

# The Nuclear Transcription Factor PKNOX2 Is a Candidate Gene for Substance Dependence in European-Origin Women

Xiang Chen<sup>1</sup>✉, Kelly Cho<sup>1</sup>✉, Burton H. Singer<sup>2</sup>, Heping Zhang<sup>1\*</sup>

**1** Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut, United States of America, **2** Emerging Pathogens Institute, University of Florida, Gainesville, Florida, United States of America

## Abstract

Substance dependence or addiction is a complex environmental and genetic disorder that results in serious health and socio-economic consequences. Multiple substance dependence categories together, rather than any one individual addiction outcome, may explain the genetic variability of such disorder. In our study, we defined a composite substance dependence phenotype derived from six individual diagnoses: addiction to nicotine, alcohol, marijuana, cocaine, opiates or other drugs as a whole. Using data from several genomewide case-control studies, we identified a strong (Odds ratio = 1.77) and significant ( $p$ -value =  $7E-8$ ) association signal with a novel gene, PBX/knotted 1 homeobox 2 (PKNOX2), on chromosome 11 with the composite phenotype in European-origin women. The association signal is not as significant when individual outcomes for addiction are considered, or in males or African-origin population. Our findings underscore the importance of considering multiple addiction types and the importance of considering population and gender stratification when analyzing data with heterogeneous population.

**Citation:** Chen X, Cho K, Singer BH, Zhang H (2011) The Nuclear Transcription Factor PKNOX2 Is a Candidate Gene for Substance Dependence in European-Origin Women. *PLoS ONE* 6(1): e16002. doi:10.1371/journal.pone.0016002

**Editor:** Vladimir Uversky, University of South Florida College of Medicine, United States of America

**Received:** October 14, 2010; **Accepted:** December 1, 2010; **Published:** January 27, 2011

**Copyright:** © 2011 Chen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported in part by National Institutes of Health Grants K02DA017713, R01DA016750, and T32 MH014235. Study of Addiction: Genetics and Environment (SAGE) was supported by National Institutes of Health Genes, Environment, and Health Initiative Grant U01 HG004422. Study of Addiction: Genetics and Environment is one of the genome-wide association studies funded as part of the Gene Environment Association Studies under Genes, Environment, and Health Initiative. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies Coordinating Center, which is supported by National Institutes of Health Grant U01 HG004446. Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism, which is supported by National Institutes of Health Grant U10 AA008401; the Collaborative Genetic Study of Nicotine Dependence, which is supported by National Institutes of Health Grant P01 CA089392; and the Family Study of Cocaine Dependence which is supported by National Institutes of Health Grant R01 DA013423. Genotyping was performed at the Johns Hopkins University Center for Inherited Disease Research and was supported by National Institutes of Health Genes, Environment, and Health Initiative Grant U01HG004438, the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and National Institutes of Health contract HHSN268200782096C, "High throughput genotyping for studying the genetic contributions to human disease." The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: heping.zhang@yale.edu

✉ These authors contributed equally to this work.

## Introduction

Substance dependence or addiction is one of the most sought-after phenomena in many populations because of its serious health and socio-economic consequences. In 2008, the Centers for Disease Control estimated that 443,000 deaths were caused by cigarette smoking and exposure to secondhand smoke [1]. In addition, alcohol misuse has been linked to attempted and successful suicide, particularly among adolescents [2]. There is strong evidence that vulnerability to substance dependence to drugs, alcohol, or smoking is a complex trait with both genetic and environmental components [3,4,5]. Therefore, a better understanding of the genetics behind vulnerability to addictions could tremendously improve overall health and quality of life in general. A useful start in this direction is given by Kreek et al. [6] In the literature, candidate genes for addiction to individual substances (alcohol, nicotine and other substances) have been identified. For example, well studied genes for alcohol dependence, such as

GABRA2, CHRM2 and ADH4, have been replicated in many samples [7,8,9,10], while several newer candidate genes (GABRG3, TAS2R16, SNCA, OPRK1 and PDYN) remain to be confirmed [11]. Multiple variants at the aldehyde dehydrogenase (ALDH) and alcohol dehydrogenase (ADH) loci have also been well documented as genes of major genetic effect especially in East-Asian populations [12,13,14,15]. A gene cluster of nicotinic acetylcholine receptors (CHRNA5, CHRNA3, and CHRNA4) and Neurexin1, also show allelic differences in heavy vs. light smokers in multiple studies [16,17,18,19]. Li [20] reported thirteen regions on chromosomes 3–7, 9–11, 17, 20, and 22, to be significantly associated with nicotine dependence in at least two independent samples, although a significant number of reported genomic regions did not reach the level of "suggestive" or "significant" linkage and failed to be replicated in other independent studies.

