Published in final edited form as:

Am J Addict. 2014 November ; 23(6): 526–539. doi:10.1111/j.1521-0391.2014.12147.x.

Genome-wide association discoveries of alcohol dependence

Lingjun Zuo, PhD^{1,*}, Lingeng Lu, PhD², Yunlong Tan³, Xinghua Pan, PhD⁴, Yiqiang Cai, PhD⁵, Xiaoping Wang⁶, Jiang Hong⁷, Chunlong Zhong⁸, Fei Wang¹, Xiang-yang Zhang⁹, Lauren A. Vanderlinden, PhD¹⁰, Boris Tabakoff, PhD¹⁰, and Xingguang Luo, PhD^{1,3,*}

¹Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

²Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA

³Biological Psychiatry Research Center, Beijing Huilongguan Hospital, Beijing, China

⁴Department of Genetics, Yale University School of Medicine, New Haven, CT, USA

⁵Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA

⁶Department of Neurology, First People's Hospital, Shanghai Jiaotong University, Shanghai, China

⁷Department of Internal Medicine, First People's Hospital, Shanghai Jiaotong University, Shanghai, China

⁸Department of Neurosurgery, Renji Hospital, Shanghai Jiaotong University, Shanghai, China

⁹Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas, USA

¹⁰Department of Pharmacology, University of Colorado School of Medicine, Aurora, CO, USA

Abstract

Objective—To report the genome-wide significant and/or replicable risk variants for alcohol dependence and explore their potential biological functions.

Methods—We searched in PubMed for all genome-wide association studies (GWASs) of alcohol dependence. The following three types of the results were extracted: (1) genome-wide significant associations in an individual sample, the combined samples, or the meta-analysis ($p < 5 \times 10^{-8}$); (2) top-ranked associations in an individual sample ($p < 10^{-5}$) that were nominally replicated in other samples ($p < 0.05$); and (3) nominally replicable associations across at least three independent GWAS samples ($p < 0.05$). These results were meta-analyzed. *cis*-eQTLs in human, RNA expression in rat and mouse brain and bioinformatics properties of all of these risk variants were analyzed.

*Corresponding authors and reprints: Lingjun Zuo, PhD and Xingguang Luo, PhD; 1010 Pleasant Hill Road, Orange, CT 06477. Xingguang.Luo@yale.edu or Lingjun.Zuo@yale.edu; Tel: 203-932-5711..

Conflict of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this paper.

Results—The variants located within *ADH* cluster were significantly associated with alcohol dependence at genome-wide level ($p < 5 \times 10^{-8}$) in at least one sample. Some associations with the *ADH* cluster were replicable across six independent GWAS samples. The variants located within or near *SERINC2*, *KIAA0040*, *MREG-PECR* or *PKNOX2* were significantly associated with alcohol dependence at genome-wide level ($p < 5 \times 10^{-8}$) in meta-analysis or combined samples, and these associations were replicable across at least one sample. The associations with the variants within *NRD1*, *GPD1L-CMTM8* or *MAP3K9-PCNX* were suggestive ($5 \times 10^{-8} < p < 10^{-5}$) in some samples, and nominally replicable in other samples. The associations with the variants at *HTR7* and *OPA3* were nominally replicable across at least three independent GWAS samples ($10^{-5} < p < 0.05$). Some risk variants at the *ADH* cluster, *SERINC2*, *KIAA0040*, *NRD1* and *HTR7* had potential biological functions.

Conclusion—The most robust risk locus was the *ADH* cluster. *SERINC2*, *KIAA0040*, *NRD1* and *HTR7* were also likely to play important roles in alcohol dependence. *PKNOX2*, *MREG*, *PECR*, *GPD1L*, *CMTM8*, *MAP3K9*, *PCNX* and *OPA3* might play less important roles in risk for alcohol dependence based on the function analysis. This conclusion will significantly contribute to the post-GWAS follow-up studies on alcohol dependence.

Keywords

Alcohol dependence; Genome-wide association; Long non-coding RNA

Introduction

Alcohol dependence is a chronic but often progressive disease that includes problems in controlling one's drinking, being preoccupied with alcohol, continuing to use alcohol even when it causes problems, having to drink more to get the same effect (physical dependence), or having withdrawal symptoms when one rapidly decreases or stops drinking. Twin, family, and adoption studies have indicated that the susceptibility to alcohol dependence is genetically influenced, with a heritability of 50–60% (1–3). At least part of the cause of alcohol dependence is attributable to gene variation.

