# Rare Non-Synonymous Exonic Variants in Addiction and Behavioral Disinhibition

**Scott I. Vrieze**[1], **Shuang Feng**[1], **Michael B. Miller**[2], **Brian M. Hicks**[3], **Nathan Pankratz**[4], **Gonçalo R. Abecasis**[1], **William G. Iacono**[2], and **Matt McGue**[2]

[1]Center for Statistical Genetics, Department of Biostatistics, University of Michigan

[2]Department of Psychology, University of Minnesota

[3]Department of Psychiatry, University of Michigan

[4]Department of Laboratory Medicine and Pathology, University of Minnesota

## Abstract

**Background**—Substance use is heritable but few common genetic variants have been associated with these behaviors. Rare non-synonymous exonic variants can now be efficiently genotyped, allowing exome-wide association tests. We identified and tested nonsynonymous variants for association with behavioral disinhibition and the use/misuse of nicotine, alcohol, and illicit drugs.

**Methods**—Comprehensive genotyping of exonic variation combined with single-variant and gene-based tests of association in 7181 individuals; 172 candidate addiction genes were evaluated in greater detail. We also evaluated the aggregate effects of nonsynonymous variants on these phenotypes using GCTA.

**Results**—No variant or gene was significantly associated with any phenotype. No association was found for any of the 172 candidate genes, even at reduced significance thresholds. All nonsynonymous variants jointly accounted for 35% of the heritability in illicit drug use and, when combined with common variants from a genome-wide array, accounted for 84% of the heritability.

**Conclusions**—Rare nonsynonymous variants may be important in etiology of illicit drug use, but detection of individual variants will require very large samples.

## Keywords

Behavioral Disinhibition; Addiction; Exome; Nonsynonymous; Tobacco; Alcohol; Drug

---

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Genome-wide association studies (GWAS) have served as the standard for identifying relationships between genotype and phenotype. GWAS is motivated by the common disease common variant hypothesis, which holds that complex disease can be accounted for by a large number of common variants with individually small effects. This appears to hold true for a variety of medical, anthropometric, and behavioral phenotypes, although massively polygenic effects are the norm(1). However, common variants do not account for all the heritability in these phenotypes, leaving many to speculate about the nature of this "missing heritability"(2).

One explanation for missing heritability is that there is substantial genetic variation that is not captured by common variants. Genome sequencing has recently enabled the discovery of tens of millions of rare genetic variants(3, 4). Relatively few sequenced individuals are needed for variant discovery; many more are required to establish genotype-phenotype associations. While the cost of sequencing is declining, it is still not cost-effective for obtaining sample sizes powerful enough to detect associations with rare variants(5). Until inexpensive sequencing is available, an alternative strategy is to use rare variant genotyping chips(6), that genotype selected rare variants at a small fraction of the cost of sequencing. In the present study, we used the Illumina HumanExome rare variant chip to genotype rare and functional exonic variants.

The exome is the protein-coding portion of the genome. The exome chip was designed by a coalition of investigators who pooled exome sequencing data to identify rare nonsynonymous SNPs. Such variants are expected to be highly deleterious and therefore rare, which poses practical problems for phenotype-genotype association studies, as statistical power is reduced for rare SNPs. However, we expect low power to be mitigated in part due to larger effects of nonsynonymous SNPs, as they have the potential to disable an entire gene.

In the present study, we tested for associations between these SNPs and nicotine dependence, alcohol consumption, alcohol dependence, drug use, and a non-substance use measure called behavioral disinhibition. These phenotypes have all been described previously(7) and have been previously analyzed with respect to common SNPs(8, 9).

Behavioral disinhibition connotes the difficulty to consider long-term gains or losses in the regulation of impulses(10, 11). High behavioral disinhibition is associated with increased risk for substance use disorders, antisocial behavior, and disinhibitory personality traits (e.g., impulsivity) (12–14). Behavioral disinhibition is hypothesized to be a risk factor for substance use generally, and to account for much of the high comorbidity among substance use disorders (15). Twin and family studies report that behavioral disinhibition is heritable (60–80%)(14, 16, 17), suggesting a non-specific genetic risk that contributes to the expression of substance use disorders (18).

Three GWAS of non-substance-use disinhibition constructs exist. One GWAS, using the present sample and phenotypes, identified no significant loci(8). Null results were also reported in a study of adult antisocial personality disorder(19). Finally, Dick et al.(20)

identified four loci associated with conduct disorder in a case-control sample of 3,963 across several ancestral groups.

