Research Article

NCK2 Is Significantly Associated with Opiates Addiction in African-Origin Men

Zhifa Liu, Xiaobo Guo, Yuan Jiang, and Heping Zhang

1 Department of Biostatistics, Yale University School of Public Health, New Haven, CT 06520, USA
2 Department of Statistical Science, School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou 510275, China
3 Department of Statistics, Oregon State University, Corvallis, OR 97331, USA

Correspondence should be addressed to Heping Zhang; heping.zhang@yale.edu

Received 31 December 2012; Accepted 18 January 2013

Substance dependence is a complex environmental and genetic disorder with significant social and medical concerns. Understanding the etiology of substance dependence is imperative to the development of effective treatment and prevention strategies. To this end, substantial effort has been made to identify genes underlying substance dependence, and in recent years, genome-wide association studies (GWASs) have led to discoveries of numerous genetic variants for complex diseases including substance dependence. Most of the GWAS discoveries were only based on single nucleotide polymorphisms (SNPs) and a single dichotomized outcome. By employing both SNP- and gene-based method of analysis, we identified a strong (odds ratio = 13.87) and significant (P-value = 1.33E-11) association of an SNP in the NCK2 gene on chromosome 2 with opiates addiction in African-origin men. Codependence analysis also identified a genome-wide significant association between NCK2 and comorbidity of substance dependence (P-value = 3.65E-08) in African-origin men. Furthermore, we observed that the association between the NCK2 gene (P-value = 3.12E-10) and opiates addiction reached the gene-based genome-wide significant level. In summary, our findings provided the first evidence for the involvement of NCK2 in the susceptibility to opiates addiction and further revealed the racial and gender specificities of its impact.

1. Introduction

Substance dependence is believed to result from a combination of genetic and environmental factors. Since substance dependence is a chronic brain disease, with high relapse rates, it causes serious social, economic, and medical consequences [1–3]. The World Health Organization (WHO) and the United Nations Office on Drugs and Crime (UNODC) reported that opiate dependence is associated with a high risk of HIV infection when opiates are injected using contaminated injection equipment [4]. Paulozzi et al. in 2006 reported that the number of deaths which involved prescription opioid analgesics increased from 2,900 in 1999 to at least 7,500 in 2004, an increase of 160% in just 5 years [5]. All available evidence indicated that the increasing numbers of deaths are significantly correlated to the increasing use of prescription drugs, especially opioid painkillers, among people during the working years of life. While exposure to drugs is the prerequisite for addiction, the most important question is as follows: who will be addicted after the exposure? Genes are believed to be a major factor, although it is most likely that there are multiple genes as well as gene-environment interactions. For this reason, understanding the genetic mechanisms behind vulnerability to drug addiction is critical to improve the quality of overall health and life. Linkage and genome-wide association studies (GWASs) have implicated many regions and genes for dependence on alcohol, tobacco, and opiates. GABRA2, CHRMM2, ADH4, PKNOX2, GABRG3, TAS2R16, SNCA, OPRK1, and PDYN have all been associated with alcohol dependence with various degrees of replication [6–21]. Associations of other candidate alcohol dependence genes, such as KIAA0040, ALDH1A1, and MANBA [18, 20, 22–25], remain to be confirmed. Several groups reported CHRNA5, CHRNA3,
CHRNB4, and CSMD1 to be associated with nicotine dependence [26–34]. Meanwhile, recent studies also reported that a group of genes, such as OPRM1 [35–37], OPRD1, OPRK1 [21, 38, 39], HTTRIB [40], SLC6A4 [41], GABRG2 [42], and BDNF [43], to be associated or in linkage with opiates addiction.

Complex diseases may involve heterogeneous genetic effects in different ethnic and gender groups [7, 44–47]. Luo et al. [44] reported that African-origin smokers become dependent at a lower threshold (number of cigarettes per day) than European-origin smokers. Hartel et al. [46] found that men are more vulnerable to addiction when compared to women. In addition, Chen et al. [7] revealed that PKNOX2 is associated with drug addiction in European-origin women. These examples underscore the necessity to consider demographic or even other covariates in genetic association studies.

