Published in final edited form as:
Psychiatr Genet. 2010 October ; 20(5): . doi:10.1097/YPG.0b013e32833a2106.

Genome-wide association study identifies genes that may contribute to risk for developing heroin addiction

David A. Nielsen[a,b], **Fei Ji**[c], **Vadim Yuferov**[a], **Ann Ho**[a], **Chunsheng He**[c], **Jurg Ott**[c,d], and **Mary Jeanne Kreek**[a]

[a]Laboratory of the Biology of Addictive Diseases, The Rockefeller University, New York, NY

[b]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, and Michael E. DeBakey V.A. Medical Center, Houston, Texas, Houston, TX

[c]Laboratory of Statistical Genetics, The Rockefeller University, New York, NY, USA

[d]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Abstract

Objectives—We have used genome-wide association studies to identify variants that are associated with vulnerability to develop heroin addiction.

Methods—DNA from 325 methadone stabilized, former severe heroin addicts and 250 control subjects were pooled by ethnicity (Caucasian and African American) and analyzed using the Affymetrix GeneChip Mapping 100K Set. Genome-wide association tests were conducted.

Results—The strongest association with vulnerability to develop heroin addiction, with experiment-wise significance ($P = 0.035$), was found in Caucasians with the variant rs10494334, a variant in an unannotated region of the genome (1q23.3). In African Americans, the variant most significantly associated with heroin addiction vulnerability was rs950302, found in the cytosolic dual specificity phosphatase 27 gene *DUSP27* (point-wise $P = 0.0079$). Furthermore, analysis of the top 500 variants with the most significant associations (point-wise $P \leq 0.0036$) in Caucasians showed that three of these variants are clustered in the regulating synaptic membrane exocytosis protein 2 gene *RIMS2*. Of the top 500 variants in African Americans (point-wise $P \leq 0.0238$), three variants are in the cardiomyopathy associated 3 gene *CMYA3*.

Conclusions—This study identifies new genes and variants that may increase an individual's vulnerability to develop heroin addiction.

Keywords

Addiction; gene; heroin; microarray; polymorphism

Introduction

Factors that contribute to the development of heroin addiction, besides the drug of abuse itself, are both genetic and environmental. Epidemiological studies indicate that genetic

Correspondence and requests for reprints to: M. J. Kreek: Laboratory of the Biology of Addictive Diseases, The Rockefeller University, Box 171, 1230 York Avenue, New York, NY 10065, Tel: 212.327.8490, Fax: 212.327.8574, kreek@rockefeller.edu.

factors, both those shared by addiction to several drugs of abuse and those specific for addiction to a given drug, contribute approximately 40-60% of the risk of developing heroin addiction (Kendler et al., 2003, Tsuang et al., 1998, Tsuang et al., 1996). One study found 38% genetic variance specific to the development of heroin addiction, a variance higher than that for marijuana, stimulants, sedatives, or psychedelics (Tsuang et al., 1998). Genetic predisposition may explain why approximately one-third of those self-exposed to opiates become addicted (Kreek, 2002).

Recently, using the Affymetrix 10K GeneChip, we found a genotype pattern of three unlinked alleles that was associated with developing heroin addiction and another pattern that was associated with protection from developing heroin addiction (Nielsen et al., 2008b). That study suggested a role for five genes in the risk of developing heroin addiction including the gene coding for the μ opioid receptor.

An alternative method of genotyping large cohorts is to analyze pools consisting of multiple DNA samples (reviewed in (Sham et al., 2002)), which was developed to reduce genotyping costs. Analysis of pooled samples still allows a comparison of allele frequencies between cases and controls. Pools are made from cases or control subjects, either as one pool containing the case samples and the other containing the controls, or as multiple pools of each category of subjects. This technique has been used successfully to find differences in allele frequencies using microarrays (Kirov et al., 2006, Liu et al., 2006, Liu et al., 2005, Johnson et al., 2006, Uhl et al., 2007, Uhl et al., 2008).

Herein, we have used a pooled sample approach with the Affymetrix Gene Mapping 100K Set to identify variants and genes that may be associated with vulnerability to develop heroin addiction in a large cohort of volunteers ascertained in New York City. The subjects for this genome-wide association study were former severe heroin addicts or control subjects, with extensively defined phenotypic characterization.

Materials and methods

Subjects and phenotyping

Two hundred and fifty control subjects and 325 former severe heroin addicts, who were consecutive volunteers in genetic studies in our laboratory at The Rockefeller University and who met the inclusion criteria defined below were included in this study. They were recruited from advertisements and from clinical resources in New York City. An informed consent approved by The Rockefeller University Hospital Institutional Review Board that gave specific consent for genetic studies was signed by all subjects. Based on the ethnic/cultural background of the subjects, their parents, grandparents, and great-grandparents, subjects were classified as Caucasian (n = 350) or African American (n = 225).

The Addiction Severity Index (ASI) (McLellan et al., 1980) was administered to and urine analyses were conducted for multiple drugs of abuse in all subjects. Former severe heroin addicts met Federal guidelines for methadone maintenance treatment (one year or more of daily multiple injections of heroin or other opiates) (Rettig and Yarmolinsky, 1995). Control subjects had no history of alcohol or illicit drug use (illicit drug use or drinking to intoxication three or more times per week for 6 months or more), no use of cannabis (three or more times per week) for more than 4 years, and no current alcohol or illicit drug use (one or more instance of drinking to intoxication or any illicit drug use (except for possible cannabis use up to 12 days) in the last 30 days).

