The *J*cientificWorldJOURNAL

Research Article

Large Scale Association Analysis for Drug Addiction: Results from SNP to Gene

Xiaobo Guo,[1, 2] Zhifa Liu,[1] Xueqin Wang,[2, 3] and Heping Zhang[1]

[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, Guangdong, China
[3] *Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong, China*

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

Many genetic association studies used single nucleotide polymorphisms (SNPs) data to identify genetic variants for complex diseases. Although SNP-based associations are most common in genome-wide association studies (GWAS), gene-based association analysis has received increasing attention in understanding genetic etiologies for complex diseases. While both methods have been used to analyze the same data, few genome-wide association studies compare the results or observe the connection between them. We performed a comprehensive analysis of the data from the Study of Addiction: Genetics and Environment (SAGE) and compared the results from the SNP-based and gene-based analyses. Our results suggest that the gene-based method complements the individual SNP-based analysis, and conceptually they are closely related. In terms of gene findings, our results validate many genes that were either reported from the analysis of the same dataset or based on animal studies for substance dependence.

1. Introduction

Genome-wide association studies (GWAS) have become a powerful tool in the identification of susceptible loci for numerous diseases [1]. A typical strategy in GWAS is to analyze single nucleotide polymorphisms (SNPs) individually and select the top SNPs by setting a stringent threshold for the *P* value. Then the top SNPs were mapped into functional regions such as a gene or pathway to facilitate further investigation of the corresponding gene and disease. Based on SNP-based association analysis, many genetic variants underlying complex diseases or traits were detected [2, 3]. Due to the large number of SNPs with each of which entails an association test, it is essential to control the type I error or false discovery rate [4]. A predefined *P* value $< 5 \times 10^{-8}$ is usually used as the threshold to declare a genome-wide significance SNP, which also limits the discoveries of the genes that are important to the disease. Also importantly, susceptible SNPs generally explain a small fraction of the risk—a phenomenon commonly referred to as the "missing heritability" [5, 6]. To alleviate this

problem, alternative methods have emerged to complement the simple SNP-based methods. Among those methods, gene-based analysis [7–9], which jointly analyzes the SNPs within genes, is a promising solution to improve the power of GWAS. Compared with the SNP-based approach, gene-based association analysis has certain advantages. First, gene is a unit of heredity and function, and hence the gene-based association approaches can provide direct insights into the heredity and functional mechanisms of complex traits [10]. Second, from the statistical perspective, the gene-based association approaches reduce the number of association tests in the order of millions to about 20,000 gene-based tests, which dramatically reduces the chance of false discovery. In addition, the gene-based methods are not affected by the heterogeneity of a single locus. Hence, the results are highly consistent across populations [11], which enhances the likelihood of replication.

Gene-based methods have been successfully applied to GWAS of complex diseases, including Crohn's disease [7], type 1 diabetes [12], and melanoma [8]. Despite the above-noted features of the gene-based association approach, there

are few comparisons of genetic association analyses between SNP and gene-based methods. Here, we compare and relate these two approaches using the data from the Study of Addition: Genetics and Environment (SAGE) [13].

Recent studies show that there are many candidate genes associated with substance dependence. For example, GABRA2, CHRM2, ADH4, PKNOX2, GABRG3, TAS2R16, SNCA, OPRK1, and PDYN are well studied for alcohol addiction and have been replicated in many samples [13–28]. However, other candidate genes, such as KIAA0040, ALDH1A1, DKK2, and MANBA [25, 27, 29, 30], remain illusive. For addiction to nicotine, CHRNA5, CHRNA3, CHRNB4, and CSMD1 have been replicated in many studies [31–39].

Based on the analysis of the SAGE data, we report a number of susceptible loci at the SNP and/or gene levels, which validate many susceptibility loci that have been reported to be associated with substance dependence [13, 14, 25, 27, 29, 37, 38, 40–44]. Meanwhile, both SNP- and gene-based analyses reveal three novel risk genes: NCK2, DSG3, and PUSL1.