Many of the reported genetic variants have been identified through single SNP association tests. Despite many of the successes, a single SNP tends to have a small effect, and the single SNP-based association tests require a very stringent significance level, which is likely a key factor to the so-called “missing heritability” problem [48, 49]. To overcome some of these limitations, gene-based analysis [50–52] has emerged to jointly analyze the SNPs within genes. Gene-based methods are less affected by the heterogeneity of a single locus; hence the results may be more robust across populations [53], which increases the likelihood of replication. Hence, we performed both single SNP-based and gene-based association analyses for the data from the Study of Addiction: Genetics and Environment (SAGE) [6] which includes well-characterized phenotypic data on substance dependence including addiction to nicotine, alcohol, marijuana, cocaine, opiates, and other drugs. In our analysis, we find a genome-wide significant association of NCK2 gene on chromosome 2 with opiates dependence in African-origin men at both the SNP and gene levels. NCK2 is a member of NCK family of adaptor proteins, which is associated with tyrosine-phosphorylated growth factor receptors of their cellular substrates [54]. However, to the best of our knowledge, NCK2 has not been reported to be associated with any drug addiction outcomes in humans.

2. Materials and Methods

Phenotypes for multisubstance dependency and genome-wide SNP data from SAGE [6] were downloaded from dbGaP (http://www.ncbi.nlm.nih.gov/gap). SAGE is a large case-control association study which investigates the genetic variants for drug addiction. The samples were collected from three large-scale genome-wide association studies: Collaborative Study on the Genetic of Alcoholism (COGA), the Family Study of Cocaine Dependence (FSCD), and the Collaborative Genetic Study of Nicotine Dependence (COGEND) [16, 44, 55, 56]. The original data set contains 4,121 subjects with six categories of substance dependence data: addiction to alcohol, cocaine, marijuana, nicotine, opiates, and other drugs. Lifetime dependence on these six substances is diagnosed by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). The genotyping was performed by the illumina Human 1M platform. In this study, we followed a quality control/quality assurance process similar to previous analyses [7, 57]. Individuals with call rates <90% and SNPs with minor allele frequency MAF <1% were excluded from the analysis. The P value for the Hardy-Weinberg equilibrium was set up by >0.0001. These steps reduced the level of noise in genotypes and increased the efficiency of analysis. There are 60 duplicate genotype samples and 9 individuals with ethnic backgrounds other than African origin or European origin. All of those individuals were removed from the subject list. Finally, there were a total of 3,627 unrelated samples with 859,185 autosomal SNPs for our final analysis. To alleviate the confounding by population substructure, we stratified the sample by race and sex. Finally, there are four sub-samples: 1,393 European-origin women, 1,131 European-origin men, 568 African-origin women and 535 African-origin men. The distribution 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.

3. Methods

Figure 1 displays the flow chart of our analytic strategy, and the details of the association analysis methods are described later.

3.1. Statistical Analysis for Single Trait. The SNP-based association is performed by the standard allelic test and logistic regression to obtain the P values for individual SNPs, and PLINK software (version 1.07) was used for analysis [58]. Meanwhile, a list of SNP pairs in linkage disequilibrium (LD) \((r^2 > 0.2)\) is calculated for the gene-based association test.

For the gene-based analysis, we used the open-source tool: Knowledge-Based Mining System for Genome-Wide Genetic Studies (KGG, version 2.0)—based on the SNP association test results and LD files produced by PLINK. The procedure was performed as the following. We first calculate the effective number \(m_e\) of independent \(P\) value among \(m\) SNPs within a gene. Then, we sort the SNPs and calculate the effective number \(m_e(j)\) of independent \(P\)-values among the
top \( j \) significant SNPs. Finally, the modified Simes test [51] was employed to obtain a gene-based \( P \) value as follows,

\[
P_G = \min \left( \frac{m_{(j)}}{P_{(j)}} \right),
\]

where \( P_{(j)} \) is the \( j \)th most significant among the \( m \) SNPs within a gene. We refer the interested readers to [51] for details.

