Retraction and Correction

RETRACTION

GENETICS, SOCIAL SCIENCES
Retraction for “PKNOX2 gene is significantly associated with substance dependence in European-origin women,” by Xiang Chen, Kelly Cho, Burton H. Singer, and Heping Zhang, which published online August 31, 2009, in Proc Natl Acad Sci USA (10.1073/pnas.0908521106).

The authors wish to retract this paper because its publication violates the Gene Environment Association Studies Genes and Environment Initiative Study of Addiction: Genetics and Environment (SAGE) dataset’s embargo policy. The SAGE data access agreement states that investigators agree not to submit findings of the SAGE dataset(s) for publication until September 23, 2009. The authors sincerely apologize for this violation of SAGE policy.

Xiang Chen
Kelly Cho
Burton H. Singer
Heping Zhang

CORRECTION

NEUROSCIENCE

The authors note that on page 13583, in the legend for Fig. 2, an equation appeared incorrectly. The figure and its corrected legend appear below. Additionally, in Fig. 3A on page 13584, the panels labeled “Ghrelin” and “PER1” appeared incorrectly. The corrected figure and its legend appear below. These errors do not affect the conclusions of the article.

Fig. 2. Running wheel behavior of wild-type and GHSR−/− mice during ad libitum feeding, food restriction, and food deprivation conditions. (A) The bar above the actograms shows the light–dark cycle; time of food availability is shown in gray. Actograms depict activity of representative GHSR+/+ and GHSR−/− mice during ad libitum feeding (days 1–4), food restriction ZT6–ZT14 (days 4–15), ad libitum food availability (days 15–18), and food deprivation (day 19). (B) Group activity profiles show the amount of wheel running during the last 7 days of restricted feeding in GHSR+/+ (black) and GHSR−/− (gray) mice. The data are plotted in 10-min bins (mean ± SEM). **, *P < 0.002, difference between GHSR+/+ and GHSR−/− in onset time of activity. (C) Line graph of cumulative wheel-running activity (mean ± SEM) from lights on (ZT0) to time of food presentation (ZT6) shows that GHSR−/− mice (solid gray line) ran 42.4% less than GHSR+/+ (solid black line) mice. Superimposed are the curves derived from the Gaussian function f(x) = e^-(x^2)/2σ^2 (dashed lines). (D) The anticipation ratios during 7 days of restricted feeding (Top) and the persistence ratio (Middle) on the day of food deprivation show that GHSR−/− mice (gray bars) and control GHSR+/+ (black bars) mice, with significant differences between groups. (Bottom) Daily activity during the period of food restriction. *, *P < 0.002; **, *P < 0.01.
PKNOX2 gene is significantly associated with substance dependence in European-origin women

Xiang Chen1, Kelly Cho1, Burton H. Singer2, and Heping Zhang2

1Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520-8034; and 2Office of Population Research, Princeton University, Princeton, NJ 08544

Contributed by Burton H. Singer, August 1, 2009 (sent for review March 18, 2009)

Substance dependence is a complex environmental and genetic disorder that results in serious health and socioeconomic consequences. Many studies have reported and implicated genes associated with various substance dependence outcomes, including addiction to nicotine and addiction to alcohol. Using data from several genome-wide case-control studies, we conducted a genome-wide association study of a composite substance dependence phenotype derived from six individual diagnoses: addiction to nicotine, alcohol, marijuana, cocaine, opiates, or other drugs as a whole. We identified a strong (odds ratio = 1.77) and significant (P value = 7e-8) association signal with the PBX/knotted 1 homeobox 2 (PKNOX2) gene on chromosome 11 in European-origin women with the composite phenotype. Our findings also indicate that the associations are not as significant when individual outcomes for addiction are considered, underscoring the importance of considering multiple addiction types.

Results

Table 1 summarizes the top eight significant SNPs which cluster in PKNOX2 on chromosome 11 (11q24). None of the eight SNPs violates the Hardy-Weinberg equilibrium assumption (minimum P level = 0.12). The most significant SNP (rs12284594; P = 7.13E-08) is observed in white women with an odds ratio (OR) of 1.77; thus, 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 (23). In addition, there are seven other SNPs with P values less than 3.8E-06 in white women, and the corresponding ORs are similar in magnitude (1.63–1.72). We also examined haplotypes 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 (24), or whether the correlation among related individuals was ignored. Although we also observed that these eight 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.

Additional analyses were performed to examine each substance dependence outcome separately for the top eight SNPs presented above. The corresponding P values for the eight SNPs for each substance dependence outcome are presented in Table 2. Alcohol dependence shows the strongest association (P = 3.8E-06).

The authors declare no conflict of interest.

1X.C. and K.C. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: heping.zhang@yale.edu or singer@princeton.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0908521106

See Retraction Published
September 9, 2009
1.97E-06 with rs12284594); however, none of these P values attains the genomewide significance level of 0.05.

**Discussion**

We have found a 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. PKNOX2 was identified as a TALE homeodomain-encoding gene, located at 11q24 in humans by Imoto et al. (25). They reported that the structure and subcellular localization of PKNOX2 indicate that this protein functions as a nuclear transcription factor. Several years later, PKNOX2 was identified as one of the cis-regulated genes for alcohol addiction in mice (22). However, 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 (Fig. 1). Among all subjects, 47% have been diagnosed with alcohol dependence in combination with other substance dependence outcomes. Our results, which show a strong association of this composite of binary 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 candidate gene for substance dependence in humans, particularly for white women in the SAGE sample.

