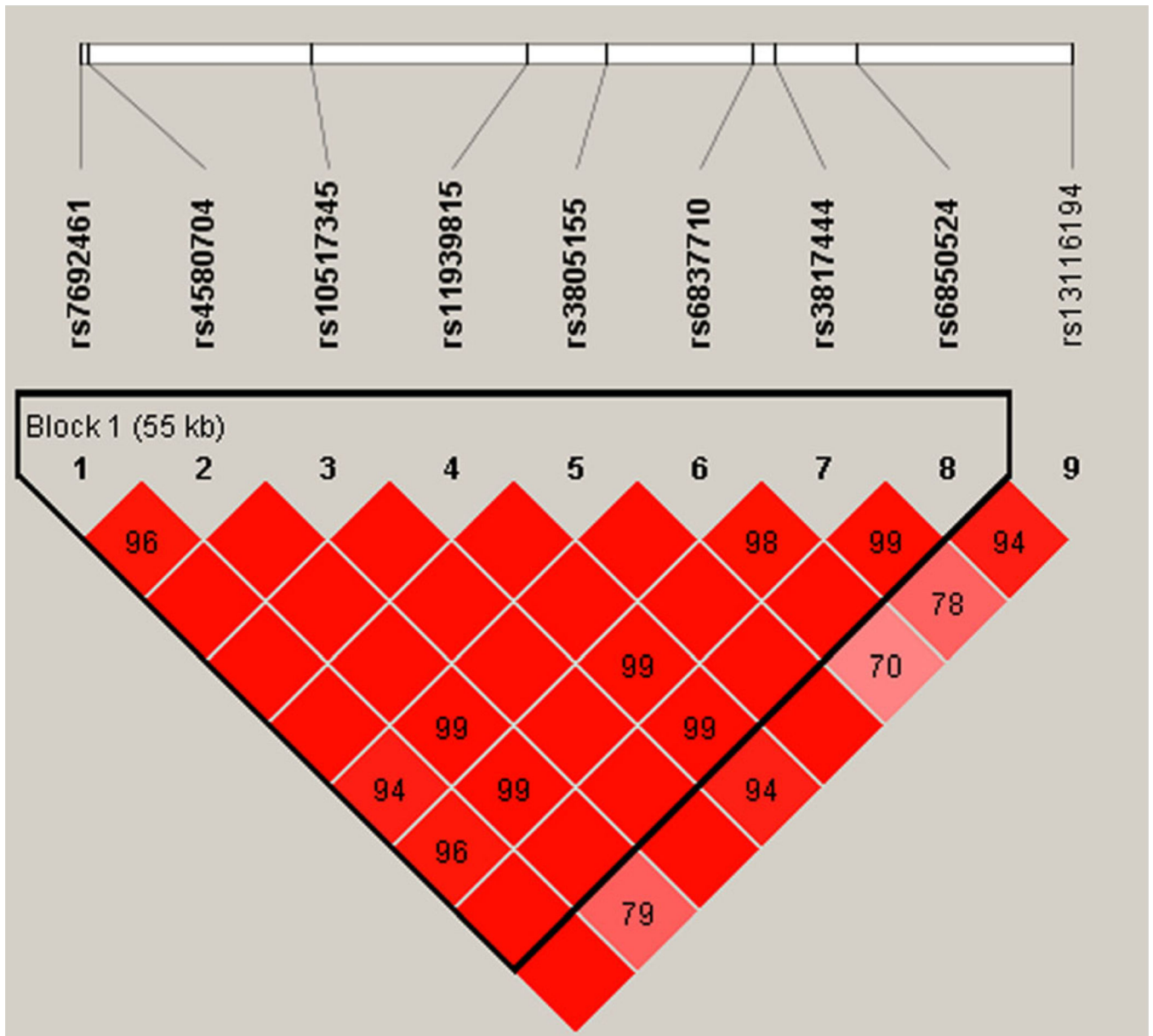
Dr. Kranzler has been a consultant or advisory board member for Alkermes, Lilly, Lundbeck, Pfizer, and Roche. He is also a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which is supported by Lilly, Lundbeck, Abbott, and Pfizer.