2. Materials and Methods

2.1. Dataset and Study Design. The dataset included 4,121 subjects in SAGE with six categories of substance dependence data: alcohol, cocaine, marijuana, nicotine, opiates, and other dependencies on drugs. The data were downloaded from dbGaP (study accession phs000092.v1.p1) [13]. SAGE [13] is a large case-control study which aims to detect susceptible genetic variants for addition. The subjects were recruited from eight study sites in seven states and the District of Columbia in the United States. All subjects' life time dependencies on these six dependencies are diagnosed by using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). All samples were genotyped on ILLUMINA Human 1 M platform at the Center for Inherited Disease Research in Johns Hopkins University. In this paper, we strictly followed the quality control/quality assurance as we did in our previous analysis [14]. Genome-wide SNP data were filtered by setting thresholds: MAF > 5% and call rate > 90%. In addition, 60 duplicate genotype samples and 9 individuals with ethnic backgrounds other than African origin or European origin were excluded in our analysis. Finally, 3,627 unrelated samples with 859,185 autosomal SNPs passed the quality control procedures. To avoid population stratification, the dataset was stratified into four subsamples: 1,393 white women, 1,131 white men, 568 black women, and 535 black men. To capture most of the gene coding and regulatory variants, SNPs are considered being mapped to a gene if their physical locations are within 20 kilobases (kb) 5′ upstream and 10 kilobases (kb) 3′ downstream of gene coding regions [26]. In addition, SNPs are also assigned to a gene if they are in strong LD ($r^2 > 0.9$) with the initially assigned SNPs within the gene [10]. Together, around 533,639 SNPs were assigned to 18,699 protein coding genes (28.6 ± 47.7 (mean ± SD) SNPs per gene).

Following the conventional standards, we used $5.0E - 8$ and $2.5E - 6$ as the genome-wide significant thresholds for SNP-based and gene-based methods, respectively [4]. To increase the power of detecting potentially important SNPs that do not meet the stringent thresholds, we also considered relaxed thresholds. Specifically, SNPs with $P < 1.0E - 5$ and genes $P < 5.0E - 4$ were considered further. These P values are referred to as relaxed significance thresholds below. The selected SNPs were then mapped into the corresponding genes by the mapping rule proposed above.

2.2. Genetic Association Test at SNP and Gene Levels. We took several steps in testing the associations between genetic variants (SNP or gene) and substance dependenice. First, the P value of each SNP was evaluated by the logistic regression, and then the correlation coefficients (r^2) of all SNP pairs were calculated. The computation was performed in PLINK software (version 1.07) [45]. In the second step, we implemented the gene-based analysis in the open-source tool: Knowledge-Based Mining System for Genome-Wide Genetic Studies (KGG, version 2.0) [46] based on the association test results and LD files obtained from PLINK. Simes procedure (GATES) was employed in the gene-based association test [7]. Specifically, assume that m SNPs are assigned to a gene; an association test such as through the traditional logistic regression or linear regression is used to examine the association between the phenotype and each single SNP. This step yields m P values for m SNPs. GATES combines the available m P values within a gene by using a modified Simes test to give a gene-based P value. The summary P value is defined as

$$P_G = \text{Min}\left(\frac{m_e p_{(j)}}{m_{e(j)}}\right), \tag{1}$$

where $p_{(j)}$ is the jth smallest P value among the m SNPs; m_e is the effective number of independent P values among m SNPs within the gene, and $m_{e(j)}$ is the effective number of independent P values among the top j SNPs. The effective number of independent P values was derived by accounting for the LD structure among the specified SNPs; we refer to [7] on the calculation.

In order to compare the performance of the SNP-based and gene-based methods, in the SNP-based method, we selected those SNPs whose P values were less than $1.0E - 5$ and then mapped them into the corresponding genes. This allows us to compare the susceptible genes identified by both methods discussed above.

3. Results

3.1. Detecting Susceptibility Loci at the Relaxed Significance Level. Table 1 summarizes the susceptible genes identified by the SNP-based association test and gene-based association test at the relaxed significance level. In total, 207 genes passed the relaxed gene-based threshold, whereas only 64 genes with SNPs passed the relaxed SNP-based threshold.

Next, we performed a literature search on the genetic regions which contain the identified genes and filtered the

TABLE 1: Summary statistics for susceptibility loci identified by gene-based method and SNP-based method.

	Alcohol		Cocaine		Marijuana		Nicotine		Opiates		Other	
	G	S	G	S	G	S	G	S	G	S	G	S
Black men	4	3	4	1	6	2	5	2	8	2	9	5
Black women	4	3	8	5	9	3	7	3	3	1	6	3
White men	16	3	9	2	10	3	4	1	11	3	3	1
White women	20	5	12	2	10	2	11	1	4	5	24	3

G refers to gene-based method. S refers to SNP-based method.