Sample Pooling

Each pool was made in duplicate from the DNA of 25 subjects. Genomic DNA from each subject was diluted to approximately 10 μg/ml and the concentration determined using the Rediplate 96 PicoGreen assay (Invitrogen, Carlsbad CA). For each pool, 50 ng DNA of each subject was used, and the final pooled DNA was diluted to 300 μl with TE (0.1 mM EDTA, 10 mM Tris, pH 8.0). Each duplicate pool was made separately. There were six pools from Caucasian controls, eight pools from Caucasian former heroin addicts (cases), four pools from African American controls, and five pools from African American cases (Table 1).

Microarrays

Two separate hybridizations were performed for each of the DNA pools using the Affymetrix 50K Hind 240 and the 50K Xba 240 arrays (Affymetrix, Santa Clara, CA). Five hundred ng DNA (in 120 μl) from each pool was ethanol precipitated, resuspended in 10 μl TE, and processed according to manufacturer's protocol (Affymetrix). Processed arrays were scanned using an Affymetrix GeneChip Scanner 3000. Data was acquired using the GeneChip Operating System (GCOS) and analyzed using the GTYPE genotype Analysis Software (Affymetrix). Variant annotations are available from the NetAffx Analysis Center (http://www.affymetrix.com/analysis/index.affx).

Data analysis

Separate analyses of Caucasians and African Americans were conducted to avoid errors due to population stratification. For each DNA microarray (one of two duplicate pools), a "background" value was calculated as the average fluorescence intensity from the 5% of cells with the lowest values; the "background" value was then subtracted from the intensity value of each cell. The background subtracted intensity values were then normalized by dividing by a "ceiling" value, which was the mean intensity of the 5% of cells with the highest values. The normalized A and B probe intensities were averaged from ten "perfect match" intensity values for each A and B probe. The methods are outlined in Johnson, *et al.* (Johnson et al., 2006). The ratio of averaged and normalized A probe intensity to the sum of averaged and normalized A plus B probe intensity was determined for each variant. The pooled A allele frequency was the average ratio from the duplicate arrays, and this value was used below.

The GeneChip Mapping 100K Set contains 116,204 variants. Since our pools contained both males and females, 2,361 X chromosomal variants could not be evaluated. No Y chromosomal variants are represented on the 100K Set. Hence, analyses were performed on autosomal variants only. In addition, the 100K Set contains 644 variants with no annotation. After exclusion due to low allele frequency (< 0.03 within a single ethnic group), 113,135 variants in the Caucasians and 113,174 variants in the African Americans were evaluated further.

To test for differences in allele frequency between the cases and controls for each of the variants within each ethnic group, a two sample non-parametric t-test was conducted. Multivariate permutation (Simon et al., 2004) was used to correct for multiple testing, and experiment-wise P values are reported. To perform permutation testing for the experiment-wise P value, the class labels were permuted and the t statistic values for each of the markers were recalculated. The maximum t statistic (corresponding to minimum P value) of all ~110,000 tests (one test for each marker) from this permutation was taken. This procedure was repeated for 3,003 permutations of the data in the Caucasian group. The value 3,003 was obtained by selecting 6 controls (or 8 cases) out of 14 pooled samples. The originally observed t statistic was compared to the distribution of t statistics composed of 3003 maximum t statistics to obtain the experiment-wise P value. For example, an experiment-

wise *P* value of 0.035 means that 105 out of 3,003 permuted maximum *t* statistics were higher than or equal to our observed *t* statistic. The *P* value obtained by this method is called the multivariate *P* value.

There is a high degree of correlation between many of the variants in a genome-wide association study due to linkage disequilibrium across the chromosomes. If we corrected for multiple testing using Bonferroni or False Discovery Rate (FDR), we would be discounting the correlation between the markers and over correcting our *P* value. Permutation testing allows us to maintain the correlation structure between the variants and calculate a global cut-off where any *P* values that are smaller than that observed value will have an experiment-wise significance. This approach allows the correlation among variants and is therefore less conservative than the Bonferroni approach. *P* values are reported for point-wise (nominal) significance and experiment-wise (corrected) significance.

Variant analysis

The finding of a significant association between a variant and a phenotype may be due several factors. The variant itself may modify function by altering the coding sequence of the gene, the stability of the resulting mRNA (Duan, Wainwright et al. 2003), or the regulation of gene expression. Regulatory variants may be found far upstream of genes. For example, a number of variants have been identified upstream of the *SOX9* gene which are associated with the palatal lesion Pierre Robin sequence (Benko, Fantes et al. 2009). One variant is located 1.44 million nucleotides upstream of *SOX9* and alters several predicted transcription binding sites. Other examples include two variants found upstream of the *SHH* gene. One is located one million nucleotides upstream of the *SHH* gene (Lettice, Heaney et al. 2003) and was found to be associated with preaxial polydactyly, while the other is located 470,000 nucleotides upstream and was associated with holoprosencephaly (Jeong, Leskow et al. 2008). Using 11,446 genes in a Bayesian hierarchical model, the Pritchard group found that 5% of the quantitative trait loci for gene expression (eQTLs) were located more than 20,000 nucleotides upstream of the transcription start sites (Veyrieras, Kudaravalli et al. 2008).