Investigations of substance use phenotypes have identified multiple loci. For cigarette smoking, GWAS meta-analyses have identified SNPs in cholinergic nicotinic receptor subunit gene clusters *CHRNA5-CHRNA3-CHRNB4*(21–23) and *CHRNB3-CHRNA6*(24), as well as nicotine metabolizing enzyme genes *CYP2A6-CYP2B6*(24). One locus was associated with alcohol consumption in the *AUTS2* gene and subsequently validated in a mouse model(25).

In the present study we evaluated deleterious exonic variants in an attempt to identify novel loci and refine known associations within the genes listed above. In addition, although we do not review the expansive candidate gene literature here, we provide results for commonly-studied candidate genes in systems ranging from neurotransmitters to alcohol metabolism to genes involved in stress reaction.

## Method

### Participants

Individuals were participants in studies conducted at the Minnesota Center for Twin and Family Research(26), which utilized a community-based ascertainment and accelerated longitudinal cohort designs. The sample included 7181individuals nested within 2299 pedigrees designed to include two parents and two offspring. Some pedigrees were incomplete because a participant failed to contribute a viable DNA sample. Offspring family members included monozygotic twins, dizygotic twins, non-twin biological offspring and siblings, or adoptive offspring and siblings. There were 1771 mothers (mean age=42, SD=5.3), 2055 fathers (45, 5.7), and 3355 offspring (47% male). The study is longitudinal, with repeated assessments on the offspring. For offspring, we maximized the size of the sample by selecting assessments as close to age 17 as possible (mean age=18.0, SD=0.8, range=[16.5, 21.0]). All research reported in this manuscript was approved by the University of Minnesota Institutional Review Board.

### Phenotypes

Phenotypes were composite measures derived from factor analysis and factor score extraction of questionnaire and diagnostic symptom data of substance use, substance use disorders, and behavioral disinhibition. An extensive description of the composites and constituent measures has been provided elsewhere(7). Please see Table S1 in Supplement 1 for information on the constituent measures used to create these phenotype composites. Substance use and symptoms of substance use disorders were assessed, through in-person clinical interview with trained interviewers, using an expanded version of the Substance Abuse Module(27)of the Composite International Diagnostic Interview(28). The nicotine dependence composite included frequency (days per month), quantity (cigarettes per day), and DSM symptoms of dependence. Alcohol consumption included frequency (never to multiple times a day), maximum consumption (number of drinks in 24 hours), and lifetime intoxications. Alcohol dependence included abuse and dependence symptoms from multiple

diagnostic systems organized into facets of social and occupational problems (e.g., drinking and driving), withdrawal and tolerance, and compulsive use and impairment in major life activities (e.g., drinking throughout the day; wanted to stop but couldn't). Illicit drug use included frequency of marijuana use, a sum score of the number of different drug classes used (marijuana, amphetamines, barbiturates, cocaine, gas, hallucinogens, heroin, inhalants, opiates, PCP, tranquilizers), and symptoms of abuse and dependence for the most frequently used substance. Behavioral disinhibition included symptoms of antisocial personality disorder and non-diagnostic dissocial behaviors (being arrested, suspended from school, and early initiation of sexual intercourse) assessed via structured interviews and questionnaire measures of delinquency and disinhibited personality traits (impulsivity, aggression, risk taking).

In addition to the five traits discussed above, we also evaluated the nicotine dependence, alcohol dependence, and alcohol consumption phenotypes among individuals with adequate exposure to these substances. Nicotine exposure was defined as having done at least one of the following: smoked 2+ cigarettes on days they reported smoking or smoked 4+ days per month in the past year or having 1+ lifetime symptom(s) of dependence. Alcohol exposure was defined as having done one of the following: drinking at least once a month in the past year or ever having 6+ drinks in 24 hours or at least 2 lifetime intoxications or having 1+ lifetime symptom(s) of dependence. Minor modifications to these criteria resulted in only small fluctuations in the selected sample. Using these criteria, 3412 individuals from 1694 families were considered exposed to nicotine, and 5897 individuals from 2211 families were considered exposed to alcohol. We refer to these phenotypes throughout as nicotine (exposed), drinking (exposed), and alcohol dependence (exposed). We did not consider exposed versions of illicit drug use for several reasons, including the complex process of conditioning measures for each of 11 substances on the use of that substance, then aggregating those conditioned measures. It was unclear to us the meaning of a factor produced in this way.

Phenotypes were each separately corrected for covariates and inverse-normalized. Covariate main effects included age, $age^2$, sex, generational status (parent/child), year of birth, first 10 genetic principal components, as well as interactions of birth year $\times$ generation, sex $\times$ generation, and age $\times$ generation. Birth year and age were quantitative covariates. With the exception of principal components, each covariate was highly significantly associated with most phenotypes (Table S4 in Supplement 1). Phenotypes were highly correlated and showed significant within-family correlations (Tables S2 and S3 in Supplement 1). Scatterplot matrices of the phenotypes, before and after inverse normalization are displayed in Figures S1 and S2 (see Supplement 1). Univariate phenotype histograms, prior to inverse-normalization, are displayed in Figure S3 (see Supplement 1).