The development of alcohol dependence has been proposed to be related to the metabolism of alcohol, so that the genes in alcohol metabolism pathway, e.g., alcohol dehydrogenase (*ADH*) and aldehyde dehydrogenase (*ALDH*) genes, are candidates for alcohol dependence. Alcohol dependence is, however, considered to be a brain disorder, so that the genes related to the neurotransmitters are also hypothesized to be candidate genes. These hypotheses help in finding many risk genes for alcohol dependence. Most of the risk genes implicated have been from (1) classical neurotransmitter signaling systems, including the dopaminergic (e.g., *MAOA*, *COMT*, and *NCAM1-TTC12-ANKK1-DRD2*), serotonergic (e.g., *SLC6A4* and *HTR2B*), GABAergic (e.g., *GABRA2* and *GABRG1*), and cholinergic systems (e.g., *CHRM2* and *CHRNA5-CHRNA3-CHRNA4*) (4); (2) non-classical neurotransmitter signaling systems (e.g., *CRHR1*) (5); (3) the ethanol metabolic pathway (e.g., *ADH1B*, *ADH1C*, *ADH4*, *ADH5* and *ALDH2*) (6–8); and (4) the opioidergic signaling pathway (e.g., *OPRM1*, *OPRD1* and *OPRK1*) (9) (reviewed by Zuo et al. (10)). With the development of genotyping technology and the significantly decreased costs, scientists have scanned the whole genome to search

for risk genes since 2010. These data-driven genome-wide association studies (GWASs) found novel risk genes, breaking through the limitation of the need for an *a priori* hypothesis.

GWASs reported numerous associations. Correction for multiple testing at genome-wide level ($\alpha=5\times 10^{-8}$) is highly necessary to reduce false positives. Furthermore, a true association is usually replicable in different samples; and thus replication in independent samples, especially in genetically-distinct samples, is also a reliable way to reduce false positives. In spite of these, a significant and replicable association does not necessarily reflect a true cause-result relationship. If a variant is functional, the possibility of a cause-result relationship will increase. In this work, we extensively reviewed the findings of GWASs published so far, to report the genome-wide significant and/or replicable risk genes for alcohol dependence with potential biological functions, which is highly necessary for the follow-up studies in the post-GWAS era.

Materials and Methods

We presented previous GWAS findings in Table 1. Some previous GWASs combined heterogeneous samples and then controlled for population stratification and admixture effects. These combined effects in pooled samples are also presented in the “meta/combined” column (not underlined) in Table 1. Furthermore, we extended and validated these findings using other public data available to us, including meta-analysis, *cis*-acting expression of quantitative locus (*Cis*-eQTL) analysis, rat brain transcriptome analysis, bioinformatics analysis and SNP-disease association analysis.

Selection of GWASs and presentation of findings

We searched in PubMed and reviewed 45 GWASs of alcohol dependence (i.e., “alcoholism”, “alcohol dependence” and “alcohol addiction”). All diagnoses met DSM-IV criteria for alcohol dependence. The vast majority of GWASs produced results that were not significant at genome-wide level and were not replicated in any other studies and thus were not presented in this study. Only the following three types of results that had odd ratios, directions of effects and p values in these papers were selected and presented in Table 1: 1) genome-wide significant associations ($p<5\times 10^{-8}$) in an individual sample, the combined samples of heterogeneous samples by controlling for admixture, or the meta-analysis of completely independent samples; 2) top-ranked associations in an individual sample ($p<10^{-5}$) that were nominally replicated (at individual marker level) in other samples ($p<0.05$); or 3) nominally replicable associations across at least three independent GWAS samples ($p<0.05$). The ethnicity of each sample is shown in Table 1. Except for the Australian samples (10, 11) and an American sample (12) that were family-based, all samples were case-control designed. Most studies controlled three covariates including sex, age and principal components (PCs) of ancestries in association analysis (data not shown).