Significant associations may also be due to the variant being in linkage disequilibrium with a functional variant. While linkage disequilibrium (LD) decreases with increasing distance between markers, studies of some genes have shown that LD may be quite high past 100,000 nucleotides (Collins et al., 1999, Reich et al., 2001). In this study, if an annotated gene was found within 100,000 nucleotides of a variant, the gene's location relative to that variant is indicated. Mammalian conservation was determined using the "Vertebrate Multiz Alignment & PhastCons Conservation (28 species)" and the "Evolutionary and Sequence Pattern Extraction through Reduced Representation" (ESPERR) (King et al., 2005) to evaluate predicted regulatory potential as implemented in the UCSC Genome Browser (March 2006 assembly, http://gemome.ucsc.edu/). The "Transcription Element Search System" (TESS) was used to scan sequences for predicted transcription factor binding sites (Schug and Overton, 1977).

Results

In this study, we analyzed duplicate pooled DNA samples using the GeneChip Mapping 100K Set. Pools were formed by combining an equal amount of DNA from 25 subjects (Table 1). A total of 200 Caucasian former heroin addicts (cases), 150 Caucasian controls, 125 African American cases, and 100 African American controls were used to make the pools. More than 113,000 variants were examined in each ethnicity (see Data analysis).

Association analyses

The ten variants with the smallest experiment-wise *P* values of association of allele frequency with heroin addiction in Caucasians and African Americans are listed in Tables 2 and 3, respectively. The variant rs10494334 (located at 1q23.3) had the smallest point-wise *P* value ($P = 0.0003$) for association of allele frequency with heroin addiction in the Caucasian group, and this association was significant experiment-wise when corrected for multiple testing ($P = 0.035$) (Table 2). An in-depth analysis of this intergenic variant and surrounding sequence revealed no annotated gene within 100,000 nucleotides or any indication of this variant altering known or predicted function. Variant rs2323218 was ranked second by ascending *P* value and is located 8,989 nucleotides downstream of the *T* gene, which codes for an embryonic nuclear transcription factor (point-wise $P = 0.0003$) (Edwards et al., 1996).

In African Americans, the variant with the smallest *P* value for association of allele frequency with heroin addiction was variant rs950302 (point-wise $P = 0.0079$, Table 3). This variant is in the second intron of the dual specificity phosphatase 27 (putative) gene *DUSP27*. This gene encodes a recently identified member of the cytosolic dual specificity phosphatase family, which may be involved in energy metabolism (Friedberg et al., 2007).

Chromosomal regions with multiple variants

We searched for chromosomal regions that contained at least three variants associated with heroin addiction that were in close proximity with each other. Variants were sorted by ascending *P* value in each ethnic group (Caucasian and African American) and the 500 variants with the smallest *P* value (point-wise $P \leq 0.0037$ in the Caucasians and $P \leq 0.0238$ in the African Americans) were evaluated to determine if any set of at least three variants were within 100,000 nucleotides of each other (Table 4, point-wise $P \leq 0.0020$ for Caucasians and point-wise $P \leq 0.0079$ for African Americans). In Caucasians, three variants from this list, spanning 14,000 nucleotides, are found in an intron of the regulating synaptic membrane exocytosis protein 2 gene *RIMS2*. The cardiomyopathy associated 3 gene *CMYA3* in African Americans has three variants in a single intron that spans 3,000 nucleotides; another set of three variants are located at chromosome 13q13.1, spanning a 9,000 nucleotide region with no known function.

Candidate gene association

In another approach, we examined 240 genes that may play a role in the vulnerability to develop addiction (Nielsen et al., 2008b). On the 100K GeneChip, 2,081 variants were within 100,000 nucleotides of 167 of these genes (see Supplementary Material, Table S1). We also examined 153 other genes from a list of the Gershon laboratory (Hattori et al., 2005). One thousand one hundred fifty-nine variants are within 100,000 nucleotides of 111 of these genes (see Supplementary Material, Table S2). Of these, ten variants with the smallest *P* values in each of the two ethnic groups examined (Caucasian and African American) are listed in Table 5. In Caucasians, the variant with the smallest point-wise *P* value was in the second intron of the metabotropic glutamate receptor 8 gene *GRM8* (point-wise $P = 0.0003$). The variant with the second smallest *P* value was 44,000 nucleotides upstream of the neural cell adhesion molecule 1 *NCAM1* (point-wise $P = 0.0003$) (Table 5A). In the African American group, the variant with the smallest *P* value was in the gene encoding the cAMP-specific phosphodiesterase 4B *PDE4B* (point-wise $P = 0.0079$) and the variant with the second smallest *P* value was in the *N*-methyl D-aspartate 2A ionotropic glutamate receptor gene *GRIN2A* (point-wise $P = 0.0079$) (Table 5B). None of these variants were found to have experiment-wise significant association with heroin addiction.