TABLE 2: Summary of the candidate genes identified by the gene-based and SNP-based methods.

Chr	Gene	Source	P value (gene-based)[a]	Min P value (SNP-based)[b]	Detected SD[c]	Reported SD	Reference
1	KIAA0040	White women	$3.75E - 05$	$2.60E - 06$	Alcohol	Alcohol	[27, 44]
2	HAAO	White women	$4.40E - 04$	$3.02E - 05$	Cocaine	Alcohol	[41]
2	NCK2	Black men	$2.70E - 06$	$1.10E - 07$	Opiates	NA	NA
3	SH3BP5	White men	$1.20E - 04$	$4.24E - 06$	Cocaine	Alcohol	[13]
4	MANBA	White men	$4.63E - 04$	$3.47E - 05$	Alcohol	Alcohol	[29]
7	RELN	White men	$8.53E - 04$	$5.32E - 06$	Cocaine	Smoking	[37]
8	CSMD1	Black women	$1.23E - 02$	$8.50E - 06$	Nicotine	Smoking	[37, 38]
11	LRP5	White men	$4.01E - 05$	$1.58E - 06$	Opiates	Smoking	[42]
11	PKNOX2	White women	$1.84E - 04$	$2.20E - 06$	Alcohol	Alcohol	[13, 27, 41]
12	IFNG	White women	$1.16E - 04$	$1.57E - 05$	Opiates	Smoking	[37]
18	FAM38B	Black women	$9.24E - 04$	$5.61E - 06$	Cocaine	Smoking	[40]
18	PTPRM	Black women	$2.21E - 3$	$9.50E - 06$	Marijuana	Alcohol	[43]
22	MAPK1	Black women	$2.79E - 04$	$3.52E - 05$	Marijuana	Alcohol	[25]

[a]P value (gene-based): the P value obtained by the gene-based association test;
[b]min P value (SNP-based): the minimal P value of the SNPs within the corresponding gene;
[c]SD: substance dependence.

susceptible genetic regions which have been reported to associate with substance dependence for further investigation. In Table 2, we listed the filtered genes, their associated substance dependence type, the P values for the gene-based method, the minimal P value of SNPs within a gene, and their literature references and reported substance dependence.

In Figure 1, we plot the filtered genes obtained from the SNP-based and gene-based analyses by the position on the chromosomes against their log-transformed P values, $-\log_{10}(P)$. Each point for the SNP-based analysis in Figure 1 corresponds to the smallest SNP-based P value within the gene.

Overall, five genes, NCK2 (opiates dependence in black men), SH3BP5 (cocaine dependence in white men), LRP5 (opiates dependence in white men), KIAA0040 (alcohol dependence in white women), and PKNOX2 (alcohol dependence in white women), were identified by both the SNP-based and gene-based methods as meeting either of the relaxed significance levels for a specific dependence and within a gender-racial group. Four genes, MAPK1 (marijuana dependence in black women), MANBA (alcohol dependence in white men), HAAO (cocaine dependence in white women), and IFNG (opiates dependence in white women), met the threshold by the gene-based method only. We found that the significant signal of gene MAPK1 was mainly driven by SNPs: rs7290469 ($P = 3.25E-5$), rs9610271 ($P = 4.19E - 5$), rs9610417 ($P = 5.38E - 5$), and rs2876981
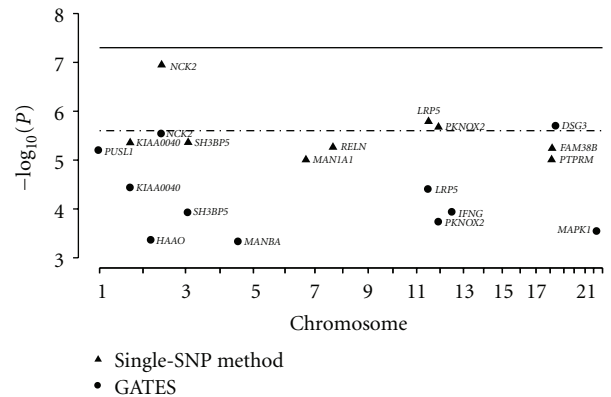

FIGURE 1: Comparison of candidate genes associated with substance dependence by the SNP- and gene-based analyses. A triangle represents the $-\log 10$ transformed P value of the marked gene from the gene-based analysis, and a dot represents the $-\log 10$ transformed the minimal P value of the SNPs within the marked gene. The solid and dashed ones are the genome-wide thresholds of SNP- and gene-based significance, respectively.