## REFERENCES

- Abarca C, Albrecht U, Spanagel R. Cocaine sensitization and reward are under the influence of circadian genes and rhythm. *Proc Natl Acad Sci USA*. 2002; 99(13):9026–9030. [PubMed: 12084940]
- Abdi, H.; Valentin, D. Multiple correspondence analysis. In: Salkind, N., editor. *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: SAGE; 2007.
- Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998; 62(5):1198–1211. [PubMed: 9545414]
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV*. Washington, DC: American Psychiatric Association; 1994.
- Andretic R, Chaney S, Hirsh J. Requirement of circadian genes for cocaine sensitization in *Drosophila*. *Science*. 1999; 285(5430):1066–1068. [PubMed: 10446052]
- Basu D, Ball SA, Feinn R, Gelernter J, Kranzler HR. Typologies of drug dependence: Comparative validity of a multivariate and four univariate models. *Drug Alcohol Depend*. 2004; 73(3):289–300. [PubMed: 15036551]
- Callahan PM, Cunningham KA. Modulation of the discriminative stimulus properties of cocaine by 5-HT1B and 5-HT2C receptors. *J Pharmacol Exp Ther*. 1995; 274(3):1414–1424. [PubMed: 7562516]

- Cami J, Farre M. Drug addiction. *N Engl J Med*. 2003; 349(10):975–986. [PubMed: 12954747]
- Cardoso JMN, Barbosa A, Ismail F, Pombo S. NETER alcoholic typology (NAT). *Alcohol Alcohol (Oxford, Oxfordshire)*. 2006; 41(2):133–139.
- Chan G, Gelernter J, Oslin D, Farrer L, Kranzler HR. Empirically derived subtypes of opioid use and related behaviors. *Addiction*. 2011; 106(6):1146–1154. [PubMed: 21306596]
- Frankel PS, Cunningham KA. m-Chlorophenylpiperazine (mCPP) modulates the discriminative stimulus effects of cocaine through actions at the 5-HT<sub>2C</sub> receptor. *Behav Neurosci*. 2004; 118(1):157–162. [PubMed: 14979792]
- Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, Poling J, Wilcox M, Farrer L, Kranzler HR. Genomewide linkage scan for cocaine dependence and related traits: Significant linkages for a cocaine-related trait and cocaine-induced paranoia. *Am J Med Genet Part B*. 2005a; 136B(1):45. [PubMed: 15909294]
- Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, Poling J, Wilcox M, Farrer L, Kranzler HR. Genomewide linkage scan for cocaine dependence and related traits: Significant linkages for a cocaine-related trait and cocaine-induced paranoia. *Am J Med Genet Part B*. 2005b; 136B(1):45–52. [PubMed: 15909294]
- Gelernter J, Panhuysen C, Wilcox M, Hesselbrock V, Rounsaville B, Poling J, Weiss R, Sonne S, Zhao H, Farrer L, et al. Genomewide linkage scan for opioid dependence and related traits. *Am J Hum Genet*. 2006; 78(5):759. [PubMed: 16642432]
- Herman AI, Balogh KN. Polymorphisms of the serotonin transporter and receptor genes: Susceptibility to substance abuse. *Subst Abuse Rehabil*. 2012; 3(1):49–157. [PubMed: 22933845]
- Hodgkinson CA, Yuan Q, Xu K, Shen PH, Heinz E, Lobos EA, Binder EB, Cubells J, Ehlers CL, Gelernter J, et al. Addictions biology: Haplotype-based analysis for 130 candidate genes on a single array. *Alcohol Alcohol*. 2008; 43(5):505–515. [PubMed: 18477577]
- 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]
- Kranzler HR, Wilcox M, Weiss RD, Brady K, Hesselbrock V, Rounsaville B, Farrer L, Gelernter J. The validity of cocaine dependence subtypes. *Addict Behav*. 2008; 33(1):41–53. [PubMed: 17582692]
- Kuo P-H, Aggen SH, Prescott CA, Kendler KS, Neale MC. Using a factor mixture modeling approach in alcohol dependence in a general population sample. *Drug Alcohol Depend*. 2008; 98(1):105–114. [PubMed: 18586414]
- LeRoux, B.; Rouanet, H. Multiple correspondence analysis. Los Angeles, CA: Sage; 2009.
- Malison RT, Kranzler HR, Yang B-Z, Gelernter J. Human clock, PER1 and PER2 polymorphisms: Lack of association with cocaine dependence susceptibility and cocaine-induced paranoia. *Psychiatr Genet*. 2006; 16(6):245–249. [PubMed: 17106427]
- McClung CA, Sidiropoulou K, Vitaterna M, Takahashi JS, White FJ, Cooper DC, Nestler EJ. Regulation of dopaminergic transmission and cocaine reward by the Clock gene. *Proc Natl Acad Sci USA*. 2005; 102(26):9377–9381. [PubMed: 15967985]
- Murtagh F. Multiple correspondence analysis and related methods. *Psychometrika*. 2007; 72(2):275–277.
- Perreau-Lenz S, Zghoul T, Spanagel R. Clock genes running amok. Clock genes and their role in drug addiction and depression. *EMBO Rep*. 2007; 8:S20–S23. [PubMed: 17726437]
- Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, Farrer L, Kranzler HR. Diagnostic reliability of the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend*. 2005; 80(3):303–312. [PubMed: 15896927]
- Pierucci-Lagha A, Gelernter J, Chan G, Arias A, Cubells JF, Farrer L, Kranzler HR. Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend*. 2007; 91(1):85–90. [PubMed: 17590536]
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155(2):945–959. [PubMed: 10835412]
- Purcell, S. Package: PLINK v1.07. 2009. <http://pngumghharvardedu/purcell/plink>

- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, de Bakker P, Daly MJ, et al. PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet.* 2007; 81(3):559–575. [PubMed: 17701901]
- Roncero C, Ros-Cucurull E, Daigre C, Casas M. Prevalence and risk factors of psychotic symptoms in cocaine-dependent patients. *Actas Esp Psiquiatr.* 2012; 40(4):187–197. [PubMed: 22851479]
- SAS Institute, Inc. SAS(r) 9.2 enhanced logging facilities. Cary, NC: SAS Institute, Inc; 2008.
- Saxon AJ, Oreskovich MR, Brkanac Z. Genetic determinants of addiction to opioids and cocaine. *Harv Rev Psychiatry.* 2005; 13(4):218–232. [PubMed: 16126608]
- Spanagel R, Weiss F. The dopamine hypothesis of reward: Past and current status. *Trends Neurosci.* 1999; 22(11):521–527. [PubMed: 10529820]
- Substance Abuse and Mental Health Services Administration. NSDUH Series H-41, HHS Publication No (SMA) 11-4658. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2011. Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings.
- Sun J, Bi J, Chan G, Oslin D, Farrer L, Gelernter J, Kranzler HR. Improved methods to identify stable, highly heritable subtypes of opioid use and related behaviors. *Addict Behav.* 2012; 37(10):1138–1144. [PubMed: 22694982]
- Tsuang MT, Lyons MJ, Eisen SA, Goldberg J, True W, Lin N, Meyer JM, Toomey R, Faraone SV, Eaves L. Genetic influences on DSM-III-R drug abuse and dependence: A study of 3,372 twin pairs. *American Journal of Medical Genetics Part B, Neuropsychiatr Genet.* 1996; 67(5):473–177.
- Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L. Co-occurrence of abuse of different drugs in men: The role of drug-specific and shared vulnerabilities. *Arch Gen Psychiatry.* 1998; 55(11):967–972. [PubMed: 9819064]
- Uhl, GR.; Elmer, GI.; LaBuda, MC.; Pickens, RW. Genetic influences in drug abuse. New York: Raven Press; 1995.
- Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NL, Yu W. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet.* 2010; 87(3):325–340. [PubMed: 20817139]
- Yuferov V, Butelman ER, Kreek MJ. Biological clock: Biological clocks may modulate drug addiction. *Eur J Hum Genet.* 2005; 13(10):1101–1103. [PubMed: 16094306]



**FIG. 1.**

LD structure of nine SNPs spanning the CLOCK gene in the AA genotyped sample. Shades of pink/red represent  $D' < 1$  and  $L_{OD} \geq 2$  with the value of  $D'$  written in the boxes, bright red represents  $D' = 1$  and  $L_{OD} \geq 2$ .