## Genotypes

The sample was genotyped on the Illumina HumanExome BeadChip array and the Illumina 660W Quad. Extensive description of the 660W-Quad genome-wide array in this sample was published previously(26) and is extremely similar to the description of the exome array, so we omit it for space. The Illumina HumanExome chip included 247,737 SNPs. All

samples were clustered together in GenomeStudio, and markers showing evidence of poor quality, such as lower call rates, poor cluster separation, deviation from Hardy-Weinberg equilibrium, or rare (0  MAF<0.0001) and incomplete call rate (<1.0), were manually reviewed and reclustered or discarded (n~5,000). After removing SNPs identified by Illumina as having technical defects we had 247,186 SNPs for further QC testing. A total of 1,250 markers were dropped in our QC screen: 503 SNPs had a no-call rate exceeding 1%, 52 had a Hardy-Weinberg test $p$-value less than the Bonferroni cutoff of $2e^{-7}$, 28 autosomal markers were associated with participant sex ($p<2e^{-7}$) and 793 SNPs had two or more errors of any of these kinds: incorrect call in a CEPH sample, discordant calls in a pair of duplicated samples or an MZ twin pair, Mendelian inconsistency, heterozygosity for an X-chromosome marker in a male sample, or nonmissing Y-chromosome genotype for a female sample.

In total 7,350 samples were run through the HumanExome array. After QC screening we were left with useable genotype data for 7,244 subjects, including only one member of each MZ twin pair. Samples were dropped for a no-call rate of 0.5% or higher (145 samples), 16 samples were dropped because of sample mix-ups, four because of consent issues, and for all duplicate samples, only the one with the highest call rate was retained. After creating records for 1,120 MZ cotwins by copying genotype data of their MZ twins, the sample size useable for GWAS increased to 8,364 subjects, including all ethnic groups.

Markers from the exome chip and genome-wide chip were combined into an integrated panel of 629,573 polymorphic SNPs. There were 84,678,352 overlapping genotypes between the arrays in this sample (11,792 overlapping SNPs times 7181 individuals; we include the MZ co-twins in these numbers). Of these, 84,654,782 were non-missing on both chips (99.9722%). Of non-missing variants, 84,654,329 (99.9995%) were concordant. The 453 discordant genotypes were set to missing.

Next, restricting the sample to individuals of European ancestry who have phenotype data reduced the total sample available for the present study to 7181. We generated 10 principal components using Eigenstrat, and determined ancestry based on a combination of these principal components and self-reported ancestry as described previously(26).

In the 7181 subjects there were 629,573 polymorphic autosomal SNPs. Of these, 133,517 had MAF<.05. For nicotine exposed there were 615,184 polymorphic SNPs with 119,226 rare SNPs. For alcohol dependence exposed and alcohol consumption exposed there were 627,472 polymorphic SNPs with 131,386 rare SNPs.

Markers were annotated with ANNO(29) using GENCODE v11 transcripts from the ENCODE project. In total there were 111,592 polymorphic nonsynonymous variants in exomic regions, with 2973 start- or stop-changing and 1762 splice-site variants. Descriptive information for the non-synonymous marker set is displayed in Table 1.

### Candidate Genes

In addition to evaluating all SNPs on the integrated array, we also evaluated 172 candidate addiction genes from two published sources(30, 31), and a shorter list of "high priority"

genes implicated in smoking and drinking from GWAS meta-analyses of tobacco and alcohol use(22, 24, 25, 32). Of the 172 genes, 151 contained polymorphic nonsynonymous variants in our sample. We report burden test results on the full 151, and also report single variant tests of nonsynonymous variants in the shorter list of high priority genes (i.e., CHRNA5, CHRNA3, CHRNA6, CHRNB4, CYP2A6, CYP2B6, ELGN2, DBH, ALDH2, ADH1B, and AUTS2). The meta-analyses evaluated smoking and alcohol consumption, measures readily available in large epidemiological samples, and we restricted our tests to similar phenotypes in the present study. Specifically, we tested these 11 genes on nicotine (exposed), alcohol dependence (full sample and exposed individuals), and drinking (full sample and exposed individuals).