($P = 7.51E - 5$). The P values for these SNPs are slightly greater than the relaxed SNP-based threshold ($P < 1.0E - 5$), and hence the SNP-based method failed to detect them.

TABLE 3: Summary of genome-wide significant genes at the gene level (P value $< 1.0E - 5$) and their top SNPs with P value $< 1.0E - 3$.

Population	Substance dependence	Gene	Gene's P value	Top SNPs	SNP's P value
Black men	Opiates	NCK2	$2.70E - 06$	rs2377339	$1.10E - 07$
				rs7589342	$1.45E - 04$
				rs12995333	$1.89E - 04$
				rs12053259	$2.31E - 04$
				rs6747023	$3.90E - 04$
				rs879900	$7.72E - 04$
White men	Nicotine	DSG3	$1.99E - 06$	rs6701037	$1.20E - 07$
				rs1057302	$3.93E - 07$
				rs6425323	$2.94E - 04$
				rs1057239	$3.35E - 04$

Furthermore, four other genes, FAM38B (cocaine dependence in black women), PTPRM (marijuana dependence in black women), CSMD1 (nicotine dependence in black women), and RELN (cocaine dependence in white men), contain at least one SNP that met the SNP-based relaxed threshold of significance. The gene-based P values for FAM38B, PTPRM, and RELN are $9.27E - 4$, $2.21E - 3$, and $8.53E - 4$, respectively, which are greater than yet at the same order as the relaxed threshold (P value $< 5.0E - 4$). For CSMD1, 1,934 SNPs were mapped into it. Its signal was mainly determined by only five SNPs: rs2624087 (P value = $8.50E - 6$), rs4875371 (P value = $4.0E - 4$), rs2623607 (P value = $6.89E - 4$), rs10503267 (P value = $7.22E - 4$), and rs4875372 (P value = $8.18E - 4$). Because there were only 5.3% of the SNPs (103 SNPs) with P value less than 0.05, the overall association from the gene became less significant.

3.2. Genome-Wide Significant Loci.

Since none of the SNPs attained the genome-wide significance for any dependence by the SNP-based method, in this section we will only focus on the results from the gene-based method.

Table 3 presents the genes with gene-based P value $< 1.0E - 5$. This method identified one genome-wide significant gene, DSG3 (P value = $1.99E - 6$) for nicotine dependence in white men. The P value of gene NCK2: $2.70E - 6$ is very close to the genome-wide significant threshold, which provided very strong evidence for the association of opiates in black men. As shown in Table 3, both NCK2 and DSG3 contained SNPs with strong signals; they are rs2377339 (P value = $1.09E - 7$) for NCK2 gene and rs6701037 (P value = $1.20E - 7$) and rs1057302 (P value = $3.93E - 7$) for DSG3 gene. However, none of these SNPs reached genome-wide significance.

4. Discussion

In this paper, we thoroughly analyzed the SAGE data from the SNP-based and gene-based methods, and compared the results obtained from these two methods. Specifically, for each sex-racial group, we performed association analysis for the six categories of substance dependence separately. The gene-based method appears to be more powerful in detecting susceptibility loci.

Most of the genes identified in our study are supported by various reports in the literature related to the genetics of substance dependence [47, 48]. Based on some of the genes that we identified, here common genetic variants among different substance dependencies may exist [49].

Overall, we did not detect any genome-wide significant SNP when using the SNPs-based method. However, one gene, DSG3, is genome-wide significantly ($P = 2.70E - 6$) associated with nicotine dependence in the white men, according to the gene-based method. Another gene, NCK2, is nearly genome-wide significant ($P = 2.7E - 6$) in its association with substance dependence.

The SNP-based method and gene-based method are closely related. In fact, the SNP-based method can be viewed as a gene-based method using the extreme function, namely, the minimal P value of the SNPs within a gene, whereas the typical gene-based method uses a weighted approach. The advantages and limitations of these two approaches are similar to those between the extreme function and a weighted average.