TABLE I

Demographic Characteristics by Group [N (%)]

Characteristic	Group 1, 3,370 (33.8)	Group 2, 1,241 (12.5)	Group 3, 180 (1.81)	Group 4, 3,258 (32.7)	Group 5, 1,916 (19.2)	$\chi^2$ (df)
Age	39.5 (14.2)	3971 (10.0)	52.39 (6.4)	40.89 (8.1)	39.20 (9.5)	591.8 (4)
Sex [N (%)]						262.7 (4)
Men	1,466 (43.5)	750 (60.4)	104 (57.8)	1,816 (55.7)	1,226 (64.0)	
Women	1,904 (56.5)	491 (39.6)	76 (42.2)	1,442 (44.3)	690 (36.0)	
Race [N (%)]						638.7 (12)
AA	1,303 (38.7)	549 (44.2)	94 (52.2)	1,969 (60.4)	580 (30.3)	
EA	1,612 (47.8)	490 (39.5)	66 (36.7)	793 (24.3)	897 (46.8)	
Hispanic	201 (6.0)	107 (8.6)	7 (3.9)	243 (7.5)	240 (12.5)	
Other	253 (7.5)	94 (7.6)	13 (7.2)	253 (7.8)	199 (10.4)	
Education [N (%)]						1,495.4 (12)
No HS	60 (1.8)	73 (5.9)	5 (2.8)	138 (4.2)	168 (8.8)	
Some HS	382 (11.3)	432 (34.8)	51 (28.3)	1,172 (36.0)	832 (43.4)	
HS graduate	656 (19.5)	397 (32.0)	63 (35.0)	950 (29.2)	472 (24.6)	
Beyond HS	2,271 (67.4)	338 (27.2)	61 (33.9)	997 (30.6)	442 (23.1)	
Marital status [N (%)]						634.5 (8)
Never married	1,731 (51.4)	789 (63.6)	47 (26.1)	1,920 (58.9)	1,135 (59.2)	
Married	1,025 (30.4)	144 (11.6)	20 (11.1)	398 (12.2)	210 (11.0)	
Div/Sep/Wid	614 (18.2)	308 (24.9)	113 (62.8)	940 (28.9)	571 (29.8)	



**TABLE II**

Lifetime Prevalence of Substance Use and Psychiatric Disorders by Group

<b>Disorders</b>	<b>Group 1, 3,370 (33.8)</b>	<b>Group 2, 1,241 (12.5)</b>	<b>Group 3, 180 (1.81)</b>	<b>Group 4, 3,258 (32.7)</b>	<b>Group 5, 1,916 (19.2)</b>	<b><math>\chi^2</math> (df = 4)</b>
<b>Substance use disorders</b>						
Cocaine dependence	0 (0.0)	770 (62.1)	156 (86.7)	3,207 (98.4)	1,907 (99.5)	788.4
Nicotine dependence	663 (19.7)	707 (57.0)	112 (62.2)	2,331 (71.6)	1,568 (81.8)	2,164.8
Alcohol dependence	765 (22.7)	537 (43.3)	105 (58.3)	2,042 (62.7)	1,288 (67.2)	1,361.6
Opioid dependence	274 (8.1)	559 (45.0)	65 (36.1)	946 (29.0)	1,389 (72.5)	1,723.1
Sedative dependence	32 (1.0)	64 (5.2)	12 (6.7)	190 (5.8)	301 (15.7)	313.5
Stimulant dependence	15 (0.5)	43 (3.5)	13 (7.2)	167 (5.1)	216 (11.3)	203.5
Other substance dependence	66 (2.0)	160 (12.9)	18 (10.0)	340 (10.4)	747 (39.0)	1,019.6
<b>Psychiatric disorders</b>						
ASPD	177 (5.3)	135 (10.9)	9 (5.0)	484 (14.9)	423 (22.1)	316.9
MDE	400 (11.9)	148 (11.9)	37 (20.6)	560 (17.2)	395 (20.6)	92.2
PTSD	195 (5.8)	115 (9.3)	30 (16.7)	576 (17.7)	413 (21.6)	323.0
OCD	35 (1.0)	18 (1.5)	3 (1.7)	93 (2.9)	74 (3.9)	49.5
Social phobia	89 (2.6)	42 (3.4)	12 (6.7)	148 (4.5)	130 (6.8)	54.9
Agoraphobia	87 (2.6)	41 (3.3)	15 (8.3)	195 (6.0)	161 (8.4)	97.7
Panic disorder	86 (2.6)	45 (3.6)	16 (8.9)	183 (5.6)	201 (10.5)	148.4
Compulsive gambling	95 (2.8)	72 (5.8)	19 (10.6)	364 (11.2)	236 (12.3)	198.2