### Single Variant Analysis

Single variant analysis was conducted for each SNP and each phenotype. To account for family structure we used a mixed model(33) where the kinship matrix was fixed according to the known pedigree structure. The model was estimated with an implementation of FaST-LMM(34). Mixed models allow an estimate of phenotypic heritability, where the phenotypic variance-covariance matrix is partitioned into:

$$Cov(\mathbf{y}) = \sigma_g^2 \mathbf{K} + \sigma_e^2 \mathbf{I},$$

where $Cov(\mathbf{y})$ is the N×N phenotypic matrix, $\mathbf{K}$ is the N×N kinship matrix fixed according to the known pedigree structure, $\sigma_g$ is the heritability variance component, $\sigma_e$ is the error variance, and $\mathbf{I}$ is an N×N identity matrix.

### Burden Tests of Nonsynonymous SNPs

Statistical power in association studies is in part a function of minor allele frequency and it is expected that individual tests of rare variants will lack sufficient statistical power. To address this, we implemented a series of burden tests that grouped nonsynonymous variants within a gene and test for an association between the grouping value and the phenotype(35, 36). Different burden tests perform more or less well depending on the true genetic effect(35). Accordingly, we selected two burden tests with different assumptions: a variable threshold collapsing and multivariate count method (VTCMC(37, 38)), and the sequence kernel association test (SKAT(39)).

VTCMC assumes the same direction of effect for all rare variants within a gene. It sums the number of observed rare alleles within the gene for each person, and tests the association between the gene-based sumscore and the phenotype. Since the optimal MAF cutoff for "rare" variants is unknown, the VTCMC uses the cutoff that yields the strongest effect. Since the SNP effects and the MAF cutoff are chosen in the same data, the resulting p-value is asymptotically corrected for capitalization on chance.

SKAT is a kernel-based method(39–41) that allows for variants within a gene to have different directions of effect. We used the linear kernel function reported in Wu et al.(39), which sums the weighted genetic similarity across markers within a region (here, a gene). The square root of the weight followed a Beta(MAF; $a_1$, $a_2$) distribution with parameters $a_1$=1 and $a_2$=25, which weights rarer variants (e.g., MAF< 1%) more heavily than more

common variants (1% < MAF < 5%). The method tests the null hypothesis that the distribution of the SNP effects within a gene has zero mean and covariance $\tau\mathbf{K}$, where $\mathbf{K}$ is the kernel value and $\tau$ is the variance component attributable to the SNPs. The value of $\tau$ thus depends only on the magnitude of the SNP effects, and not their direction of effect.

### Power

The effective sample size was somewhere between the number of independent families (2299), and the number of individuals (7181). For single variant tests on the primary traits, this resulted in 80% power to identify a SNP effect with $r^2$ between .005 and .015 assuming alpha=$3.4\times10^{-7}$, the Bonferroni correction for all variants with MAF<.05. While power appears reasonable, it is important to note that for a very rare single variant, such as one with MAF=.1%, an $r^2$ of .005 corresponds to a 0.5 standard deviation increase in the trait for each rare allele an individual possesses, a very large effect. Power for burden tests depends on, among other things, assumptions about direction of effect, the number of SNPs within a region, and the proportion of causal SNPs. We therefore refer the reader to prior publications on VTCMC and SKAT(38, 39).

### Heritability Due to Rare Variants

Genome-wide Complex Trait Analysis (GCTA) is a method to calculate the phenotypic variance accounted for by all SNPs in aggregate. Genetic relationships among all unrelated individuals in a sample are used in a mixed model to account for variance in the phenotype of interest. While GCTA has been applied regularly to common SNPs(1), including reports for the phenotypes in this sample(9), to our knowledge its application has not been extended to rare exonic variants. Further, since the sample is also genotyped on a genome-wide array, we were able to compare multiple marker sets: the set of all rare (MAF<.05) nonsynonymous variants; the set of all common variants (MAF .05), regardless of annotation.

Using GCTA version 1.04(42) we computed a genetic relatedness matrix for each of the two SNP sets, rare and common. Because the sample is composed of families, and familial relationships include non-additive and shared environmental effects, we restricted the GCTA method only to nominally unrelated individuals by using individuals with a genetic relatedness less than .025 based on common SNPs, a threshold that is standard practice in these analyses. The genetic relatedness estimate used is a distance metric analogous to the average SNP correlation between two individuals. To test the effect of this cutoff, we evaluated selected results at many cutoffs ranging from .01 to 1.0 (Figure S12 in Supplement 1).

## Results

### Single Variant Tests

Genomic control values for the eight phenotypes were acceptable at 1.07, 1.02, 1.06, 1.05, 1.05, 1.01, 1.03, and 1.06, for single variant tests of nicotine dependence, alcohol consumption, alcohol dependence, illicit drugs, behavioral disinhibition, nicotine dependence (exposed), alcohol consumption (exposed), and alcohol dependence (exposed),

respectively. The covariate r$^2$ was, .13, .28, .23, .14, .19, .40, .30, and .21. In the single variant genetic association tests no individual variant was significant after Bonferroni correction for number of SNPs tested, whether that correction included all SNPs or was restricted to rare variants. QQ plots are provided in Figures S4–S11 (see Supplement 1).