In the gene-based method, SNPs within 20 kilo bases (kb) 5’ upstream and 10 kilo bases (kb) 3’ downstream of a gene’s coding regions [59] were assigned to the gene. In addition, we included other SNPs if they are in strong LD \( (r^2 > 0.9) \) with the initially mapped SNPs within the gene [60].

Since there are about 20,000 protein coding genes in human genome, we used 0.05/20,000 = 2.5 \( \times \) 10\(^{-6}\) as the genome-wide significance threshold for the gene-based association test. In contrast, we used 5.0 \( \times \) 10\(^{-8}\) as the genome-wide significance threshold for the SNP-based association test [61].

3.2. Codependence Association Analysis. Although logistic regression is commonly used to study a binary outcome, it is not suitable to evaluate comorbidity involving multiple outcomes. We use a nonparametric association test based on Kendall’s tau [62] to study the comorbidity. The Kendall's tau-based association test proceeds as follows.

Suppose that we observe a \( p \)-dimensional vector of traits \( Y_i = (Y_i^{(1)}, \ldots, Y_i^{(p)})^T \), genotype \( G_i \), and a \( q \)-dimensional vector of covariates \( Z_i = (Z_i^{(1)}, \ldots, Z_i^{(q)})^T \) for the \( i \)th subject in a population-based study with \( n \) subjects, and \( (Y_i, G_i, Z_i) \) : \( i = 1, \ldots, n \) are independent samples. For subjects \( i \) and \( j \), let \( Y_i \) and \( Y_j \) be their vectors of traits, respectively, and analogously, \( G_i \) and \( G_j \), and \( Z_i \) and \( Z_j \) are their genotypes and covariates. Generalized from Kendall’s tau, a \( U \) statistic is defined to measure the association between \( Y \) and \( G \) as follows:

\[
U = \left( \frac{n}{2} \right)^{-1} \sum_{i<j} (Y_i - Y_j) (G_i - G_j).
\]

Without considering the covariates and conditioning on all phenotypes, \( U \) follows an asymptotically normal distribution in the absence of association [63]. To accommodate covariates, a weighted \( U \) statistic has been developed [64, 65]. We refer to Jiang and Zhang [64] for a detailed description of the method. For the purpose of comparison, we present the results with and without considering age as the covariate. Recall that our analysis is stratified by ethnicity and gender.

4. Results

4.1. Association Analysis at SNP Level. Table 2 summarizes the top four significant SNPs (with \( P < 1.0 \times 10^{-6} \)) in gene NCK2 on chromosome 2 (2q12) for opiates dependence in African-origin men. We identified a genome-wide significant SNP (rs2377339 with \( P = 1.33 \times 10^{-11} \)) for the opiates dependence in African-origin men by the allelic test. Logistic regression also yielded strong evidence for the association between the SNP rs2377339 (\( P = 1.01 \times 10^{-7} \)) and opiates dependence although the \( P \)-value did not reach the genome-wide significance threshold. In addition, Table 2 presents the association results for the other five addictions with the four candidate SNPs. None of the four SNPs appeared significantly associated with the other five substance addictions.

4.2. Association Analysis at Gene Level. The gene-based association results are displayed in the last two rows of Table 2. Specifically, we included 39 SNPs in NCK2. The \( P \) values from the gene-NCK2-based tests that were obtained through the standard allelic test and logistic regression are 3.12 \( \times \) 10\(^{-10}\) and 2.70 \( \times \) 10\(^{-8}\), respectively. The gene-based \( P \) value from the standard allelic test reached the genome-wide significance level at gene level. The gene-based \( P \) value through logistic regression is very close to the gene-based genome-wide significance level. Therefore, both methods provided significant evidence that supports the association between the NCK2 gene and opiates dependence in African-origin men. For the addiction of the other five substances in African-origin men, nicotine dependence had the most significant association with the NCK2 gene \( (P = 9.56 \times 10^{-3}) \).