In the past, much effort has been devoted to the emphasis on individually different substance dependence outcomes. However,

substance dependence as a whole, combining addiction to nicotine, alcohol, marijuana, cocaine, opiates and other drugs, has not been thoroughly investigated in association studies. A composite substance dependence phenotype may be the key to finding a common genetic predisposition of substance dependence as a whole. This common genetic predisposition may not be apparent when individual addiction conditions are considered. In the literature, Li and Burmeister [21] provide a good review of comorbidity in the genetics of addiction. The availability of the Gene Environment Association Studies Genes and Environment Initiative Study of Addiction: Genetics and Environment (SAGE) data provides an unprecedented opportunity to study the genetics of a composite trait: namely, addiction to at least two of the six substances under study (nicotine, alcohol, marijuana, cocaine, opiates and other drugs).

In this report, we present a genomewide significant association ( $\alpha = 0.05$ ) of PKNOX2 gene on chromosome 11 with composite substance dependence in European-origin women. We have identified a cluster of markers in the region of PKNOX2 gene that are strongly associated with a composite addiction phenotype rather than with any single addiction type. Furthermore, we investigate potential sex-specificity and racial differences in the association. The nuclear transcription factor PKNOX2 has been previously identified as one of the *cis*-regulated genes for alcohol addiction in mice [22]. However, to our knowledge, PKNOX2 has not been reported to be associated with any substance addiction outcomes in human populations. Thus we present PKNOX2 as a novel candidate gene for substance dependence in humans.

## Methods

### Study of Addiction: Genetics and Environment (SAGE)

#### Data

We obtained the genomewide single nucleotide polymorphisms (SNP) data from the database of Genotype and Phenotype (dbGaP). The data were from the Study of Addiction: Genetics and Environment (SAGE) (Bierut et al. 2010). We included 4,121 subjects for whom the addiction to the six categories of substances and genomewide SNP data (ILLUMINA Human 1M platform) were available. SAGE is a case-control study of mostly unrelated individuals aimed at identifying genetic associations for addiction. Cases and controls were selected from three large, complementary cohorts: Collaborative Study on the Genetics of Alcoholism (COGA, initiated in 1989), Family Study of Cocaine Dependence

(FSCD, 2000–2006), and Collaborative Genetic Study of Nicotine Dependence (COGEN, initiated in 2000). These three studies have been previously described [8,23,24,25]. Lifetime dependence on nicotine, alcohol, marijuana, cocaine, opiates or other dependence on other drugs was diagnosed in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). As stated above, we studied a composite addiction phenotype according to whether a subject was addicted to substances in at least two of the aforementioned categories.

### Subject Characteristics and Study Design

To reduce the level of noise in genotypes and increase the efficiency of analysis, we filtered SNPs by setting thresholds for minor allele frequency (MAF) and call rate (i.e., MAF >5% and call rate >90%). In addition, we excluded 60 duplicate genotype samples and removed 9 subjects with ethnic backgrounds other than African-origin (Black) or European-origin (White). Table 1 lists the descriptive statistics of the sample included in our study. Almost all subjects were either Black (30.3%) or White (69.4%).

In our final analysis, a total of 3,627 unrelated subjects with 830,696 autosomal SNPs were included. The final subset considered in the analysis consisted of 45.9% men and 54.1% women with mean ages of 39.4 and 38.6 years old, respectively. Because substance dependence is a complex disease with both genetic and environmental components, we analyzed the male and female subsets separately. In addition, we performed separate analysis for Blacks and Whites in light of the possibility that underlying genetic variants may be different in different ethnic groups. Thus our study included four sub-samples: 1,393 White women, 1,131 White men, 568 Black women and 535 Black men. Overall a total of 1,513 subjects were defined as having two or more substance addictions according to DSM-IV. Of these, there were 316, 585, 237 and 375 subjects in the Black male, White male, Black female and White female subsets, respectively. The proportions of subjects diagnosed with lifetime dependence on substances in each of the six categories – nicotine, alcohol, marijuana, cocaine, opiates or other drugs – are presented in Table 1. The top three most widely used substances among the six were alcohol, nicotine and cocaine, in that order (Figure 1).