Meta-analysis

To combine the effects of completely independent samples (underlined in Table 1), meta-analysis (fixed effect model) was performed to derive the combined p values (underlined in

Table 1) using the program METAL (13). This program is designed to facilitate meta-analysis of large datasets in a convenient, rapid and memory efficient manner. It is especially appropriate when data from the individual studies cannot be analyzed together because of differences in ethnicity, phenotype distribution, gender or constraints in sharing of individual level data imposed. It combines p-values across studies, taking sample size and direction of effect (i.e., $\log(\text{OR})$) into account. First, for each marker, a reference allele is selected and a z-statistic characterizing the evidence for association is calculated. The z-statistic summarizes the magnitude and the direction of effect relative to the reference allele and all studies are aligned to the same reference allele. Next, an overall z-statistic and combined p-value are then calculated from a weighted sum of the individual statistics. Weights are proportional to the square-root of the number of individuals contributing to that particular p-value examined in each sample and selected such that the squared weights sum to 1.0. When the same sample was studied by different GWASs, which usually had similar results, only one of the results (underlined in Table 1) was included in the meta-analysis.

Cis-eQTL analysis

There is a substantial gap between SNP-alcohol dependence associations and understanding how these SNPs contribute to alcohol dependence. The functional effects of SNPs on alcohol dependence can be mediated through several mechanisms. Variation in gene expression is an important mechanism underlying susceptibility to alcohol dependence. The abundance of a gene transcript is directly modified by polymorphism in regulatory elements. Consequently, transcript abundance might be considered as a quantitative trait that can be mapped with considerable power, which has been named expression QTLs (eQTLs). To examine the relationships between all risk SNPs (listed in Table 1) and mRNA expression levels of the proximate protein-coding genes, we examined the expression data in human lymphoblastoid cell lines from 270 unrelated HapMap individuals including six different populations (14), in 93 autopsy-collected frontal cortical brain tissue samples with no defined neuropsychiatric condition, and in 80 peripheral blood mononuclear cell (PBMC) samples collected from living healthy donors (15). The expression data were normalized across all samples for a tissue type (four separate normalizations) per Affymetrix PLIER protocol with a sketch-quantile normalization procedure (Affymetrix Expression Console). Differences in the distribution of mRNA expression levels between SNP genotypes were analyzed using a linear regression model implemented in the program SNPEXPRESS (15) by correcting for age, sex and principle component (PC) scores, separately by tissue. These PCs were calculated from all expression data for both brain and PBMC samples and the first ten PCs were included in the eQTL analysis.

Rat brain transcriptome analysis

We also assessed the mRNA expression of these risk genes in brains of two strains of rats (i.e., BN-Lx/ CubPrin and SHR/OlaPrin) using RNA-Seq technology. After ribosomal RNA depletion, the total RNAs were reversely transcribed to cDNAs that were then sequenced on the Illumina HiSeq2000 system. RNA-Seq data were available from the PhenoGen website (<http://phenogen.ucdenver.edu>) (16). After reads were trimmed for adapters and quality, they were aligned to the RGSC5.0/rn5 version of the rat genome using TopHat2 (17). RNA expression was quantitated for genes identified in a *denovo* genome-guided transcriptome

reconstruction using CuffLinks and all uniquely aligned reads (18). Read fragments per kilobase of transcript per million reads (FPKM) for the candidate genes were calculated. To confirm these RNA-Seq results in rats, gene expression data were also generated using whole brain tissue of naïve mice using Affymetrix mouse whole genome oligonucleotide microarrays (i.e., GeneChip Mouse Genome 430 v2.0 Array, Affymetrix, Santa Clara, CA) (16). A BXD recombinant inbred panel of 30 strains (parental strains are C57BL/6J & DBA/2J) was assessed and the microarray data were normalized using MAS5 method. When a specific mRNA detected by RNA-Seq in rats was present in >90% of the mouse samples, this expression was confirmed.

Bioinformatics analysis

All of the risk SNPs listed in Table 1 were analyzed using UCSC Genome Browser data or other bioinformatics analysis software (e.g., FuncPred from <http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm>) to see whether they were located in the transcription factor binding sites (TFBS), within methylated CpG islands, within copy number variations (CNVs) or in exonic splicing silencer (ESS) or enhance (ESE). These locations were helpful for us to predict whether these SNPs were potentially functional. Additionally, Polyphen (19) was applied to predict whether these risk SNPs affected the protein function or structure, and MFOLD (20) was applied to predict whether these risk SNPs altered the secondary RNA structure.