In further exploratory analysis, we evaluated nonsynonymous variants with minor allele count (MAC) 4 and MAF .05. We did this because common variants (MAF>.05) have been tested elsewhere and there is very limited power to detect effects for variants with MAC<4, which is why burden tests are used(35). Evaluating this restricted SNP set lowers the Bonferroni threshold to $8.06 \times 10^{-7}$, and resulted in one SNP at 5:73207116 in the GRNEF gene associated with nicotine dependence (reference allele=G; alternate=A; MAC=12, effect=1.49; p=$6.9 \times 10^{-7}$). The QQ plot is in Figure S14 (see Supplement 1). We report this SNP only a suggestive finding. Using Snipper(43), we found no evidence in the literature that this region has previously been linked to relevant phenotypes.

## Burden Tests

The number of genes with sufficient variation to enable burden testing ranged from 15,761 to 16,195 depending on the phenotype, resulting in Bonferroni corrections of $p < 3.2e^{-6}$. At this threshold no gene was significantly associated with any of the phenotypes.

## Candidate Genes

Detailed results for the 151 candidate genes are presented in Table S5 (see Supplement 2). Even when the Bonferroni threshold is relaxed to .0003 for 151 tests, no gene was significant. For each phenotype, the distribution of p-values across all 151 genes was uniform, with means within a few percent of .50.

We conducted 116 single variant tests of 78 SNPs from a "high priority" list of 11 genes that harbor a variant identified by GWAS meta-analyses of smoking or drinking. After correcting for only 116 tests, only one rare (MAC=6) nonsynonymous variant at 9:136521726 in *DBH* was significantly associated with the nicotine (exposed) phenotype. The remaining p-values were uniformly distributed and non-significant. The full table and QQ plot of 116 tests are in Table S4 and Figure S10 (see Supplement 1).

## SNP-based Heritability

Results from GCTA are displayed in Table **2**. In aggregate, rare nonsynonymous SNPs accounted for 19% of the heritability in illicit drugs (*p*=.01), but did not significantly account for variance in any other phenotype. Marginal effects were observed for aggregate nonsynonymous effects on behavioral disinhibition and alcohol dependence. Common SNPs, on the other hand, accounted for 27% and 26% of the variance in illicit drugs and behavioral disinhibition, which is less than half of the pedigree-based heritability of those traits. When considering all variants, the SNPs significantly account for 46% and 40% of the variance in illicit drugs and behavioral disinhibition. When considering the heritability of these traits is 55% and 58%, respectively, we find that the combination of rare nonsynonymous and genome-wide common variants account for 84% of the heritability in the illicit drug phenotype, and 69% of the behavioral disinhibition phenotype.

## Discussion

We report the results from a genetic association test between nearly 100,000 rare nonsynonymous exonic SNPs with 5 measures of behavioral disinhibition, as well as three measures of substance use in individuals exposed to the relevant substance. Tests of genetic association, whether single variant tests or gene-based burden tests, revealed no significant associations. Additional inspection of 172 candidate genes, of which 151 contained polymorphic variation in this sample, also revealed no significant association. One SNP was statistically significant only after restricting tests to nonsynonymous variants within genes previously implicated in drinking and smoking by GWAS meta-analysis. This variant was associated with nicotine dependence among individuals exposed to nicotine. The variant is located in the dopamine beta-hydroxylase gene (*DBH*), a long-standing candidate gene for addiction due to dopamine's clear involvement in the rewarding effects of drugs of abuse(44). A variant within this gene was previously found to be associated with smoking cessation by the TAG consortium(22). Our nicotine phenotype is not cessation per se and that, along with the nominally-associated p-value, suggests that this is a spurious finding. This variant will require substantial verification in larger samples.

One limitation of the association tests is in the nicotine phenotype selection. It is known, for example, that variants within nicotinic receptor genes on chromosome 15 influence heaviness of smoking in smokers. Our nicotine dependence (exposed) phenotype is a measure of heaviness of smoking among individuals who have ever smoked, but the lowest p-value within nicotinic receptor genes was *p*=.01 for rs16969968, a variant known to be associated with heaviness of smoking in current smokers(21, 22). One possible explanation for our nonsignificant finding is that our exposure criteria were insufficient. This study did not ask, for example, whether an individual had ever smoked 100+ cigarettes, a common proxy for measures of lifetime regular smoking. An additional complication is that approximately half of the sample reported here is around 18 years old, and common SNPs in nicotinic receptor genes such as rs16969968 appear to influence smoking more strongly in older individuals(45, 46). It is also possible that we may have missed strong associations with nicotine dependence because we did not evaluate additional nicotine phenotypes due to our focus on the behavioral disinhibition hypothesis.