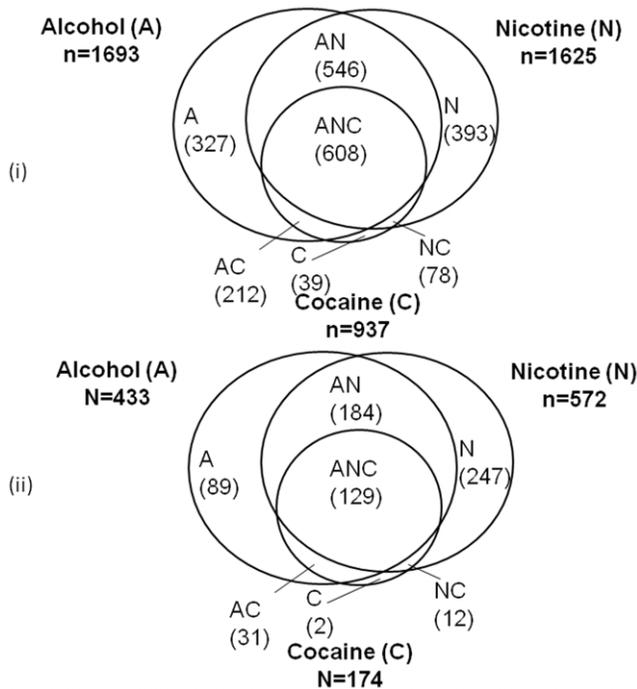
### Statistical Analysis

We took precaution to investigate and control for potential confounding by stratification and admixture due to disequilibrium

**Table 1.** Descriptive statistics of the sample stratified by sex and race.

	Black Men	White Men	Black Women	White Women	Overall
n	535	1131	568	1393	3627
Age (SD) yr.	40.9 (8.2)	38.7 (10.3)	39.7 (6.7)	38.2 (9.1)	39.0 (9.1)
Height (SD) m	1.78 (0.08)	1.79 (0.07)	1.64 (0.08)	1.65 (0.07)	1.71 (0.10)
Weight (SD) kg	89.7 (18.3)	88.7 (16.8)	85.0 (21.9)	72.4 (18.8)	81.1 (20.1)
Alcohol (%)	62.1	62.3	39.4	31.1	46.7
Cocaine (%)	46.4	27.3	36.3	12.5	25.8
Marijuana (%)	25.4	25.2	13.7	8.7	17.1
Nicotine (%)	47.5	46.7	47.7	41.1	44.8
Opiates (%)	8.2	9.9	6.2	4.8	7.1
Other drugs (%)	11.4	18.0	6.5	9.4	11.9
No drug (%)	27.1	31.0	38.9	50.1	39.0

doi:10.1371/journal.pone.0016002.t001



**Figure 1. Number of substance dependent subjects according to DSM-IV for the top three addiction categories: alcohol (A), nicotine (N) and cocaine (C).** (i) is based on the overall sample and (ii) is based on the White women subset. doi:10.1371/journal.pone.0016002.g001

between pairs of unlinked loci [26] given the different ethnic populations in our data. To avoid potential population stratification, we first stratified our analysis by race and sex. Then we performed formal population stratification analysis for each subset using PLINK software (version 1.04) [27]. The results confirmed that each subset comes from a homogenous population; thus no further adjustment was needed to control for potential confounding by stratification and admixture.

We used allelic Chi-square tests with 1 degree of freedom in our analysis, stratified by race and sex. Haploview (version 4.0) [28] was used to analyze the linkage disequilibrium in the PKNOX2 gene region and the association between the haplotypes and the composite phenotype. We performed additional analyses by examining and comparing the results of including and excluding 214 related subjects in the data. For mixtures of unrelated and related subjects, we used PedGenie [29] to perform association analyses. PedGenie first performs the allelic Chi-square tests treating all individuals independently, then takes the pedigree information into account to assess statistical significance through permutation analysis.

#### Determination of Significance Threshold for GWAS

Using genotypes from the Wellcome Trust Case-Control Consortium, Dudbridge and Gusnanto [30] studied the genome-wide significance threshold for the UK Caucasian population. They subsampled the genotypes at different densities and estimated the threshold for 5% family-wise error using permutations. They then extrapolated to infinity density and estimated that the genome-wide significance threshold for this population is 7.2E-8. We used this genome-wide significance threshold (7.2E-8) for the Caucasian population (white men and white women) in our analysis.