SNP-disease association test

The long non-coding RNA (LncRNA) (>200nt) proximate to each risk gene was explored in NCBI dbSNP and UCSC Genome Browser databases. Associations between SNPs located within these LncRNAs and alcohol dependence were tested in three datasets. These datasets included one European-American SAGE+COGA cohort (1,409 cases with alcohol dependence and 1,518 controls), one African-American SAGE+COGA cohort (681 cases and 508 controls) (dbGaP# phs000092.v1.p1 and phs000125.v1.p1) and one African-American Yale cohort (1429 cases and 498 controls) (dbGaP# phs000425.v1.p1) (<http://www.ncbi.nlm.nih.gov/gap>). Affected subjects met lifetime DSM-IV criteria for alcohol dependence (21). All subjects gave written informed consent to participating in protocols approved by the relevant institutional review boards (IRBs). All subjects were de-identified in this study that was approved by Yale IRB. Detailed demographic information for these cohorts is available elsewhere (22–29). The SAGE+COGA and Yale cohorts were genotyped on the Illumina Human 1M and HumanOmni1_Quad_v1-0_B beadchips, respectively. The association analysis was performed using unconditional logistic regression models implemented in PLINK (30). Diagnosis and alleles each served as the dependent and independent variables, respectively, with sex, age and the first 10 principal components (PCs) of ancestries as covariates. These PCs were estimated from all autosomal SNPs that had a call rate >95%, using principal component analysis (PCA) implemented in the program EIGENSTRAT (31). These SNPs were stringently cleaned using the QC criteria as we previously published (25) before PCA. Each individual received scores on each principal component. Because related subjects, non-EA and non-AA subjects, and any population group outliers have been excluded during data cleaning, their effects on PCA were removed.

Although different sets of autosomal SNPs were used to estimate these PCs in different datasets, the first 10 PCs all explained >95% of variance of the samples in different studies.

Results

The variants located within *ADH* cluster on Chr4 (32–35) were found to be significantly associated with alcohol dependence at genome-wide level ($p < 5 \times 10^{-8}$) in at least one sample. Some associations with the *ADH* cluster were replicable across six independent GWAS samples including Koreans (34), other Asians (34), Germans (32, 33), European-Americans (33, 35, 36) and African-Americans (22, 36). The variants located within or between five other loci were found to be significantly associated with alcohol dependence at genome-wide level ($p < 5 \times 10^{-8}$) in meta-analysis or in the pooled samples, and these associations were replicable across at least one sample ($p < 0.05$). They were serine incorporator 2 gene (*SERINC2*) (10) and *KIAA0040* on Chr1 (11, 22, 25), melanoregulin gene (*MREG*) and peroxisomal trans-2-enoyl-CoA reductase gene (*PECR*) on Chr2 (32, 33), and PBX/knotted 1 homeobox 2 gene (*PKNOX2*) on Chr11 (11, 22, 37). Additionally, the variants within or between five loci were found to be suggestively ($5 \times 10^{-8} < p < 10^{-5}$) associated with alcohol dependence in some samples, and nominally replicable in other samples ($p < 0.05$). They were Nardilysin gene (*NRD1*) on Chr1 (11, 22), glycerol-3-phosphate dehydrogenase 1-like gene (*GPD1L*) and CKLF-like MARVEL transmembrane domain containing 8 gene (*CMTM8*) on Chr3 (33), mitogen-activated protein kinase kinase kinase 9 gene (*MAP3K9*) and pecanex homolog gene (*PCNX*) on Chr14 (32). Finally, two loci were found to be nominally associated with alcohol dependence across at least three independent GWAS samples ($10^{-5} < p < 0.05$). They were 5-hydroxytryptamine (serotonin) receptor 7 gene (*HTR7*) on Chr10 (12, 38) and optic atrophy 3 gene (*OPA3*) on Chr19 (23, 39) (Table 1).