The aggregate effect of rare nonsynonymous SNPs was significant for the illicit drug phenotype, with marginal effects for the other phenotypes. Most applications of GCTA rely on the relationship between tagged SNPs on a genome-wide array and untyped causal variants. For rare nonsynonymous variants we anticipate only very weak, or nonexistent, linkage disequilibrium, which allows the tentative conclusion that nonsynonymous variants typed on the exome chip are directly associated with illicit drug use in this community-representative sample, and awaits replication elsewhere.

There are several alternative explanations for the GCTA results. First, for rare variants it may not always be true that global ancestry measures (such as PCs) can account for genomically-local population stratification. That is, two individuals may have similar overall ancestry but possess some haplotypes from diverse ancestries(47, 48), which may artificially inflate GCTA estimates from local ancestry around some exons. If that confound existed in

this data we might expect population stratification to inflate all GCTA estimates from rare variants, which was not observed. Second, the illicit drug phenotype has the strongest mother-father correlation (.56; Table S2 in Supplement 1), raising the possibility of confounding due to assortative mating. Note, however, that substantial mother-father correlations exist for other phenotypes (notably .40 for alcohol consumption) but these showed no GCTA effect for rare nonsynonymous variants. Third, we speculate that rare nonsynonymous variation may contribute strongly to more severe phenotypes. It appears, for example, that the strongest nonsynonymous GCTA results were for illicit drugs, behavioral disinhibition, and alcohol dependence, and lesser effects observed for nicotine dependence and alcohol consumption, the latter two considered to be relatively more normative behaviors. Future work may consider this possibility.

Different sampling approaches, denser genotyping (e.g., sequencing), and larger samples will be required to identify rare variant associations with the phenotypes investigated here. Larger (meta-) samples will allow us to draw stronger conclusions about genetic etiology in addiction and behavioral disinhibition and we are excited to contribute to consortium efforts to achieve these ends.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

1. Yang JA, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nature Genet. 2011; 43(6): 519–U44. [PubMed: 21552263]

2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. [PubMed: 19812666]

3. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56–65. Epub 2012/11/07. [PubMed: 23128226]

4. Fu WQ, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013; 493(7431):216–220. [PubMed: 23201682]

5. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012; 337(6090):64–69. Epub 2012/05/19. [PubMed: 22604720]

6. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genetics. 2012; 8(8):e1002793. Epub 2012/08/10. [PubMed: 22876189]

7. Hicks BM, Schalet BD, Malone SM, Iacono WG, McGue M. Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. Behav Genet. 2011; 41(4):459–475. Epub 2010/12/15. [PubMed: 21153693]

8. McGue M, Zhang Y, Miller MB, Basu S, Vrieze SI, Hicks BM, et al. A Genome-Wide Association Study of Behavioral Disinhibition. In press at Behavior Genetics. 2013

9. Vrieze SI, McGue M, Miller MB, Hicks BM, Iacono WG. Three Mutually Informative Ways to Understand the Genetic Relationships Among Behavioral Disinhibition, Alcohol Use, Drug Use, Nicotine Use/Dependence, and Their Co-occurrence: Twin Biometry, GCTA, and Genome-Wide Scoring. Behav Genet. 2013; 43(2):97–107. Epub 2013/01/31. [PubMed: 23362009]

10. Gorenstein EE, Newman JP. Disinhibitory psychopathology: A new perspective and model for research. Psychological Review. 1980; 87(3):301–315. [PubMed: 7384345]

11. Iacono WG, Malone SM, McGue M. Behavioral disinhibition and the development of early-onset addiction: Common and specific influences. Annu Rev Clin Psycho. 2008; 4:325–348.

12. Sher KJ, Bartholow BD, Wood MD. Personality and substance use disorders: a prospective study. J Consult Clin Psych. 2000; 68(5):818–829. Epub 2000/11/09.

13. Quay HC. Inhibition and attention deficit hyperactivity disorder. Journal of Abnormal Child Psychology. 1997; 25(1):7–13. [PubMed: 9093895]

14. Young SE, Stallings MC, Corley RP, Krauter KS, Hewitt JK. Genetic and environmental influences on behavioral disinhibition. Amercan Journal of Medical Genetics. 2000; 96(5):684–695. Epub 2000/10/31.