## Results

The top 8 significant SNPs are summarized in Table 2. They cluster in PKNOX2 on chromosome 11 (11q24). None of the 8 SNPs violates the Hardy Weinberg equilibrium assumption (minimum *p*-level = 0.12). Among them, rs12284594 is the most significant SNP (*p*-value = 7.13E-08) observed in White women with an odds ratio (OR) of 1.77, suggesting that those who have the risk allele (G) for rs12284594 are at significantly increased risk of being diagnosed with at least two of the six categories of substance dependence. This *p*-value reaches the accepted genome-wide significance level [30]. In addition, there are 7 other SNPs with *p*-values less than 3.8E-06 in White women with similar ORs (1.63 - 1.72). We further examined association of haplotypes with the composite phenotype in this region, but they did not enhance the strength of the associations; hence these results are not reported here. Similarly, when related individuals were included in the analysis, the strength of association was not enhanced, whether the analysis was performed using PedGenie [29], or whether the correlation among related individuals was ignored (data not shown). Although we also observed that these 8 SNPs confer increased risk in White men, Black men and Black women, they fail to reach genome-wide significance. Hence, detailed results are not presented here for these groups.

We performed additional analyses to examine each substance dependence outcome separately for the top 8 SNPs presented above. Table 3 shows the corresponding *p*-values for the 8 SNPs for each substance dependence outcome. Alcohol dependence shows the strongest association (*p*-value = 1.97E-6 with rs12284594); however none of these *p*-values attains the genome-wide significance level of 0.05.

## Discussion

We have found a novel, genome-wide significant association of a composite substance dependence phenotype with a SNP in the PKNOX2 gene in White women. PKNOX2, PBX/knotted 1 homeobox 2, belongs to the three-amino-acid loop extension (TALE) homeobox family. Homeodomain proteins are highly conserved transcription regulators. Imoto et al. [31] identified PKNOX2 as a novel TALE homeodomain-encoding gene, located at 11q24 in humans and it functions as a nuclear transcription factor indicated by its structure and sub-cellular localization. Later, PKNOX2 was identified as one of the *cis*-regulated genes for alcohol addiction in mice [22]. However,

**Table 2. Summary of the 8 most significant SNPs in PKNOX2 gene showing genome-wide significant association with substance dependence in White women.**

SNP	P-value	OR
rs1426153 (G)	1.84E-06	1.66
rs11220015(A)	1.97E-06	1.65
rs11602925(G)	1.24E-06	1.67
rs750338(C)	4.22E-07	1.63
rs12273605(T)	3.83E-06	1.71
rs10893365(C)	2.27E-07	1.72
rs10893366(T)	6.87E-07	1.70
<b>rs12284594(G)</b>	<b>7.13E-08</b>	<b>1.77</b>

The high-risk allele is in the parenthesis. doi:10.1371/journal.pone.0016002.t002

**Table 3.** Associations of the 8 most significant SNPs in PKNOX2 with six individual substance dependence outcomes (p-values).

SNP	Nicotine	Alcohol	Marijuana	Cocaine	Opiates	Others
rs1426153 (G)	0.0159	5.75E-5	7E-4	3E-4	0.0113	1E-4
rs11220015(A)	0.0163	6.86E-5	0.0010	3E-4	0.0037	4.18E-5
rs11602925(G)	0.0136	4.24E-5	7E-4	3E-4	0.0059	5.29E-5
rs750338(C)	0.0491	4.26E-5	0.0013	2E-4	0.0112	2.22E-5
rs12273605(T)	0.0921	3E-4	3.53E-5	1E-4	0.0680	3.11E-5
rs10893365(C)	0.0411	1.72E-5	8.58E-6	2.91E-5	0.0699	2.58E-5
rs10893366(T)	0.0621	1.37E-5	8.80E-6	8.63E-5	0.0905	5.35E-5
<b>rs12284594(G)</b>	0.0239	1.97E-6	8.54E-6	4.39E-5	0.0533	2.45E-5

doi:10.1371/journal.pone.0016002.t003

PKNOX2 has not been reported to be associated with any substance dependence phenotype in humans to date.

The composite dichotomous substance dependence variable reflects cases with two or more addictions where the top three categories are alcohol, nicotine and cocaine (Figure 1). Among the cases, 47% have been diagnosed with alcohol dependence in combination with other substance dependence outcomes. Our results, which show a strong association of this composite substance dependence variable with PKNOX2 gene in a human sample, support the experimental findings in mice by Mulligan et al [22]. Thus our findings make an important contribution in reporting PKNOX2 as a novel candidate gene for substance dependence in humans, particularly for White women in the SAGE sample.