cis-eQTL analysis showed that rs2173201 between *ADH1B* and *ADH1C*, and rs4478858 at *SERINC2* (10) had nominally significant regulatory effects on the expression of *ADH1C* and *SERINC2* transcripts, respectively, in human brain ($p = 0.044$ for rs2173201, and $p = 0.024$ for rs4478858) and/or PBMC samples ($p = 0.026$ for rs4478858). Six *ADH* markers had nominally significant regulatory effects on the expression of *ADH* transcripts in the HapMap CEU parent sample ($2.7 \times 10^{-6} \leq p \leq 0.024$). All of the 5 SNPs at *KIAA0040* had nominally significant regulatory effects on the expression of *KIAA0040* transcripts across 2 to 5 HapMap samples ($2.0 \times 10^{-4} \leq p \leq 0.028$). rs2842576 at *NRD1* had a nominally significant regulatory effect on the expression of *NRD1* in HapMap YRI parent sample ($p = 0.028$). rs7916403 at *HTR7* had nominally significant regulatory effects on the expression of *HTR7* in both HapMap CEU samples ($p = 0.017$ in CEU Children population and 9.7×10^{-3} in CEU Parent population, respectively) (Table 2). After correction for multiple testing, many of these effects remained significant ($\alpha = 1.3 \times 10^{-3}$).

When the expression of the aforementioned candidate genes was examined in the brains of mouse and rat, the results in Table 3 were obtained. The only *Adh* gene expressed in rodent brain tissue was *Adh5*. One can consider this gene as part of the *Adh* cluster. The qualitative levels of expression of the other candidate genes in rodent brains are displayed in Table 3. The genes that had expression levels above background (FPKM > 1) included *SERINC2*,

MREG, PECR, PKNX2, NRD1, GPD1L, MAP3K9, PCNX, HTR7 and *OPA3*. Other genes that were not detected in rodent brains included *KIAA0040* and *CMTM8*.

Bioinformatics analysis shows that 4 SNPs are located in the transcription factor binding sites (TFBS), including rs4147541 at 5' regulatory region of *ADH1C*, rs4478858 at *SERINC2* and rs1057239 and rs1894709 at *KIAA0040*. They may affect the local DNA conformation, and thereby influence the binding of TFBS (40). Two SNPs, rs2275436 at *SERINC2* and rs6425323 at *KIAA0040*, are located within CpG islands, whose methylation status may affect the expression of the neighbor genes (41). All of the three significant SNPs at *SERINC2* are located within CNVs. rs1229984 (Arg47Asn) at *ADH1B* is located in exonic splicing silencers (ESS) or enhancers (ESE). rs2066702 (Arg370Cys) at *ADH1B* is predicted to be with high confidence to affect *ADH1B* function or structure. Furthermore, 8 SNPs are predicted to significantly or highly significantly alter the RNA secondary structures, including rs28864441, rs2173201 and rs4147541 at *ADH* cluster, rs2275436 at *SERINC2*, rs750338 at *PKNX2*, rs2842576 at *NRD1*, rs9825310 between *GPD1L* and *CMTM8*, and rs7916403 at *HTR7*. Eight SNPs are predicted to mildly alter the RNA secondary structures, including rs1789891 and rs4699748 at *ADH* cluster, rs4478858 at *SERINC2*, rs6701037 and rs1894709 at *KIAA0040*, rs1344694 between *MREG* and *PECR*, rs36563 between *MAP3K9* and *PCNX* and rs8111589 at *OPA3* (Table 2). They may affect the downstream activities of the RNA molecules (42).

The LncRNA proximate to each risk gene is listed in Table 4. Twelve among the total 16 LncRNAs are large intergenic non-coding RNAs (LincRNAs). Fourteen of them are located within 100kb from the proximate risk genes, with lengths of 1,055bp to 231,974bp. The LncRNAs NR_037884.1 is antisense to *ADH* cluster. Two promoter LncRNAs, XR_110219.3 and TCONS_00018282, are bidirectional to *MAP3K9* and *HTR7*, respectively (Table 4). Furthermore, many SNPs within 10 LncRNAs were nominally significantly ($p < 0.05$) associated with alcohol dependence (Table 5), including 14 SNPs within NR_037884.1 in *ADH* cluster (2.6×10^{-3} p 0.049), 2 SNPs within TCONS_00001448 (1.4×10^{-4} p 0.025), 2 SNPs within NR_034112.1 (1.9×10^{-4} p 2.2×10^{-3}), 3 SNPs within NR_027085.1 (4.0×10^{-3} p 0.038), and 1 SNP within TCONS_00000180 proximate to *SERINC2* ($p = 4.0 \times 10^{-3}$), 3 SNPs within TCONS_00000690 proximate to *KIAA0040* (0.016 p 0.048), 15 SNPs within NR_037195.1 proximate to *MREG* (6.3×10^{-4} p 0.047), 1 SNP within XR_110219.3 at the promoter of *MAP3K9* ($p = 0.017$), and 1 SNP within TCONS_00018282 at the promoter of *HTR7* ($p = 0.032$). Most of these associations became suggestive after correction for multiple testing.