All the variants in Table 5, except three, are located within the annotated genes listed. We found the linkage disequilibrium, D' and r^2, between each of the three variants that were close to, but not within, the annotated gene and a second variant within the annotated gene using data from HapMap (release 27) (www.hapmap.org). Between variant rs10492065 downstream of *KCNA1* and rs4766311 in *KCNA1* (26,423 nucleotides apart), D' = 0.97 and $r^2 = 0.027$ in the Caucasian (CEPH) population, and between rs3825786 downstream of *PLA2G4F* and rs2280248 in *PLA2G4F* (2,660 nucleotides apart), D' = 1 and $r^2 = 0.149$ in the Yoruba population. The linkage disequilibrium was low between rs2180 upstream of the *OLIGO2* gene and rs6517137 in *OLIGO2* (56,500 nucleotides apart) (D' = 0.14, $r^2 = 0.001$) in the CEPH population). However, this variant may itself be functional.

Discussion

We used a pooled DNA approach to identify single nucleotide polymorphisms associated with vulnerability to develop heroin addiction in Caucasians and African Americans. We found differences in the results between the ethnic groups studied. Other studies have found significant associations in some specific ethnicities, but not in others, most likely due to differences in allele frequencies and/or disease prevalence among ethnic groups [reviewed in (Cardon and Palmer, 2003)].

In Caucasians, variant rs10494334, located on chromosome 1, had the smallest *P* values for association with vulnerability to develop heroin addiction, which was significant after correcting for multiple testing. This variant has no annotated genes within 100,000 nucleotides. However, there is a region of approximately 50 nucleotides, located 600 nucleotides from this variant, that is highly conserved. This region is rich in predicted transcription factor binding sites and could be a region controlling expression of many genes (Morley et al., 2004).

The variant with the next smallest *P* value in the Caucasians was in the *T* gene that encodes a transcription factor, the T protein, required for posterior mesoderm differentiation and axial development (Edwards et al., 1996). A variant in the *T* gene ($TIVS_7$-2) has been reported to be associated with neural tube defects (Morrison et al., 1996, Shields et al., 2000, Jensen et al., 2004).

Three variants in close proximity in the *RIMS2* gene were found in the 500 variants with the smallest *P* values in the Caucasian group. *RIMS2* codes for three isozymes, RIM2α, RIM2β, and RIM2γ, members of the RIM family of Rab3-interacting molecules (Wang and Sudhof, 2003). RIMs are found in the active zone of presynaptic nerve terminals, the sites on the plasma membrane where synaptic vesicle exocytosis occurs (Wang et al., 2000).

Several variants were found to be clustered in the cardiomyopathy associated 3 gene *CMYA3* in the African Americans. *CYMA3*, also known as *XIRP2*, the xin actin-binding repeat containing 2 gene, is primarily expressed in striated muscle. Another variant in *CMYA3*, rs3749004 (not in our study), has been reported to be associated with autism (Faham et al., 2005).

A finding of particular interest in this study corroborated our earlier study using the 10K GeneChip with Caucasians only (Nielsen et al., 2008b). Different variants in the two studies implicate the metabotropic glutamate receptor 8 gene *GRM8* in the vulnerability to develop heroin addiction. In the earlier study, variant rs1034576, located 15,742 nucleotides downstream from *GRM8*, had the second smallest *P* value for the candidate genes. In our current study, this variant had the 339th smallest *P* value. In the Caucasians in our current study, variant rs6467108 (not on the 10K GeneChip) had the smallest *P* value of the candidate genes and is located in an intron of *GRM8*. Binding of glutamate to mGluR8

inhibits cAMP production and mGluR8 has been suggested to be a presynaptic receptor that modulates glutamate release (Scherer et al., 1997). Glutamate neurotransmission in the ventral tegmental area and in the nucleus accumbens core were shown to be required for heroin seeking and reinforcement in rodents (Xi and Stein, 2002, LaLumiere and Kalivas, 2008). Variants in *GRM8* were found to be associated with schizophrenia in Japanese (Takaki et al., 2004).

We recently reported an association of a haplotype of tryptophan hydroxylase 2 gene *TPH2* with heroin addiction vulnerability in African Americans, as well as a significant association of a genotype pattern of *TPH1* and *TPH2* variants with heroin addiction in Hispanics (Nielsen et al., 2008a). In the current study, we found that in African Americans, variant rs10506645 in *TPH2* was in the top ten of our list of candidate genes. This result supports our previous findings.

Other association studies of drug and alcohol addiction have used a similar DNA pooling approach. With a 1,494 variant chip, the Uhl group identified several variants that were associated with drug abuse vulnerability (Uhl et al., 2001). They extended their study with pools of African- and European-American cohorts using the 10K GeneChip and found 38 “nominally reproducibly positive” variants associated with non-specific substance abuse (Liu et al., 2005). Then, using the 100K GeneChip Set, they identified 51 “clustered positive” regions associated with alcohol dependence in European-Americans (Johnson et al., 2006). Based on these studies and a study that used 500K GeneChips, 89 genes were suggested to play a role in addiction vulnerability (Liu et al., 2006). In another study that used both 100K and 500K GeneChips, 39 genes that were identified as “clustered positive” were found to be in association with methamphetamine dependence (Uhl et al., 2008). Comparisons between studies are difficult to make as different platforms and dissimilar subject phenotyping were used.