Interestingly, among our most significant SNPs, we do not observe those genes previously reported for alcoholism or nicotine. Rather we find a new set of genes among the top SNPs. When each substance dependence outcome was individually analyzed for association with the 8 most significant SNPs, we found no association that reached the genomewide significance. This suggests that substance dependence or addiction as a whole has different risk genes compared to any single addiction outcome. It may also mean that there is more power in detecting common genes acting upon co-morbid addiction outcomes as a whole.

For many complex diseases, different ethnic groups have vastly different underlying genetics, and these differences may confound association results when they are pooled together as one in the analysis. Previously, racial differences in the prevalence of substance abuse have been reported [32,33,34]. More recently, Luo et al. [24] have reported that genetic differences between Black smokers and White smokers influence the nature of their nicotine dependence. Their analysis suggested that Black smokers become dependent at a lower threshold (number cigarettes per day) than Whites. On the other side, in the presence of subjects in different ethnic populations in the data, it is crucial to investigate and control for potential confounding by stratification and admixture due to disequilibrium between pairs of unlinked loci [26]. Thus we investigated these two major ethnic groups separately in our analysis. In addition, we stratified our analysis by gender; based on the premise that gender may be a confounding factor for the substance dependence outcome – men may be socially more

prone to environmental influences promoting substance use, and thus more vulnerable to addiction, compared to women [35]. Our results from the two ethnic groups do not corroborate each other, which underscores the underlying genetic differences in White and Black samples. In fact, strong association signals are observed only in the White woman sample. With a heterogeneous population like SAGE, one must be cautious in analyzing and interpreting the results.

The identification of PKNOX2 as a candidate gene for substance use disorders underscores two important issues: (a) this has not been possible in the past due to limited sample size; and (b) we have considered a composite trait of six substance dependence outcomes as a whole. The association becomes less significant if individual substance addictions are considered. Thus, this result highlights the importance of studying highly comorbid disorders or those which might otherwise have a common pathway. However, our study is limited to the information in the available data, and we acknowledge the difficulty in operationalizing substance dependence; whether our operationalization of addiction to two or more substances, truly reflects the strength of the addiction phenotype is open to question. Indeed, it may simply reflect the extent of access to drugs. We also recognize that dependence on one substance shows different characteristics from dependence on another, and it is valuable and necessary to study them as individual entities. However, our call for more attention to comorbidity and the combinatorial study of these disorders should be viewed as a valuable complementary effort.

## Acknowledgments

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

## Author Contributions

Conceived and designed the experiments: XC KC HZ. Performed the experiments: XC KC HZ. Analyzed the data: XC KC HZ. Contributed reagents/materials/analysis tools: XC KC HZ. Wrote the paper: XC KC BHS HZ.

## References

- Centers for Disease Control and Prevention (CDC) (2009) State-specific smoking-attributable mortality and years of potential life lost—United States, 2000–2004. *MMWR Morb Mortal Wkly Rep* 58: 29–33.
- Makhija NJ, Sher L (2007) Preventing suicide in adolescents with alcohol use disorders. *Int J Adolesc Med Health* 19: 53–59.
- Uhl GR, Elmer GI, Lanuda MC, Pickens RW (1995) Genetic influences in drug abuse. In: Bloom FE, Kupfer DJ, eds. *Psychopharmacology: The Fourth Generation of Progress*. New York: Lippincott Williams & Wilkins. pp 1793–1806.
- True WR, Heath AC, Scherrer JF, Xian H, Lin N, et al. (1999) Interrelationship of genetic and environmental influences on conduct disorder and alcohol and marijuana dependence symptoms. *Am J Med Genet* 88: 391–397.
- Merikangas KR, Stolar M, Stevens DE, Goulet J, Preisig MA, et al. (1998) Familial transmission of substance use disorders. *Arch Gen Psychiatry* 55: 973–979.
- Kreek MJ, Nielsen DA, Butelman ER, LaForge KS (2005) Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. *Nat Neurosci* 8: 1450–1457.