Discussion

By reviewing the genome-wide association studies of alcohol dependence and exploring the potential biological functions of the reported risk variants and genes, we concluded that the *ADH* cluster, *SERINC2*, *KIAA0040*, *NRD1* and *HTR7* were likely to play important roles in risk for alcohol dependence, but *PKNX2*, *MREG*, *PECR*, *GPD1L*, *CMTM8*, *MAP3K9*, *PCNX* and *OPA3* might play less important roles in this disease risk. The *ADH* cluster is majorly expressed in the liver and *SERINC2*, *NRD1* and *HTR7* are majorly expressed in the brain, which is consistent with the hypotheses for alcohol dependence introduced above.

ADH cluster was the only locus that was significantly associated with alcohol dependence at the genome-wide level in at least one individual sample. Most of the risk *ADH* variants were potentially functional. The most robust association was found at the *ADH* cluster. At least three GWASs detected significant association signals in Asians (34), Germans (33), European-Americans and African-Americans (36) at genome-wide level, and other four GWASs in Koreans (34), Germans (32) and European-Americans (22, 33, 35) and African-Americans (22) replicated them. There have also been numerous candidate gene studies that reported strong associations between *ADH* cluster and alcohol dependence (summarized by Luo et al. (43) 2006). Recent meta-analysis demonstrated that rs1229984 (Arg48His) at *ADH1B* ($p=10^{-36}$) and rs698 (Ile350Val) at *ADH1C* ($p=10^{-8}$) were highly strongly associated with alcohol dependence (6, 7), supporting the GWAS findings. ADHs are primary enzymes that are responsible for the oxidation of alcohol in the liver. Their activities have been well-known to influence the risk of becoming alcoholic.

Most genome-wide significant risk variants (except rs1229984 and rs2066702) identified by GWASs are typically located in non-coding regions (Table 1), without clear effects on protein structure. Non-coding genetic variants can have effects on the function of proximate genes by altering the transcription, splicing or stability of the mRNAs or ncRNAs. Recent evidence suggests that lncRNAs are involved in a wide variety of cellular functions, including epigenetic silencing, transcriptional regulation, RNA processing, RNA modification (44–46), and implied in plasticity in the nervous system (47), neuropathological process (48), neurotransmission (49), and stress response (47). Dysregulation of many lncRNAs has been found to contribute to substance use disorders including alcohol, nicotine, heroin and cocaine dependence. For example, *NEAT2*, an lncRNA regulating synapse formation (50), was up-regulated in human brains of alcoholics (51); *NEAT2*, *NEAT1*, *MIAT* and *MEG3* were up-regulated in human brains (NAc) of heroin abusers (52); and *NEAT2*, *MIAT*, *MEG3* and *EMX2OS* were elevated in the NAc of cocaine abusers (52); smokers had dramatically elevated *H19* expression in airway epithelium (53); demethylation of *H19* was correlated to chronic alcohol use in males (54); and many lncRNAs played roles in cocaine-induced neural plasticity in NAc and risk for cocaine dependence (48). Therefore, we also explored the lncRNAs proximate to the risk genes in this study. Interestingly, a large antisense-overlapping lncRNA (>200k nt) covers *ADHs* 5, 4, 6 and 1A. Fourteen SNPs located within this lncRNA were associated with alcohol dependence. Such an lncRNA frequently uses diverse transcriptional and post-transcriptional gene regulatory mechanisms to carry out a wide variety of biological roles (47). It usually has a tendency to undergo fewer splicing events and typically shows lower abundance than sense mRNA transcripts (55). Additionally, a second smaller sense lncRNA (1,055 nt) is located between *ADH4* and *ADH6*. Such an lncRNA usually collaborates with chromatin modifying proteins (PRC2, CoREST and SCMx) to regulate expression of proximate genes (56). Accordingly, we can postulate that these lncRNAs within *ADH* cluster might regulate the expression of the *ADH* genes and might be functional components of the pathways through which the *ADH* variants influenced risk for alcohol dependence. Alteration of the expression for these lncRNAs in alcoholics deserves to be tested in the future, to investigate the *cis*-acting regulatory effects of risk variants within or proximate to these lncRNAs on the expression levels.