Several linkage studies suggest the involvement of specific chromosomal regions in the vulnerability to develop heroin addiction (Gelernter et al., 2006, Glatt et al., 2006, Glatt et al., 2008). The initial linkage study by the Tsuang group using 192 Chinese families and 386 short tandem repeat (STR) markers found a region on chromosome 4 at D4S1644 with evidence for linkage with heroin dependence (point-wise $P = 0.014$) (Glatt et al., 2006). In their follow-up study, which included the original 192 families with a total of 397 Chinese families and 385 STR markers, the linkage signal on chromosome 4 at D4S1644 with greater significance (point-wise $P = 0.004$) (Glatt et al., 2008). Variant rs10518620, ranked seventh in Table 2 in our study and nominally associated with heroin addiction in the Caucasian group (point-wise $P = 0.0003$), is located between D4S1644 and the nearby marker D4S2394, which had nominal significance (point-wise $P = 0.013$) in the later Tsuang study. In the linkage study of Gelernter, which used 409 STR markers in 393 families, a linkage peak at D17S785 was found (LOD = 3.06, empirical $P = 0.0002$) (Gelernter et al., 2006). Variant rs9271, which was ranked sixth in Table 2 in our current study and was found to be nominally associated with heroin addiction in the Caucasians (point-wise $P = 0.0003$), is located six million nucleotides from D17S785.

There are two recent linkage studies on opioid dependence using single nucleotide polymorphisms (Lachman et al., 2007, Yu et al., 2008). The Gelernter group found eight variants with point-wise significance in an association study of opiate dependence (Yu et al., 2008), and Lachman *et al.* found one region at chromosome 14q in their Hispanic group that was “suggestive” of genome-wide evidence for linkage (Lachman et al., 2007). We did not confirm these linkage findings in our study.

Future association studies in other cohorts of well-defined ethnicity and carefully defined phenotypes will be required to replicate our findings, which have identified several new candidate genes that may be involved in vulnerability to develop heroin addiction. If confirmed, these findings could lead to new targets for strategies for prevention and the pharmacological treatment of heroin addiction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dorothy Melia, R.N., Kathy Bell, R.N., Elizabeth Ducat, N.P., Lisa Borg, M.D., Brenda Ray, N.P., Pauline McHugh, M.D., James Schluger, M.D., and Heather Hofflich, D.O. for recruiting, screening, and assessment of study subjects. Connie Zhao, Ph.D. for the processing of the microarrays, and Sara Hamon for review and discussion of the statistical analyses. We are grateful to the late K. Steven LaForge for his role in the planning of these genetic studies.

This work was supported in part by NIH-NIDA Grants K05-DA00049 (M.J.K.) and RO1-DA12848 (M.J.K.), NIH-RR UL1RR024143 (B.C.), NIH-MH R01-44292 (J.O.), and NSFC grant 30730057 and 30700442 from the Chinese Government (J.O.).

References

Benko S, Fantes JA, Amiel J, Kleinjan DJ, Thomas S, Ramsay J, et al. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat Genet. 2009; 41:359–64. [PubMed: 19234473]

Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet. 2003; 361:598–604. [PubMed: 12598158]

Collins A, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA. 1999; 96:15173–15177. [PubMed: 10611357]

Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, Gejman PV. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum Mol Genet. 2003; 12:205–216. [PubMed: 12554675]

Edwards YH, Putt W, Lekoape KM, Stott D, Fox M, Hopkinson DA, et al. The human homolog T of the mouse T(Brachyury) gene; gene structure, cDNA sequence, and assignment to chromosome 6q27. Genome Res. 1996; 6:226–33. [PubMed: 8963900]

Faham M, Zheng J, Moorhead M, Fakhrai-Rad H, Namsaraev E, Wong K, et al. Multiplexed variation scanning for 1,000 amplicons in hundreds of patients using mismatch repair detection (MRD) on tag arrays. Proceedings of the National Academy of Sciences. 2005; 102:14717–14722.

Friedberg I, Nika K, Tautz L, Saito K, Cerignoli F, Godzik A, et al. Identification and characterization of DUSP27, a novel dual-specific protein phosphatase. FEBS Lett. 2007; 581:2527–2533. [PubMed: 17498703]

Gelernter J, Panhuysen C, Wilcox M, Hesselbrock V, Rounsaville B, Poling J, et al. Genomewide linkage scan for opioid dependence and related traits. Am J Hum Genet. 2006; 78:759–769. [PubMed: 16642432]

Glatt SJ, Lasky-Su JA, Zhu SC, Zhang R, Zhang B, Li J, et al. Genome-wide linkage analysis of heroin dependence in Han Chinese: Results from Wave Two of a multi-stage study. Drug Alcohol Depend. 2008; 98:30–4. [PubMed: 18538955]

Glatt SJ, Su JA, Zhu SC, Zhang R, Zhang B, Li J, et al. Genome-wide linkage analysis of heroin dependence in Han Chinese: Results from wave one of a multi-stage study. Am J Med Genet B Neuropsychiatr Genet. 2006; 141B:648–652. [PubMed: 16856125]

Hattori E, Liu C, Zhu H, Gershon ES. Genetic tests of biologic systems in affective disorders. Mol Psychiatry. 2005; 10:719–740. [PubMed: 15940293]

Jensen LE, Barbaux S, Hoess K, Fraterman S, Whitehead AS, Mitchell LE. The human T locus and spina bifida risk. Hum Genet. 2004; 115:475–82. [PubMed: 15449172]

Jeong Y, Leskow FC, El-Jaick K, Roessler E, Muenke M, Yocum A, et al. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. Nat Genet. 2008; 40:1348–53. [PubMed: 18836447]

Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J, et al. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: Validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2006; 141B:844–853.

Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. Am J Psychiatry. 2003; 160:687–695. [PubMed: 12668357]

King DC, Taylor J, Elnitski L, Chiaromonte F, Miller W, Hardison RC. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res. 2005; 15:1051–1060. [PubMed: 16024817]

Kirov G, Nikolov I, Georgieva L, Moskvina V, Owen MJ, O'donovan M C. Pooled DNA genotyping on Affymetrix SNP genotyping arrays. BMC Genomics. 2006; 7:27. [PubMed: 16480507]

Kreek, MJ. Gene diversity in the endorphin system: SNPs, chips, and possible implications. In: Yudell, M.; Dasalle, R., editors. The Genomic Revolution: Unveiling the Unity of Life. Joseph Henry Press; Washington, DC: 2002.

Lachman HM, Fann CS, Bartzis M, Evgrafov OV, Rosenthal RN, Nunes EV, et al. Genomewide suggestive linkage of opioid dependence to chromosome 14q. Hum Mol Genet. 2007; 16:1327–1334. [PubMed: 17409192]

Lalumiere RT, Kalivas PW. Glutamate release in the nucleus accumbens core is necessary for heroin seeking. J Neurosci. 2008; 28:3170–7. [PubMed: 18354020]

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003; 12:1725–35. [PubMed: 12837695]

Liu QR, Drgon T, Johnson C, Walther D, Hess J, Uhl GR. Addiction molecular genetics: 639,401 SNP whole genome association identifies many "cell adhesion" genes. Am J Med Genet B Neuropsychiatr Genet. 2006; 141B:918–925. [PubMed: 17099884]

Liu QR, Drgon T, Walther D, Johnson C, Poleskaya O, Hess J, et al. Pooled association genome scanning: validation and use to identify addiction vulnerability loci in two samples. Proc Natl Acad Sci USA. 2005; 102:11864–11869. [PubMed: 16091475]

Mclellan AT, Luborsky L, Woody GE, O'brien CP. An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. J Nerv Ment Dis. 1980; 168:26–33. [PubMed: 7351540]

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, et al. Genetic analysis of genome-wide variation in human gene expression. Nature. 2004; 430:743–7. [PubMed: 15269782]

Morrison K, Papapetrou C, Attwood J, Hol F, Lynch SA, Sampath A, et al. Genetic mapping of the human homologue (T) of mouse T(Brachyury) and a search for allele association between human T and spina bifida. Hum Mol Genet. 1996; 5:669–74. [PubMed: 8733136]

Nielsen DA, Barral S, Proudnikov D, Kellogg S, Ho A, Ott J, et al. TPH2 and TPH1: Association of Variants and Interactions with Heroin Addiction. Behav Genet. 2008a; 38:133–50. [PubMed: 18181017]

Nielsen DA, Ji F, Yuferov V, Ho A, Chen A, Levran O, et al. Genotype patterns that contribute to increased risk for or protection from developing heroin addiction. Mol Psychiatry. 2008b; 13:417–28. [PubMed: 18195715]

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. Linkage disequilibrium in the human genome. Nature. 2001; 411:199–204. [PubMed: 11346797]

Rettig, RA.; Yarmolinsky, A., editors. Federal Regulation of Methadone Treatment. National Academy Press; Washington, DC: 1995.

Scherer SW, Soder S, Duvoisin RM, Huizenga JJ, Tsui LC. The human metabotropic glutamate receptor 8 (GRM8) gene: a disproportionately large gene located at 7q31.3-q32.1. Genomics. 1997; 44:232–6. [PubMed: 9299241]

Schug, J.; Overton, GC. TESS: Transcription Element Search Software on the WWW. Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania; Philadelphia: 1977.

Sham P, Bader JS, Craig I, O'donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. Nat Rev Genet. 2002; 3:862–871. [PubMed: 12415316]

Shields DC, Ramsbottom D, Donoghue C, Pinjon E, Kirke PN, Molloy AM, et al. Association between historically high frequencies of neural tube defects and the human T homologue of mouse T (Brachyury). Am J Med Genet. 2000; 92:206–11. [PubMed: 10817656]

Simon RM, Korn EL, Mcshane LM, Radmacher MD, Wright GW, Zhao Y. Design and Analysis of DNA Microarray Investigations Series: Statistics for Biology and Health. 2004

Takaki H, Kikuta R, Shibata H, Ninomiya H, Tashiro N, Fukumaki Y. Positive associations of polymorphisms in the metabotropic glutamate receptor type 8 gene (GRM8) with schizophrenia. Am J Med Genet B Neuropsychiatr Genet. 2004; 128B:6–14. [PubMed: 15211621]

Tsuang MT, Lyons MJ, Eisen SA, Goldberg J, True W, Lin N, et al. Genetic influences on DSM-III-R drug abuse and dependence: a study of 3,372 twin pairs. Am J Med Genet. 1996; 67:473–477. [PubMed: 8886164]

Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, et al. Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities. Arch Gen Psychiatry. 1998; 55:967–972. [PubMed: 9819064]