15. Vollebergh WAM, Iedema J, Bijl RV, de Graaf R, Smit F, Ormel J. The structure and stability of common mental disorders - The NEMESIS Study. Archives of General Psychiatry. 2001; 58(6):597–603. [PubMed: 11386990]

16. 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 Psychiat. 2003; 160(4):687–695. [PubMed: 12668357]

17. Vrieze SI, Hicks BM, Iacono WG, McGue M. Decline in Genetic Influence on the Co-Occurrence of Alcohol, Marijuana, and Nicotine Dependence Symptoms From Age 14 to 29. Am J Psychiat. 2012 Epub 2012/09/18.

18. Hicks BM, Krueger RF, Iacono WG, McGue M, Patrick CJ. Family transmission and heritability of externalizing disorders: A twin-family study. Arch Gen Psychiat. 2004; 61:922–928. [PubMed: 15351771]

19. Tielbeek JJ, Medland SE, Benyamin B, Byrne EM, Heath AC, Madden PA, et al. Unraveling the genetic etiology of adult antisocial behavior: a genome-wide association study. Plos One. 2012; 7(10):e45086. Epub 2012/10/19. [PubMed: 23077488]

20. Dick DM, Aliev F, Krueger RF, Edwards A, Agrawal A, Lynskey M, et al. Genome-wide association study of conduct disorder symptomatology. Mol Psychiatr. 2011; 16(8):800–808.

21. Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen XN, Cichon S, et al. Multiple Independent Loci at Chromosome 15q25.1 Affect Smoking Quantity: a Meta-Analysis and Comparison with Lung Cancer and COPD. Plos Genetics. 2010; 6(8)

22. Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature Genet. 2010; 42(5):441–U134. [PubMed: 20418890]

23. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nature Genet. 2010; 42(5):436–U75. [PubMed: 20418889]

24. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nature Genet. 2010; 42(5):448–U135. [PubMed: 20418888]

25. Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. P Natl Acad Sci USA. 2011; 108(17):7119–7124.

26. Miller MB, Basu S, Cunningham J, Eskin E, Malone SM, Oetting WS, et al. The Minnesota Center for Twin and Family Research genome-wide association study. Twin Research and. Human Genetics. 2012; 15(6):767–774.

27. Robins, LN.; Babor, TF.; Cottler, LB. Composite International Diagnostic Interview: Expanded Substance Abuse Module. St. Louis: Authors; 1987.

28. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiat. 1988; 45(12):1069–1077. [PubMed: 2848472]

29. Zhan XW. ANNO. 2013 https://github.com/zhanxw/anno:github.com.

30. Hodgkinson CA, Yuan QP, Xu K, Shen PH, Heinz E, Lobos EA, et al. Addictions biology: Haplotype-based analysis for 130 candidate genes on a single array. Alcohol Alcoholism. 2008; 43(5):505–515. [PubMed: 18477577]

31. Saccone SF, Bierut LJ, Chesler EJ, Kalivas PW, Lerman C, Saccone NL, et al. Supplementing High-Density SNP Microarrays for Additional Coverage of Disease-Related Genes: Addiction as a Paradigm. Plos One. 2009; 4(4)

32. Luczak SE, Glatt SJ, Wall TL. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians. Psychol Bull. 2006; 132(4):607–621. [PubMed: 16822169]

33. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nature Genet. 2010; 42(4):348-U110. [PubMed: 20208533]

34. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. Nat Methods. 2011; 8(10):833-U94. [PubMed: 21892150]

35. Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet. 2010; 11(11):773–785. [PubMed: 20940738]

36. Asimit J, Zeggini E. Rare variant association analysis methods for complex traits. Annual review of genetics. 2010; 44:293–308. Epub 2010/11/05.

37. Li BS, Leal SM. Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. Am J Hum Genet. 2008; 83(3):311–321. [PubMed: 18691683]

38. Price AL, Kryukov GV, de Bakker PIW, Purcell SM, Staples J, Wei LJ, et al. Pooled Association Tests for Rare Variants in Exon-Resequencing Studies. Am J Hum Genet. 2010; 86(6):832–838. [PubMed: 20471002]

39. Wu MC, Lee S, Cai TX, Li Y, Boehnke M, Lin XH. Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. Am J Hum Genet. 2011; 89(1):82–93. [PubMed: 21737059]

40. Schaid DJ. Genomic Similarity and Kernel Methods II: Methods for Genomic Information. Hum Hered. 2010; 70(2):132–140. [PubMed: 20606458]

41. Schaid DJ. Genomic Similarity and Kernel Methods I: Advancements by Building on Mathematical and Statistical Foundations. Hum Hered. 2010; 70(2):109–131. [PubMed: 20610906]

42. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88(1):76–82. Epub 2010/12/21. [PubMed: 21167468]