We carefully examined brain tissue of rats and mice using RNA-Seq and Affymetrix 3' level arrays respectively for the presence of mRNA coding for all known ADHs. We saw evidence for expression of *ADH5* (Table 3), which was consistent with the work of Galter et al. (57) which examined the distribution of Class I, Class III and Class IV ADH's in tissues of mouse, rat and human brain and also found that the Class III ADH (equivalent to *ADH5*) was the major one appearing in the brain of mice, rats and humans. *ADH5* in brain is not ubiquitously distributed but is evident in the dopaminergic neurons of the substantia nigra, the red nucleus, the hippocampus, cortex and the granule and Purkinje cells of the cerebellum. The common name for *ADH5* is glutathione dependent formaldehyde dehydrogenase (chi polypeptide), and a role ascribed to this enzyme in brain has been in vitamin A metabolism, particularly the transformation of retinol to retinal. The retinal is transformed to retinoic acid which plays a major role in regulation of gene transcription. *ADH5* has not been considered to be important in the metabolism of ethanol because of its reported high K_m (low affinity) for ethanol (58). However, recent data have demonstrated that *ADH5* activity with ethanol increases significantly if the enzyme activity is tested in a hydrophobic environment such as that encountered in hepatocytes or neurons (59). The K_m for ethanol is reduced in hydrophobic solutions (59) and *ADH5* becomes an important participant in the metabolism of ethanol at BAC of > 50 mM (~ 200 mg%) (60). It is interesting to note that the pharmacokinetic parameters of the early absorptive phase of ethanol metabolism show that ethanol (particularly if consumed on an empty stomach) partitions into brain and significantly higher concentrations of ethanol can be measured in brain as compared to blood in the early phases of ethanol administration. The implications for conversion of ethanol to acetaldehyde in brain have recently been detailed for rats bred to consume high levels of ethanol (61).

SERINC2 on Chr1 is also a genome-wide significant and replicable risk gene for alcohol dependence in the population of European descent. *SERINC2* encodes serine incorporator 2 (Serinc2) that belongs to a cell membrane carrier protein family of Serinc1–5. High concentration of Serinc2 is seen in neurons of the human hippocampus and cerebral cortex (62) and also in the rodent brain in the present study. It serves as an effector molecule that incorporates serine into membranes and facilitates the synthesis of two serine-derived membrane lipids, phosphatidylserine and sphingolipids (long-chain amino alcohols) (63). Phosphatidylserine supplement has been approved by the U.S. Food and Drug Administration to treat memory deficit disorders such as Alzheimer's disease and other forms of dementia, to support healthy cognitive function during aging, and to remediate cognitive deficits as a result of heavy drinking and cigarette smoking. *SERINC2* might also influence alcohol dependence via the glutaminergic pathway that is known to play important roles in alcohol intoxication and withdrawal (64). *SERINC2* is enriched with numerous structural variants (see NCBI dbSNP). The three genome-wide significant risk SNPs are also located within CNVs and may thus be functional. Furthermore, *SERINC2* is close to four LncRNAs. Eight SNPs within these LncRNAs were associated with alcohol dependence, suggesting an additional potential regulatory pathway for *SERINC2* in influencing alcohol dependence.

KIAA0040 on Chr1 has long been considered to be a lincRNA until recently when it was reported to code for an HLA-DR11-restricted T-cell epitope. It is expressed in multiple tissues and organs including brain. In the present study, all significant risk SNPs at *KIAA0040* had significant eQTL signals in human lymphoblastoid cell lines. Bioinformatics analysis also supported that these variants might regulate the gene expression. *TNN* and *TNR* flank *KIAA0040*. *TNN*, which encodes tenascin-N, is involved in neurite outgrowth and cell migration in hippocampal explants. *TNR*, which encodes tenascin-R, is an extracellular matrix protein expressed primarily in the central nervous system and has been related to multiple brain diseases. *KIAA0040* is close to *ASTN1* that has been associated with alcohol dependence or substance dependence by two studies (65, 66). *ASTN1* has also been implicated in autism, schizophrenia, bipolar disorder and attention deficit hyperactivity disorder (ADHD) (reviewed by Hill et al. (66) 2012).