Uhl GR, Drgon T, Liu QR, Johnson C, Walther D, Komiyama T, et al. Genome-wide association for methamphetamine dependence: convergent results from 2 samples. Arch Gen Psychiatry. 2008; 65:345–55. [PubMed: 18316681]

Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE. Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. BMC Genet. 2007; 8:10. [PubMed: 17407593]

Uhl GR, Liu QR, Walther D, Hess J, Naiman D. Polysubstance abuse-vulnerability genes: genome scans for association, using 1,004 subjects and 1,494 single-nucleotide polymorphisms. Am J Hum Genet. 2001; 69:1290–1300. [PubMed: 11704927]

Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 2008; 4:e1000214. [PubMed: 18846210]

Wang Y, Sudhof TC. Genomic definition of RIM proteins: evolutionary amplification of a family of synaptic regulatory proteins(small star, filled). Genomics. 2003; 81:126–37. [PubMed: 12620390]

Wang Y, Sugita S, Sudhof TC. The RIM/NIM family of neuronal C2 domain proteins. Interactions with Rab3 and a new class of Src homology 3 domain proteins. J Biol Chem. 2000; 275:20033–44. [PubMed: 10748113]

Xi ZX, Stein EA. Blockade of ionotropic glutamatergic transmission in the ventral tegmental area reduces heroin reinforcement in rat. Psychopharmacology (Berl). 2002; 164:144–50. [PubMed: 12404076]

Yu Y, Kranzler HR, Panhuysen C, Weiss RD, Poling J, Farrer LA, et al. Substance dependence low-density whole genome association study in two distinct American populations. Hum Genet. 2008; 123:495–506. [PubMed: 18438686]

Table 1

Pool composition

Pools of unique subjects	Total pools[a]	Ethnicity	Category	No. of subjects
8	16	Caucasian	Severe former heroin addicts	200
6	12	Caucasian	Controls	150
5	10	African American	Severe former heroin addicts	125
4	8	African American	Control	100
Total: 23	46			575

[a]Each pool was created separately (i.e., in duplicate) from DNA of 25 different subjects to yield a total of 46 pools.

Table 2

Top 10 variants ranked by ascending experiment-wise *P* value based on analysis of the association of allele frequency with heroin addiction in the Caucasian group

Rank Order	Variant	Allele frequency, A[a]		Experiment-wise *P*[b]	Cytoband	Variant location[c]	Distance (bases)	Gene	Gene description
		Control	Case						
1	rs10494334	0.18	0.23	0.035	1q23.3				
2	rs2323218	0.18	0.25	0.221	6q27	down[d]	8,989	*T*	T, brachyury homolog
3	rs7923687	0.83	0.79	0.426	10q25.1				
4	rs6121489	0.03	0.05	0.523	20q13.33	intron[e]		*CDH4*	cadherin 4, type 1, R-cadherin (retinal)
5	rs3849399	0.78	0.71	0.605	2p24.1				
6	rs9271	0.75	0.68	0.620	17q25.3	3′UTR[f]		*WDR45L*	DR45-like
7	rs10518620	0.89	0.84	0.632	4q28.3				
8	rs3893249	0.79	0.72	0.640	2p24.1				
9	rs1711055	0.30	0.25	0.736	15q22.1	intron		*AQP9*	aquaporin 9
10	rs724729	0.73	0.63	0.741	15q14				

[a] A allele is defined by Affymetrix and can be found at the NetAffx Analysis Center (http://www.affymetrix.com/analysis/index.affx).

[b] Point-wise *P* value was 0.0003 for all the alleles listed. The *t* statistic of the observed data was calculated and then the case/control labels of each subject were permuted. For each permutation, a *t* statistic was obtained. For a sample size of 14 with 8 cases and 6 controls, there are 3,003 total possible permutations. The value 0.0003 was obtained when the observed *t* statistic is the smallest among 3,003 *t* statistics (1/3,003 = 0.0003). The originally observed *t* statistic was compared to the distribution of *t* statistics composed of 3,003 maximum *t* statistics to obtain the experiment-wise *P* value.

[c] Variant location is given when variant is found with 100,000 nucleotides of an annotated gene.

[d] down = distance the variant is located downstream of the polyadenylation site of the specified gene.

[e] intron = variant is located in the intron of specified gene.

[f] 3′UTR = variant is located in the 3; untranslated region of the specified gene.

Data is from the NetAffx web site; verified and corrected using the USCS Genome Browser.

Table 3

Top 10 variants ranked by ascending experiment-wise *P* value based on analysis of the association of allele frequency with heroin addiction in the African American group

Rank Order	Variant	Allele frequency, A		Experiment wise P^a	Cytoband	Variant location[b]	Distance (bases)	Gene	Gene description
		Control	Case						
1	rs950302	0.58	0.47	0.214	1q24.1	intron		*DUSP27*	dual specificity phosphatase 27 (putative)
2	rs990937	0.34	0.30	0.651	9p22.3				
3	rs10513523	0.89	0.94	0.873	3q25.32				
4	rs579533	0.47	0.37	0.976	6q25.3				
5	rs4075106	0.78	0.70	0.984	12p13.31	intron		*CD163L1*	CD163 molecule-like 1
6	rs2140014	0.68	0.82	0.992	8q21.13				
7	rs4597125	0.36	0.33	1	12q21.2	down	10,655	*PHLDA1*	pleckstrin homology-like domain, family A
8	rs1494066	0.52	0.63	1	5q15.33				
9	rs10495722	0.40	0.33	1	2p24.1				
10	rs9289317	0.81	0.84	1	3q21.3	intron		*MGLL*	monoglyceride lipase isoform 1