43. Welch R. Snipper. https://github.com/welchr/Snipper2013.

44. Volkow ND, Fowler JS, Wang GJ, Goldstein RZ. Role of dopamine, the frontal cortex and memory circuits in drug addiction: Insight from imaging studies. Neurobiol Learn Mem. 2002; 78(3):610–624. [PubMed: 12559839]

45. Vrieze SI, McGue M, Iacono WG. The interplay of genes and adolescent development in substance use disorders: leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. Hum Genet. 2012; 131(6):791–801. [PubMed: 22492059]

46. Belsky DW, Moffitt TE, Baker TB, Biddle AK, Evans JP, Harrington H, et al. Polygenic Risk and the Developmental Progression to Heavy, Persistent Smoking and Nicotine Dependence: Evidence From a 4-Decade Longitudinal Study. JAMA psychiatry. 2013:1–9. Epub 2013/03/29.

47. Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. Nature Genet. 2012; 44(3):243-U29. [PubMed: 22306651]

48. Listgarten J, Lippert C, Heckerman D. FaST-LMM-Select for addressing confounding from spatial structure and rare variants. Nature Genet. 2013; 45(5):470–471. [PubMed: 23619783]

**Table 1**

Non-synonymous Variants on Exome Chip and Distribution in Current Study

| Type of Variant | Total Polymorphic | Total Number of SNPs by MAF | | | | | | Het/hom rare genotypes per sample | |
|---|---|---|---|---|---|---|---|---|---|
| | | <.1% | .1 – .5% | .5–1% | 1–5% | >5% | Mean (SD) | Range |
| All Non-Synonymous | 111,592 | 65829 | 18174 | 5099 | 8435 | 14055 | 5931 (84.4) | [5615, 6465] |
| Stop Loss | 326 | 168 | 47 | 8 | 22 | 81 | 39 (4.4) | [23, 54] |
| Stop Gain | 2336 | 1670 | 351 | 78 | 97 | 140 | 59 (5.9) | [40, 80] |
| Start Loss | 303 | 175 | 45 | 16 | 28 | 39 | 16 (3.1) | [6, 30] |
| Start Gain | 8 | 6 | 1 | 0 | 1 | 0 | 0 (.2) | [0, 1] |
| Essential Splice | 969 | 578 | 128 | 37 | 35 | 191 | 79 (6.5) | [57, 103] |
| Normal Splice | 793 | 35 | 19 | 6 | 50 | 683 | 297 (12.1) | [253, 338] |

Note: All non-synonymous SNPs includes start/stop and splice-site variants. All variants were annotated relative to GENCODE v11 transcripts from the ENCODE project. Minor allele frequency was computed without respect to founder status.

**Table 2**

Heritability Analysis

| Trait | Pedigree Heritability | | SNP-based Heritability | | | | | | Heritability Accounted for by SNPs |
|---|---|---|---|---|---|---|---|---|---|
| | | | VC1: Rare Nonsyn SNPs (97,537) | | | VC2: Common SNPs (496,056) | | VC1+VC2 | |
| | Estimate | N | Estimate (SE) | p-value | N | Estimate (SE) | p-value | | |
| Nicotine | .49 | 7181 | 0 (.08) | .5 | 3644 | .14 (.09) | .06 | .14 | 29% |
| Alcohol Consumption | .48 | 7181 | 0 (.08) | .5 | 3644 | .12 (.09) | .09 | .12 | 25% |
| Alcohol Dependence | .58 | 7181 | .10 (.07) | .08 | 3644 | .12 (.09) | .08 | .22 | 38% |
| Illicit Drugs | .55 | 7181 | .19 (.09) | .01 | 3644 | .27 (.09) | .001 | .46 | 84% |
| Behavioral Disinhibition | .58 | 7181 | .14 (.08) | .05 | 3644 | .26 (.09) | .002 | .40 | 69% |
| Nicotine (Exposed) | .48 | 3412 | .11 (.12) | .17 | 1774 | .07 (.19) | .35 | .18 | 38% |
| Alcohol Consumption (Exposed) | .42 | 5897 | 0 (.08) | .5 | 3239 | .11 (.10) | .14 | .11 | 26% |
| Alcohol Dependence (Exposed) | .50 | 5897 | .12 (.09) | .07 | 3239 | .08 (.10) | .22 | .20 | 40% |

Note: Pedigree heritability refers to the proportion of phenotypic variance accounted for by kinships in the FaST-LMM mixed model of the full sample. The GRM cutoff was .025, which excludes closely related individuals (e.g., less than second cousins). VC1 and VC2 are variance components estimated jointly. Rare nonsynonymous SNPs were defined as having MAF < .05 (see Table **1**). Common SNPs were all SNPs with MAF > .05. More detailed information on the effect of the GCTA cutoff for Illicit Drugs and Behavioral Disinhibition is provided in Figure S12 (see Supplement 1).