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

Xiang Chen1*, Kelly Cho1*, Burton H. Singer2, Heping Zhang1*

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 homebox 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.

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 CHRNB4) 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 – nicotine, alcohol, marijuana, cocaine, opiates and other drugs – are presented in Table 1. The top three most widely used substances among the six substances in each of the six categories – nicotine, alcohol, marijuana, cocaine, opiates or other drugs – are presented in Table 1. The 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).

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    |

PKNOX2 Gene Associated with Substance Dependence

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 43.9% 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 916, 505, 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).

Table 1. Descriptive statistics of the sample stratified by sex and race.
Determination of Significance Threshold for GWAS

Using genotypes from the Wellcome Trust Case-Control Consortium, Dudbridge and Gusnanto [30] studied the genomewide 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 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.

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.

| SNP P-value OR |
|----------------|
| rs1426153 (G) 1.84E-06 1.66 |
| rs11220015(A) 1.97E-06 1.65 |
| rs16002925(G) 1.24E-06 1.67 |
| rs750338(C) 4.22E-07 1.63 |
| rs12273605(T) 3.83E-06 1.71 |
| rs10893365(C) 2.77E-07 1.72 |
| rs10893366(T) 6.87E-07 1.70 |
| rs12284594(G) 7.13E-08 1.77 |

The high-risk allele is in the parenthesis.
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 genomewise 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 other ethnicities. 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 www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id = phs000092.v1.p through dbGaP accession number phs000092.v1.p.

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 BHS HZ. Wrote the paper: XC KC BHS HZ.

References

1. 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 fatal Wkly Rep 58: 29–33.
2. Mokhaji NJ, Sher L (2007) Preventing suicide in adolescents with alcohol use disorders. J Adolesc Med 19: 53–59.
3. Uhl GR, Emler 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.
4. True WR, Heath AC, Scherrer JF, Xian H, Lin N, et al. (1999) Interrelationships of genetic and environmental influences on conduct disorder and alcohol and marijuana dependence symptoms. Am J Med Genet 88: 391–397.
5. Merikangas KR, Stolar M, Stevens DE, Gouler J, Prein MA, et al. (1998) Familial transmission of substance use disorders. Arch Gen Psychiatry 55: 973–979.
6. 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.

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 | 6.66E-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 | 5.33E-5 |
| rs112284594(G) | 0.0239 | 1.97E-6 | 8.54E-6 | 4.39E-5 | 0.0533 | 2.45E-5 |

doi:10.1371/journal.pone.0016002.t003
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. (1996) Genomewide search for genes affecting the risk for alcohol dependence. Am J Med Genet 81: 207–213.
9. Song J, Koller DL, Foroud T, Carr K, Zhao J, et al. (2003) Association of GABRA5 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 GABAA 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, Inazawa J, Kinoshita T, Takagi S, et al. (1994) Aldehyde dehydrogenase genotypes in Japanese alcoholics. Lancet 343: 741–742.
14. Maesawa 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 alcohols with the atypical aldehyde dehydrogenase 2pd. I. A comparison of the genotypes of ALDH2, ADH2, ADH3, and cytochrome P4502E1 between alcoholics and nonalcoholics. Alcohol Clin Exp Res 20: 52–55.
16. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, et al. (1995) The alcohol and drug abuse gene collection: results from the Collaborative Study on the Genetics of Alcoholism. Alcohol Health Res World 19: 229–236.
17. 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.
18. Bierut LJ, Strickland JR, Thompson JR, Alaf et al. (2008) Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. Drug Alcohol Depend 95: 14–22.
19. Luo Z, Alvarado GF, Hatsukami DK, Johnson EO, Bierut LJ, et al. (2008) Race differences in nicotine dependence in the Collaborative Genetic study of Nicotine Dependence (COGEND). Nicotine Tob Res 10: 1223–1230.
20. 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: 229–236.
21. Curtin K, Wong J, Allen-Brady K, Camp NJ (2007) PedGenie: meta genetic association testing in mixed family and case-control designs. BMC Bioinformatics 8: 486.
22. Curtin K, Wong J, Allen-Brady K, Camp NJ (2007) PedGenie: meta genetic association testing in mixed family and case-control designs. BMC Bioinformatics 8: 486.
23. Yu et al. (2006) Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. Proc Natl Acad Sci U S A 103: 6368–6373.
24. Luo Z, Alvarado GF, Hatsukami DK, Johnson EO, Bierut LJ, 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: 229–236.
26. Redden DT, Divis J, Vaughan LK, Twari 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) PedGenie: meta genetic association testing in mixed family and case-control designs. BMC Bioinformatics 8: 486.
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 (2004) Identification and characterization of human PNKX2, a novel homeobox-containing gene. Biochem Biophys Res Commun 327: 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.