TABLE III

Cocaine Use Characteristics, Cocaine-Related Effects, and Cocaine Treatment History by Group [N (%)]

Behaviors	Group 2 1,241 (12.5)	Group 3 180 (1.81)	Group 4 3,258 (32.7)	Group 5 1,916 (19.2)	$\chi^2$ (df=3)
Mean [SD] age of first cocaine use in year	21.1 (5.7)	36.5 (11.8)	21.4 (5.2)	17.9 (4.3)	1,057.8
Mean [SD] age of onset of heaviest cocaine use in year	27.2 (8.2)	45.1 (5.0)	29.3 (7.4)	25.8 (8.4)	2,158.0
Used cocaine daily or almost daily	861 (69.4)	157 (87.2)	3,026 (92.9)	1,857 (96.9)	540.8
Injected cocaine intravenously	388 (31.3)	44 (24.4)	129 (3.9)	1,417 (74.0)	1,804.4
Stayed high from cocaine for a whole day or more	594 (47.9)	131 (72.8)	2,628 (80.7)	1,623 (84.7)	576.2
Strong desire for cocaine made it hard to think of anything else	404 (32.6)	118 (65.6)	2,532 (77.7)	1,619 (84.5)	933.9
Cocaine interfered with work, school, or home life	318 (25.6)	111 (61.7)	2,475 (76.0)	1,650 (86.1)	1,132.4
Family members, friends, doctor, clergy, boss, Or people at work or school objected to cocaine use	475 (38.3)	108 (60.0)	2,683 (82.4)	1,664 (86.9)	961.2
Been arrested or had trouble with the police because Of cocaine use	310 (25.0)	61 (33.9)	1,776 (54.5)	1,233 (64.4)	465.6
Gave up or greatly reduced important activities due to cocaine use	419 (33.8)	131 (72.8)	2,853 (87.6)	1,775 (92.6)	1,418.3
Ever treated for a cocaine-related problem	432 (34.8)	109 (60.6)	2,372 (72.8)	1,421 (74.2)	585.6
Ever attended self-help group for cocaine use	363 (29.3)	90 (50.0)	2,186 (67.1)	1,369 (71.5)	601.2

TABLE IV

Summary of the Subjects and Single Nucleotide Polymorphisms (SNPs) Used in the Association Tests

Phenotype	Cases			Controls			No. of SNPs		
	Total no.	Male (%)	Average age	Total no.	Male (%)	Average age	Missing <sup>a</sup>	MAF <sup>b</sup>	Remained in the analysis
AA									
CD diagnosis	1,291	49.6	41.1					212	1,106
Group 4	779	45.7	40.4	187	56.7	42.4	32	215	1,103
Group 5	259	61.4	42.0					210	1,108
EA									
CD diagnosis	1,029	58.0	37.5					346	971
Group 4	390	52.7	37.7	293	62.1	36.3	33	350	967
Group 5	436	63.1	36.9					352	965

<sup>a</sup> SNPs were removed because of missing values on > 5% of subjects.

<sup>b</sup> SNPs were removed because of minor allele frequency [MAF] < 5%.