We should point out that both the SNP-based and gene-based methods have their own advantages and disadvantages. The SNP-based method has its unique strength in identifying genes with only a small number of significant SNPs. However, since the SNP-based method focuses on a single SNP at a time, it is less powerful to detect a gene whose SNPs have weak marginal effects, but a strong joint effect. In our analysis, 207 genes passed the relaxed gene-based threshold, whereas only 64 genes passed the relaxed SNP-based threshold.

Both the SNP-based and gene-based methods can be conducted conveniently in commonly available software, such as PLINK [45] for the SNP-based method and KGG [46] for the gene-based method. For the SNP-based analysis, PLINK is the most convenient platform. For the SAGE GWAS data, it took about 25 minutes to do a genome-wide SNP scan on a regular desktop computer (Intel Core 2, 4 GB Memory). In our gene-based analysis, we used the SNP-based association results and the linkage disequilibrium (LD) files from PLINK as the input to the KGG software. After this preparation, it took about 30 minutes to perform the gene-based association scan with the same desktop as mentioned above.

Authors' Contributions

X. Guo and Z. Liu are contributed equally to this work.

Acknowledgments

References

[1] M. I. McCarthy, G. R. Abecasis, L. R. Cardon et al., "Genome-wide association studies for complex traits: consensus, uncertainty and challenges," *Nature Reviews Genetics*, vol. 9, no. 5, pp. 356–369, 2008.

[2] P. R. Burton, D. G. Clayton, L. R. Cardon et al., "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, no. 7145, pp. 661–678, 2007.

[3] D. J. Hunter, P. Kraft, K. B. Jacobs et al., "A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer," *Nature Genetics*, vol. 39, no. 7, pp. 870–874, 2007.

[4] F. Dudbridge and A. Gusnanto, "Estimation of significance thresholds for genomewide association scans," *Genetic Epidemiology*, vol. 32, no. 3, pp. 227–234, 2008.

[5] E. E. Eichler, J. Flint, G. Gibson et al., "Missing heritability and strategies for finding the underlying causes of complex disease," *Nature Reviews Genetics*, vol. 11, no. 6, pp. 446–450, 2010.

[6] T. A. Manolio, F. S. Collins, N. J. Cox et al., "Finding the missing heritability of complex diseases," *Nature*, vol. 461, no. 7265, pp. 747–753, 2009.

[7] M. X. Li, H. S. Gui, J. S. H. Kwan, and P. C. Sham, "GATES: a rapid and powerful gene-based association test using extended Simes procedure," *American Journal of Human Genetics*, vol. 88, no. 3, pp. 283–293, 2011.

[8] J. Z. Liu, A. F. McRae, D. R. Nyholt et al., "A versatile gene-based test for genome-wide association studies," *American Journal of Human Genetics*, vol. 87, no. 1, pp. 139–145, 2010.

[9] X. Guo, Z. Liu, X. Wang, and H. Zhang, "Genetic association test for multiple traits at gene level," *Genetic Epidemiology*. In press.

[10] K. Wang, M. Li, and H. Hakonarson, "Analysing biological pathways in genome-wide association studies," *Nature Reviews Genetics*, vol. 11, no. 12, pp. 843–854, 2010.

[11] B. M. Neale and P. C. Sham, "The future of association studies: gene-based analysis and replication," *American Journal of Human Genetics*, vol. 75, no. 3, pp. 353–362, 2004.

[12] B. Lehne, C. M. Lewis, and T. Schlitt, "From SNPs to genes: disease association at the gene level," *PLoS ONE*, vol. 6, no. 6, Article ID e20133, 2011.

[13] L. J. Bierut, A. Agrawal, K. K. Bucholz et al., "A genome-wide association study of alcohol dependence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 11, pp. 5082–5087, 2010.

[14] X. Chen, K. Cho, B. H. Singer, and H. Zhang, "The nuclear transcription factor PKNOX2 is a candidate gene for substance dependence in European-origin Women," *PLoS ONE*, vol. 6, no. 1, Article ID e16002, 2011.

[15] J. Clarimon, R. R. Gray, L. N. Williams et al., "Linkage disequilibrium and association analysis of α-synuclein and alcohol and drug dependence in two American Indian populations," *Alcoholism: Clinical and Experimental Research*, vol. 31, no. 4, pp. 546–554, 2007.