[a]Point-wise *P* value was 0.0079 for all the alleles (refer to footnote [b] of Table 2)

Table 4

Sets of three or more variants within 100,000 nucleotides of each other that are in the top 500 most significant variant list

Variant	Allele frequency (A)		Absolute difference in allele frequency	*P*	Experiment-wise *P*	Rank of *P*	Distance to next variant (nucleotides)	Location
	Control	Case						
A. Caucasian								
***RIMS2*: Regulating synaptic membrane exocytosis protein 2 (chromosome 8q22.3)**								
rs2511571	0.59	0.66	0.07	0.0007	1	146	2,031	intron
rs2511576	0.63	0.71	0.08	0.0007	1	178	12,304	intron
rs1156813	0.33	0.26	0.07	0.0003	0.98	34		intron
B. African American i.								
i. *CMYA3*: cardiomyopathy associated 3 (chromosome 2q24.3)								
rs1429931	0.73	0.79	0.07	0.0238	1	364	428	intron
rs6706115	0.67	0.74	0.08	0.0079	1	136	2,913	intron
rs10497320	0.20	0.17	0.03	0.0079	1			intron
ii. No gene within 100,000 nucleotides (chromosome 13q13.1)								
rs1937387	0.63	0.54	0.09	0.0079	1	157	3,132	
rs2876780	0.77	0.66	0.11	0.0079	1	155	5,877	
rs9318868	0.73	0.63	0.10	0.0079	1			

Table 5

Genes from the hypothesis-based gene lists of the Kreek (Table S1) and Gershon (Table S2) containing the ten variants ranked by significance

Symbol	Gene	Gene location[a]	Variant	Variant location	Distance from closet annotated gene (nucleotides)	Allele associated with heroin addiction[b]	T-statistic	*P*	Rank[c]
A. Caucasian									
GRM8	glutamate receptor, metabotropic 8	7q31.33	rs6467108	intron		C	5.56	0.0003	1
NCAM1	neural cell adhesion molecule 1	11q23.1	rs1245124	intron		T	5.10	0.0003	2
GABBR2	gamma-aminobutyric acid (GABA) B receptor, 2	9q22.33	rs2779577	intron		C	4.99	0.0003	3
GABRB2	gamma-aminobutyric acid (GABA) A receptor, beta 2	5q34	rs10515827	intron		T	4.89	0.0003	4
KCNA1	potassium voltage-gated channel, shaker-related subfamily, member 1	12p13.32	rs10492065	down	21,439	T	6.08	0.0007	5
GRIK4	glutamate receptor, ionotropic, kainate 4	11q23.3	rs10502240	intron		A	5.88	0.0007	6
KCND2	potassium voltage-gated channel, Shal-related subfamily, member 2	7q31.31	rs917901	intron		G	5.50	0.0007	7
GABRG3	gamma-aminobutyric acid (GABA) A receptor, gamma 3	15q12	rs1378101	intron		A	4.97	0.0007	8
HTR4	5-hydroxytryptamine (serotonin) receptor 4	5q33.1	rs9325102	intron		C	5.02	0.0010	9
OLIG2	oligodendrocyte lineage transcription factor 2	21q22.11	rs2180	up	53,874	G	5.0	0.0010	10
B. African American									
PDE4B	phosphodiesterase 4B, cAMP-specific	1p31.3	rs10493398	intron		G	7.20	0.0079	1
GRIN2A	glutamate receptor, ionotropic, N-methyl D-aspartate 2A	16p13.2	rs10518152	intron		A	5.73	0.0079	2
GRM5	glutamate receptor, metabotropic 5	11q14.3	rs524874	intron		A	5.51	0.0079	3
GLRA3	glycine receptor, alpha 3	4q34.1	rs1491402	intron		C	4.58	0.0079	4
SERPINA6	serpin peptidase inhibitor, clade A	14q32.13	rs8023023	intron		C	4.47	0.0079	5
PLA2G4F	phospholipase A2, group IVD (cytosolic)	15q15.1	rs3825786	down	2,257	T	4.17	0.0079	6
GRIA4	glutamate receptor, ionotropic, AMPA 4	11q22.3	rs502300	intron		T	4.05	0.0079	7
DTNBP1	dystrobrevin binding protein 1	6p22.3	rs9296978	intron		G	3.84	0.0079	8

Symbol	Gene	Gene location[a]	Variant	Variant location	Distance from closet annotated gene (nucleotides)	Allele associated with heroin addiction[b]	T-statistic	*P*	Rank[c]
			rs2743868	intron		C	3.80	0.0079	9
TPH2	tryptophan hydroxylase 2	12q21.1	rs10506645	intron		A	3.80	0.0079	10

[a] Gene location from UCSC Genome Browser, March 2006 assembly (NCBI Build 36.1).

[b] Allele more frequent in cases than controls

[c] Rank is based on ascending *P* value. If the point-wise non-parametric *P*s for several variants are the same, the variant with the larger T-statistic is ranked higher.

Psychiatr Genet. Author manuscript; available in PMC 2013 November 18.