TABLE V

Top Findings From Main Effect Tests

Chr	Position (bp)	SNPID	Gene	CD diagnosis			Group 4			Group 5			
				RAF	Coef.	P-Value	RAF	Coef	P-Value	RAF	Coef.	P-Value	
AA													
X	113,890,903	rs5988072	HTR2C	0.883	-0.569	<b>8.04E-06<sup>a,b</sup></b>	0.868	-0.491	2.42E-04	0.869	-0.714	<b>5.80E-05<sup>c</sup></b>	
9	75,607,884	rs1143429	ALDH1A1	0.802	-0.461	3.35E-04 <sup>b</sup>	0.795	-0.440	2.34E-03	0.783	-0.567	1.53E-03	
4	56,358,283	rs1939815	CLOCK	0.372	0.404	4.00E-04	0.363	0.503	<b>3.87E-05<sup>a,b</sup></b>	0.420	0.308	2.52E-02	
5	151,298,449	rs991738	GLRA1	0.555	-0.391	5.72E-04	0.553	-0.421	3.02E-04 <sup>b</sup>	0.518	-0.302	3.59E-02	
4	56,364,009	rs3805155	CLOCK	0.799	-0.438	6.05E-04	0.801	-0.520	1.11E-04 <sup>b</sup>	0.767	-0.335	3.17E-02	
4	56,397,317	rs13116194	CLOCK	0.792	-0.411	7.57E-04	0.793	-0.472	2.11E-04 <sup>b</sup>	0.760	-0.325	3.18E-02	
4	56,381,997	rs6850524	CLOCK	0.641	-0.372	1.37E-03	0.649	-0.460	2.01E-04 <sup>b</sup>	0.599	-0.290	3.42E-02	
20	3,050,393	rs3761248	OXT	0.908	0.717	3.27E-03	0.913	0.709	5.19E-03	0.909	1.010	3.04E-04 <sup>b</sup>	
EA													
10	119,024,225	rs363256	SLC18A2	0.896	-0.391	7.68E-03	0.899	-0.646	2.27E-04 <sup>b</sup>	0.872	-0.364	2.86E-02	

RAF, reference allele frequency.

<sup>a</sup>Significant at  $P < 0.05$  after Bonferroni correction for multiple tests.<sup>b</sup>The smallest  $P$ -value obtained among the three traits.<sup>c</sup>Significant at  $P < 0.05$  after correction for multiple tests by permutation (empirical threshold  $P < 5.9E-05$ ).

TABLE VI

Top Findings From Pairwise Epistatic Effect Tests

Chr	SNP1			SNP2			GEE P-Value			PLINK P-Value			BOOST P-Value			
	Position (bp)	SNP ID	Gene	Chr	Position (bp)	SNP ID	Gene	CD diagnosis	Group 4	Group 5	CD diagnosis	Group 4	Group 5	CD diagnosis	Group 4	Group 5
AA																
4	47,423,595	rs4591574	GABRB1	15	27,008,264	rs8041610	GABRB3	3.2E-6	6.5E-6	1.9E-4	6.0E-7	2.0E-6	5.4E-5	3.0E-5	7.1E-5	1.5E-3
7	50,620,275	rs2229341	DOC	15	27,514,749	rs751085	GABRG3	5.8E-5	1.4E-6	3.6E-2	1.8E-4	3.7E-6	1.6E-3	4.5E-3	8.4E-5	2.2E-2
8	54,140,999	rs3802280	OPRK1	11	113,803,666	rs17116138	HTR3B	6.8E-4	1.3E-2	4.7E-6	2.3E-4	6.3E-4	2.5E-5	4.8E-3	1.6E-2	3.5E-5
8	54,140,999	rs3802280	OPRK1	11	113,801,668	rs17116121	HTR3B	7.9E-4	1.3E-2	6.6E-6	2.8E-4	5.7E-4	3.3E-5	2.3E-3	1.1E-2	5.5E-5
6	146,446,024	rs9373486	GRM1	6	154,445,215	rs10485058	OPRM1	7.2E-4	8.5E-3	7.5E-6	1.8E-4	1.3E-4	7.6E-6	8.4E-4	3.4E-2	4.9E-6
EA																
7	30,713,608	rs2284217	CRHR2	11	27,744,859	rs12273363	BDNF	2.2E-5	2.8E-6	5.0E-3	1.5E-5	3.5E-6	1.4E-3	6.7E-4	9.9E-5	3.2E-2
4	47,051,185	rs4315750	GABRB1	15	27,048,768	rs4S17759	GABRB3	1.4E-4	6.6E-6	4.2E-2	1.4E-4	8.7E-6	3.5E-4	3.1E-3	9.8E-5	1.0E-2
2	171,682,740	rs3828275	GADI1	17	43,894,798	rs81189	CRHRI	2.8E-5	4.4E-3	7.6E-6	4.0E-5	1.2E-3	3.1E-6	1.7E-3	2.9E-2	1.0E-4

Bold font numbers are the smallest P values among the three traits for each of the three methods. All numbers were rounded to the first decimal place in the mantissa of the scientific number notation.