[16] D. M. Dick, H. J. Edenberg, X. Xuei et al., "Association of GABRG3 with alcohol dependence," *Alcoholism: Clinical and Experimental Research*, vol. 28, no. 1, pp. 4–9, 2004.

[17] H. J. Edenberg, D. M. Dick, X. Xuei et al., "Variations in GABRA2, encoding the α2 subunit of the GABA a receptor, are associated with alcohol dependence and with brain oscillations," *American Journal of Human Genetics*, vol. 74, no. 4, pp. 705–714, 2004.

[18] H. J. Edenberg and T. Foroud, "The genetics of alcoholism: identifying specific genes through family studies," *Addiction Biology*, vol. 11, no. 3-4, pp. 386–396, 2006.

[19] H. J. Edenberg, J. Wang, H. Tian et al., "A regulatory variation in OPRK1, the gene encoding the κ-opioid receptor, is associated with alcohol dependence," *Human Molecular Genetics*, vol. 17, no. 12, pp. 1783–1789, 2008.

[20] T. Foroud, L. F. Wetherill, T. Liang et al., "Association of alcohol craving with α-synuclein (SNCA)," *Alcoholism: Clinical and Experimental Research*, vol. 31, no. 4, pp. 537–545, 2007.

[21] J. Gelernter, R. Gueorguieva, H. R. Kranzler et al., "Opioid receptor gene (OPRM1, OPRK1, and OPRD1) variants and response to naltrexone treatment for alcohol dependence: results from the VA Cooperative Study," *Alcoholism: Clinical and Experimental Research*, vol. 31, no. 4, pp. 555–563, 2007.

[22] T. Reich, "A genomic survey of alcohol dependence and related phenotypes: results from the Collaborative Study on the Genetics of Alcoholism (COGA)," *Alcoholism: Clinical and Experimental Research*, vol. 20, supplement 8, pp. 133A–137A, 1996.

[23] T. Reich, H. J. Edenberg, A. Goate et al. et al., "Genome-wide search for genes affecting the risk for alcohol dependence,"

American Journal of Medical Genetics, vol. 81, no. 3, pp. 207–215, 1998.

[24] J. Song, D. L. Koller, T. Foroud et al., "Association of GABAA receptors and alcohol dependence and the effects of genetic imprinting," *American Journal of Medical Genetics*, vol. 117, no. 1, pp. 39–45, 2003.

[25] B. Tabakoff, L. Saba, M. Printz et al. et al., "Genetical genomic determinants of alcohol consumption in rats and humans," *BMC Biology*, vol. 7, article 70, 2009.

[26] J. C. Wang, A. L. Hinrichs, S. Bertelsen et al., "Functional variants in TAS2R38 and TAS2R16 influence alcohol consumption in high-risk families of African-American origin," *Alcoholism: Clinical and Experimental Research*, vol. 31, no. 2, pp. 209–215, 2007.

[27] K. S. Wang, X. F. Liu, Q. Y. Zhang, Y. Pan, N. Aragam, and M. Zeng :, "A meta-analysis of two genome-wide association studies identifies 3 new loci for alcohol dependence," *Journal of Psychiatric Research*, vol. 45, no. 11, pp. 1419–1425, 2011.

[28] H. Zhang, H. R. Kranzler, B. Z. Yang, X. Luo, and J. Gelernter, "The OPRD1 and OPRK1 loci in alcohol or drug dependence: OPRD1 variation modulates substance dependence risk," *Molecular Psychiatry*, vol. 13, no. 5, pp. 531–543, 2008.

[29] G. Kalsi, P. H. Kuo, F. Aliev et al., "A systematic gene-based screen of chr4q22-q32 identifies association of a novel susceptibility gene, DKK2, with the quantitative trait of alcohol dependence symptom counts," *Human Molecular Genetics*, vol. 19, no. 20, Article ID ddq326, p. 4121, 2010.

[30] P. H. Kuo, G. Kalsi, C. A. Prescott et al., "Association of ADH and ALDH genes with alcohol dependence in the Irish Affected Sib Pair Study of alcohol dependence (IASPSAD) Sample," *Alcoholism: Clinical and Experimental Research*, vol. 32, no. 5, pp. 785–795, 2008.