4.3. Haplotypes Analysis. We also examined association of haplotypes with opiate addiction in NCK2 region. Figure 2 displays the linkage disequilibrium (LD) heat map of 14 SNPs
Table 2: Association of the most significant SNPs in the NCK2 gene with the six substances dependence in African-origin men.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Method</th>
<th>Alcohol</th>
<th>Cocaine</th>
<th>Marijuana</th>
<th>Nicotine</th>
<th>Opiates</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCK2</td>
<td>rs2377339</td>
<td>Logistic regression</td>
<td>2.46E-2</td>
<td>5.09E-2</td>
<td>4.48E-2</td>
<td>7.01E-2</td>
<td>1.10E-7</td>
<td>3.84E-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard allelic test</td>
<td>6.03E-3</td>
<td>4.34E-2</td>
<td>3.89E-2</td>
<td>6.25E-2</td>
<td>1.33E-11</td>
<td>1.78E-3</td>
</tr>
<tr>
<td>NCK2</td>
<td>rs7589342</td>
<td>Logistic regression</td>
<td>2.84E-1</td>
<td>8.79E-1</td>
<td>9.35E-1</td>
<td>8.60E-1</td>
<td>1.45E-4</td>
<td>4.26E-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard allelic test</td>
<td>2.81E-1</td>
<td>8.78E-1</td>
<td>9.34E-1</td>
<td>8.59E-1</td>
<td>5.39E-5</td>
<td>4.22E-1</td>
</tr>
<tr>
<td>NCK2</td>
<td>rs1299533</td>
<td>Logistic regression</td>
<td>1.71E-1</td>
<td>7.51E-1</td>
<td>7.16E-1</td>
<td>9.86E-1</td>
<td>1.89E-4</td>
<td>4.69E-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard allelic test</td>
<td>1.68E-1</td>
<td>7.51E-1</td>
<td>7.15E-1</td>
<td>9.86E-1</td>
<td>7.82E-5</td>
<td>4.67E-1</td>
</tr>
<tr>
<td>NCK2</td>
<td>rs12053259</td>
<td>Logistic regression</td>
<td>1.39E-1</td>
<td>9.43E-1</td>
<td>7.42E-1</td>
<td>9.01E-1</td>
<td>2.31E-4</td>
<td>4.86E-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard allelic test</td>
<td>1.33E-1</td>
<td>9.42E-1</td>
<td>7.39E-1</td>
<td>9.9E-1</td>
<td>8.67E-5</td>
<td>4.80E-1</td>
</tr>
</tbody>
</table>

NCK2 — KGG-logistic 1.83E-1 8.15E-1 5.92E-1 9.56E-3
NCK2 — KGG-Standard allelic test 1.41E-1 8.16E-1 4.83E-1 8.71E-3

Figure 2: Linkage disequilibrium heatmap near SNP rs2377339 on chromosome 2.

Table 3: Allele counts of rs2377339 in cases (opiates dependence) and controls (nonopiates dependence) in African-origin men.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td></td>
</tr>
<tr>
<td>Genotype AA</td>
<td>35</td>
</tr>
<tr>
<td>Genotype AG</td>
<td>9</td>
</tr>
<tr>
<td>Genotype GG</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
</tr>
<tr>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Genotype AA</td>
<td>483</td>
</tr>
<tr>
<td>Genotype AG</td>
<td>8</td>
</tr>
<tr>
<td>Genotype GG</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>491</td>
</tr>
</tbody>
</table>

in 28 kb region [66]. Haplotype “AGTTCAGATCTCGT” with probability 0.016 yielded a P value of 1.66 × 10^-11. The genome-wide significant association between this haplotype and opiate addiction reduces the chance of a false discovery at the peak of a single SNP.