7. Reich T (1996) A genomic survey of alcohol dependence and related phenotypes: results from the Collaborative Study on the Genetics of Alcoholism (COGA). *Alcohol Clin Exp Res* 20: 133A–137A.
8. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, et al. (1998) Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet* 81: 207–215.
9. Song J, Koller DL, Foroud T, Carr K, Zhao J, et al. (2003) Association of GABA(A) receptors and alcohol dependence and the effects of genetic imprinting. *Am J Med Genet B Neuropsychiatr Genet* 117B: 39–45.
10. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, et al. (2004) Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. *Am J Hum Genet* 74: 705–714.
11. Edenberg HJ, Foroud T (2006) The genetics of alcoholism: identifying specific genes through family studies. *Addict Biol* 11: 386–396.
12. Shen YC, Fan JH, Edenberg HJ, Li TK, Cui YH, et al. (1997) Polymorphism of ADH and ALDH genes among four ethnic groups in China and effects upon the risk for alcoholism. *Alcohol Clin Exp Res* 21: 1272–1277.
13. Higuchi S, Matsushita S, Imazeki H, Kinoshita T, Takagi S, et al. (1994) Aldehyde dehydrogenase genotypes in Japanese alcoholics. *Lancet* 343: 741–742.
14. Maezawa Y, Yamauchi M, Toda G, Suzuki H, Sakurai S (1995) Alcohol-metabolizing enzyme polymorphisms and alcoholism in Japan. *Alcohol Clin Exp Res* 19: 951–954.
15. Nakamura K, Iwahashi K, Matsuo Y, Miyatake R, Ichikawa Y, et al. (1996) Characteristics of Japanese alcoholics with the atypical aldehyde dehydrogenase 2\*2. I. A comparison of the genotypes of ALDH2, ADH2, ADH3, and cytochrome P-4502E1 between alcoholics and nonalcoholics. *Alcohol Clin Exp Res* 20: 52–55.
16. Bierut IJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, et al. (2007) Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 16: 24–35.
17. Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, et al. (2007) Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 16: 36–49.
18. Berrettini W, Yuan X, Tozzi F, Song K, Francks C, et al. (2008) Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry* 13: 368–373.
19. Bierut IJ (2007) Genetic variation that contributes to nicotine dependence. *Pharmacogenomics* 8: 881–883.
20. Li MD (2008) Identifying susceptibility loci for nicotine dependence: 2008 update based on recent genome-wide linkage analyses. *Hum Genet* 123: 119–131.
21. Li MD, Burmeister M (2009) New insights into the genetics of addiction. *Nat Rev Genet* 10: 225–231.
22. Mulligan MK, Ponomarev I, Hitzemann RJ, Belknap JK, Tabakoff B, et al. (2006) Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proc Natl Acad Sci U S A* 103: 6368–6373.
23. Bierut IJ, Strickland JR, Thompson JR, Afful SE, Cottler LB (2008) Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend* 95: 14–22.
24. Luo Z, Alvarado GF, Hatsukami DK, Johnson EO, Bierut IJ, et al. (2008) Race differences in nicotine dependence in the Collaborative Genetic study of Nicotine Dependence (COGEND). *Nicotine Tob Res* 10: 1223–1230.
25. Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li TK, et al. (1995) The Collaborative Study on the Genetics of Alcoholism. *Alcohol Health Res World* 19: 228–236.
26. Redden DT, Divers J, Vaughan LK, Tiwari HK, Beasley TM, et al. (2006) Regional admixture mapping and structured association testing: conceptual unification and an extensible general linear model. *PLoS Genet* 2: e137.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.
28. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
29. Curtin K, Wong J, Allen-Brady K, Camp NJ (2007) PedGenic: meta genetic association testing in mixed family and case-control designs. *BMC Bioinformatics* 8: 448.
30. Dudbridge F, Gusnanto A (2008) Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 32: 227–234.
31. Imoto I, Sonoda I, Yuki Y, Inazawa J (2001) Identification and characterization of human PKNOX2, a novel homeobox-containing gene. *Biochem Biophys Res Commun* 287: 270–276.
32. Karch DL, Barker L, Strine TW (2006) Race/ethnicity, substance abuse, and mental illness among suicide victims in 13 US states: 2004 data from the National Violent Death Reporting System. *Inj Prev* 12 Suppl 2: ii22–ii27.
33. Breslau N, Johnson EO, Hiripi E, Kessler R (2001) Nicotine dependence in the United States: prevalence, trends, and smoking persistence. *Arch Gen Psychiatry* 58: 810–816.
34. Kandel D, Chen K, Warner LA, Kessler RC, Grant B (1997) Prevalence and demographic correlates of symptoms of last year dependence on alcohol, nicotine, marijuana and cocaine in the U.S. population. *Drug Alcohol Depend* 44: 11–29.
35. Hartel DM, Schoenbaum EE, Lo Y, Klein RS (2006) Gender differences in illicit substance use among middle-aged drug users with or at risk for HIV infection. *Clin Infect Dis* 43: 525–531.