[31] L. J. Bierut, "Genetic variation that contributes to nicotine dependence," *Pharmacogenomics*, vol. 8, no. 8, pp. 881–883, 2007.

[32] L. J. Bierut, P. A. F. Madden, N. Breslau et al., "Novel genes identified in a high-density genome wide association study for nicotine dependence," *Human Molecular Genetics*, vol. 16, no. 1, pp. 24–35, 2007.

[33] N. Caporaso, F. Gu, N. Chatterjee et al., "Genome-wide and candidate gene association study of cigarette smoking behaviors," *PLoS ONE*, vol. 4, no. 2, Article ID e4653, 2009.

[34] L. S. Chen, E. O. Johnson, N. Breslau et al., "Interplay of genetic risk factors and parent monitoring in risk for nicotine dependence," *Addiction*, vol. 104, no. 10, pp. 1731–1740, 2009.

[35] N. L. Saccone, S. F. Saccone, A. L. Hinrichs et al., "Multiple distinct risk loci for nicotine dependence identified by dense coverage of the complete family of nicotinic receptor subunit (CHRN) genes," *American Journal of Medical Genetics, Part B*, vol. 150, no. 4, pp. 453–466, 2009.

[36] S. F. Saccone, A. L. Hinrichs, N. L. Saccone et al., "Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs," *Human Molecular Genetics*, vol. 16, no. 1, pp. 36–49, 2007.

[37] G. R. Uhl, T. Drgon, C. Johnson et al., "Genome-wide association for smoking cessation success: participants in the Patch in Practice trial of nicotine replacement," *Pharmacogenomics*, vol. 11, no. 3, pp. 357–367, 2010.

[38] G. R. Uhl, Q. R. Liu, T. Drgon et al., "Molecular genetics of successful smoking cessation: convergent genome-wide association study results," *Archives of General Psychiatry*, vol. 65, no. 6, pp. 683–693, 2008.

[39] R. B. Weiss, T. B. Baker, D. S. Cannon et al., "A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction," *PLoS Genetics*, vol. 4, no. 7, Article ID e1000125, 2008.

[40] M. J. Ahn, H. H. Won, J. Lee et al. et al., "The 18p11. 22 locus is associated with never smoker non-small cell lung cancer susceptibility in Korean populations," *Human Genetics*, vol. 131, no. 3, pp. 365–372, 2012.

[41] D. M. Dick, J. Meyers, F. Aliev et al., "Evidence for genes on chromosome 2 contributing to alcohol dependence with conduct disorder and suicide attempts," *American Journal of Medical Genetics, Part B*, vol. 153, no. 6, pp. 1179–1188, 2010.

[42] P. F. Giampietro, C. McCarty, B. Mukesh et al., "The role of cigarette smoking and statins in the development of postmenopausal osteoporosis: a pilot study utilizing the marshfield clinic personalized medicine cohort," *Osteoporosis International*, vol. 21, no. 3, pp. 467–477, 2010.

[43] G. Joslyn, A. Ravindranathan, G. Brush, M. Schuckit, and R. L. White, "Human variation in alcohol response is influenced by variation in neuronal signaling genes," *Alcoholism: Clinical and Experimental Research*, vol. 34, no. 5, pp. 800–812, 2010.

[44] L. J. Zuo, J. Gelernter, C. K. Zhang et al. et al., "Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q," *Neuropsychopharmacology*, vol. 37, no. 2, pp. 557–566, 2012.

[45] S. Purcell, B. Neale, K. Todd-Brown et al., "PLINK: a tool set for whole-genome association and population-based linkage analyses," *American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–575, 2007.

[46] http://bioinfo1.hku.hk:13080/kggweb//.

[47] K. R. Merikangas, M. Stolar, D. E. Stevens et al., "Familial transmission of substance use disorders," *Archives of General Psychiatry*, vol. 55, no. 11, pp. 973–979, 1998.

[48] W. R. True, A. C. Heath, J. F. Scherrer et al., "Interrelationship of genetic and environmental influences on conduct disorder and alcohol and marijuana dependence symptoms," *American Journal of Medical Genetics*, vol. 88, no. 4, pp. 391–397, 1999.

[49] M. D. Li and M. Burmeister, "New insights into the genetics of addiction," *Nature Reviews Genetics*, vol. 10, no. 4, pp. 225–231, 2009.