In addition, among our most significant SNPs, we do not observe those genes previously reported for alcoholism or nicotine. Rather we find a set of genes among the top SNPs. When each substance dependence outcome was individually investigated, we found no association that reached genomewide significance. This suggests that substance dependence or addiction as a whole has different risk genes compared to any single addiction outcome.

**Table 1. Summary of the 8 most significant SNPs in PKNOX2 gene showing genomewide significant association with substance dependence in White women**

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

The high-risk allele is in the parenthesis.

It may also mean that there is more power in detecting common genes acting upon co-morbid addiction outcomes as a whole.

Different ethnic groups have vastly different underlying genetics for many complex diseases, 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 (26–28). More recently, Luo et al. (29) have reported that genetic differences between black smokers and white smokers influence the nature of their nicotine dependence; they found that black smokers become dependent at a lower threshold (number cigarettes per day) than white smokers. 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 (30). 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

**Table 2. Associations of the 8 most significant SNPs in PKNOX2 with six individual substance dependence outcomes (p-values)**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nicotine</th>
<th>Alcohol</th>
<th>Marijuana</th>
<th>Cocaine</th>
<th>Opiates</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1426153 (G)</td>
<td>0.0159</td>
<td>5.75E-5</td>
<td>7E-4</td>
<td>3E-4</td>
<td>0.0113</td>
<td>1E-4</td>
</tr>
<tr>
<td>rs11220015 (A)</td>
<td>0.0163</td>
<td>6.86E-5</td>
<td>0.0010</td>
<td>3E-4</td>
<td>0.0037</td>
<td>4.18E-5</td>
</tr>
<tr>
<td>rs11602925 (G)</td>
<td>0.0136</td>
<td>4.24E-5</td>
<td>7E-4</td>
<td>3E-4</td>
<td>0.0059</td>
<td>5.29E-5</td>
</tr>
<tr>
<td>rs750338 (C)</td>
<td>0.0491</td>
<td>4.26E-5</td>
<td>0.0013</td>
<td>2E-4</td>
<td>0.0112</td>
<td>2.22E-5</td>
</tr>
<tr>
<td>rs12273605 (T)</td>
<td>0.0921</td>
<td>3E-4</td>
<td>3.53E-5</td>
<td>1E-4</td>
<td>0.0680</td>
<td>3.1E-5</td>
</tr>
<tr>
<td>rs10893365 (C)</td>
<td>0.0411</td>
<td>1.72E-5</td>
<td>8.58E-6</td>
<td>2.91E-5</td>
<td>0.0699</td>
<td>2.58E-5</td>
</tr>
<tr>
<td>rs10893366 (T)</td>
<td>0.0621</td>
<td>1.37E-5</td>
<td>8.80E-6</td>
<td>8.63E-5</td>
<td>0.0905</td>
<td>5.35E-5</td>
</tr>
<tr>
<td>rs12284594 (G)</td>
<td>0.0239</td>
<td>1.97E-6</td>
<td>8.54E-6</td>
<td>4.39E-5</td>
<td>0.0533</td>
<td>2.45E-5</td>
</tr>
</tbody>
</table>
substance dependence outcome, men may be socially more prone to environmental influences promoting substance use, and thus more vulnerable to addiction, compared to women (31). 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 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.

Methods

Study of Addiction: Genetics and Environment (SAGE) Data. We obtained the genome-wide 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), which originally included 4,121 subjects for whom the addiction to the six categories of substances and genome-wide SNP data (ILLUMINA Human 1M platform) were available. SAGE is a case-controlled study of mostly unrelated individuals aimed at identifying genetic factors of genotype and phenotype (dbGaP). The data were from the Study of Genotype and Phenotype (dbGaP). The data were from the Study of Genotype and Phenotype (dbGaP). The data were from the Study of Genotype and Phenotype (dbGaP).

Table 3. Descriptive statistics of the sample stratified by sex and race

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

Statistical Analysis. Because of the presence of different ethnic populations in our data, it was crucial for us to investigate and control for potential confounding by stratification and admixture due to disequilibrium between pairs of unlinked loci (30). 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) (34). 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. Overall a total of 1,513 subjects were included in this analysis as having two or more substance addiction according to DSM-IV. Of these, there were 316, 585, 237, and 357 subjects in the black male, white male, black female, and white female subsets, respectively. The distribution of subjects diagnosed with each substance dependence is presented in Table 3. The top three most widely used substances among the six were alcohol, nicotine and cocaine, in that order (Fig. 1). We used allelic χ² tests with one degree of freedom in our analysis, stratified by race and sex. Haploview (version 4.0) (35) 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 (24) to perform association analyses. PedGenie first performs the allelic χ² 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 (23) studied the genomewide significance threshold for the U.K. 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 genomewide significance threshold for this population is 7.2E-8. We used this genomewide significance threshold (7.2E-8) for the Caucasian population (white men and white women) in our analysis.

ACKNOWLEDGMENTS. This work was supported in part by National Institutes of Health Grants K02DA017713, R01DA16750, 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 HG004466. 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 U10 HG004438, 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 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.p through dbGaP accession number phs000092.v1.p.