4.4. Contingency Table Analysis. We further examined the relationship between SNP rs2377339 and the opiates dependence in African-origin men. Table 3 depicts the allele frequencies of SNP rs2377339. The proportion of individuals having minor allele G is 21.43% in the case group and 1.63% in the control group. The odds ratio of SNP rs2377339 is 13.87, indicating that those who have the risk allele (G) for rs2377339 are at a significantly increased risk of being diagnosed with opiates dependence.

4.5. Stratification Analysis. Furthermore, in Table 4, we investigated the racial specificity and sex difference in the association between SNP rs2377339 and opiates dependence. This scrutiny required us to include all racial and gender groups. We observed that the MAF and P values vary between different races and genders. The association between rs2377339 and opiates dependence becomes less significant in the overall cohort, after we adjusted race and gender in logistic regression.

4.6. Codependence Association Analysis. In Table 5, we also presented the association results for NCK2 and comorbidity
Table 4: Association between SNP rs2377339 and opiates dependence by race and sex.

<table>
<thead>
<tr>
<th></th>
<th>MAF</th>
<th>P value</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-origin men</td>
<td>1.59%</td>
<td>1.33E−11</td>
<td>13.87</td>
</tr>
<tr>
<td>African-origin women</td>
<td>1.14%</td>
<td>1.64E−1</td>
<td>2.82</td>
</tr>
<tr>
<td>European-origin men</td>
<td>6.77%</td>
<td>8.13E−2</td>
<td>0.55</td>
</tr>
<tr>
<td>European-origin women</td>
<td>6.64%</td>
<td>6.95E−1</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Combined* 5.04% 4.37E−1 1.17

*Logistic regression is used to adjust for sex and race.

Table 5: Association of most significant SNPs in NCK2 with codependence of six individual substances dependence outcomes (P value).

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>position</th>
<th>P value for unadjusted</th>
<th>P value for adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCK2</td>
<td>rs2377339</td>
<td>105823723</td>
<td>3.65E−8</td>
<td>2.03E−8</td>
</tr>
<tr>
<td></td>
<td>rs6740723</td>
<td>105798288</td>
<td>5.43E−3</td>
<td>2.58E−2</td>
</tr>
<tr>
<td></td>
<td>rs7589342</td>
<td>105799910</td>
<td>3.13E−3</td>
<td>1.68E−2</td>
</tr>
<tr>
<td></td>
<td>rs12995333</td>
<td>105802798</td>
<td>3.81E−3</td>
<td>1.83E−2</td>
</tr>
</tbody>
</table>

of substance dependence. The most significant signal in NCK2 was observed for SNP rs2377339 in men of African-origin with \( P = 3.65 \times 10^{-8} \) in adjusted association test and \( P = 2.03 \times 10^{-8} \) in unadjusted association test. \( P \) values of SNPs in NCK2 for other ethnicity by gender groups were far from the genome-wide significance level and, hence, are omitted here.

5. Discussion

We found a genome-wide significant association between SNP rs2377339 and opiates dependence in African-origin men. The NCK2 gene that contains SNP rs2377339 also achieved the genome-wide significance for opiates dependence at the gene level. For the addiction of the other five substances, nicotine dependence had the most significant association but not significant at the genome-wide level. For the addiction of the other five substances, nicotine dependence had the most significant association but not significant at the genome-wide level. Such interactions have been suggested in other addiction research [44, 45, 47]. Again, our result further supports the importance to examine interactions among genes, race, and sex in addiction.

Conflict of Interests

The authors declare that they have no conflict of interests.

Authors’ Contribution

Zhifa Liu and Xiaobo Guo contributed equally.

Acknowledgments

This work was supported by Grant R01 DA016750-09 from the National Institute on Drug Abuse. Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative (GEI) (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (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 (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH Contract "High Throughput...
Genotyping for Studying the Genetic Contributions to Human Disease" (HHSN268200782096C). The datasets used for the analyses described in this paper 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 no. phs000092.v1.p.

References