PECR is located within broad linkage peaks for several alcohol-related traits, including alcoholism (67), comorbid alcoholism and depression (68), level of response to alcohol (69) and amplitude of the P300 response (a neuronal feature related to decision-making) (70, 71). Expression of *PECR* is highest in the liver, followed by the kidney, muscle tissue, the lungs, and the heart but barely, if at all, in the human brain (71), although it was expressed in rodent brain in the present study. If this gene is involved in alcohol dependence, it must act in the periphery more possibly than in the central nervous system. *PECR* is a member of the short-chain dehydrogenase family of enzymes. It is a key enzyme of fatty acid metabolism. It catalyzes the reduction of medium-chain (unsaturated) enoyl-CoAs to saturated acyl-CoAs, which may then undergo oxidation in mitochondria for energy production. This pathway is particularly important in conditions of starvation, when the energy supply is switched from the use of glucose to fatty acids. This pathway originates from triglycerides that are typically increased in the blood of alcohol-dependent individuals (68, 69). Folate deficiency, often observed in alcohol dependent patients (70), may, therefore, interfere with this pathway and prevent the availability of a sufficient supply of acetyl-CoA.

NRD1 on Chr1 was expressed in rodent brain in the present study. It has been hypothesized to play specific roles in neuronal development in mice (72). Recently, Bernstein et al. (73) demonstrated that *NRD1* enzyme was widely, but unevenly, expressed in the human brain and might possibly play a role in brain development too. More recently, it has been shown that *NRD1* is a critical regulator in axonal maturation and myelination, which are essential processes for establishing an efficient neuronal signaling network (74). Involvement of *NRD1* in alcohol dependence is possibly related to this network.

The *HTR7* gene on Chr10 encodes a G-protein-coupled serotonergic receptor. Serotonin (5-HT), one of the most widely used neurotransmitters in the brain, binds to this receptor and produces physiological effects that has been implicated in the etiology of numerous psychiatric diseases, including schizophrenia, depression, obsessive-compulsive, social phobia, and eating disorders (75–77). *HTR7* is distributed in discrete areas of the brain, such as the limbic system and thalamocortical regions, overlapping the areas of neuroelectrical activity of θ EROs, which suggests a possible role of *HTR7* in pathophysiology of psychiatric disorders. The development of 5-HT7 antagonists (78) and the generation of a *Htr7* knockout mouse (79) have revealed a role for the receptor in the control of circadian

rhythms and sleep, as well as in modulating hippocampal neuronal functions such as learning and memory. 5-HT₇ receptors have a high affinity for a number of antidepressants and antipsychotics, such as clozapine and risperidone (80), underscoring their potential role in the neurophysiology of psychiatric disorders. The risk SNP at *HTR7* for alcohol dependence had significant eQTL signal in human lymphoblastoid cell lines and is predicted to highly significantly alter the RNA secondary structure. Interestingly, an antisense LncRNA is located in the promoter of *HTR7*, 450bp from the transcription start site (TSS). This is a LncRNA bidirectional to *HTR7*. A Bidirectional LncRNA transcript usually exhibits a similar expression pattern to its protein-coding counterpart (here, *HTR7*). It can open chromatin to promote the expression of *HTR7* (81). A SNP within this LncRNA was associated with alcohol dependence.

PKNOX2 belongs to the three-amino-acid loop extension (TALE) homeobox family (82). It is a highly conserved nuclear transcription regulator indicated by its structure and sub-cellular localization. *PKNOX2* was expressed in rodent brain and has been identified as one of the *cis*-regulated genes for alcohol addiction in mice (83). *MREG* encodes for elanoregulin, and may function in membrane fusion and regulate the biogenesis of disk membranes of photoreceptor rod cells. In humans, mutations in the *GPDIL* gene have been reported to cause Brugada syndrome 2, which is a progressive disorder of myocardial conduction (84). Interestingly, this gene has received support for the phenotype 'alcohol stimulated activity' in the mouse (85), as it is located in the QTL Etact2 on mouse chromosome 9. *CMTM8* belongs to the chemokine-like factor gene superfamily, a novel family that is similar to the chemokine and the transmembrane 4 superfamilies. It is widely expressed in many tissues, but the exact function of the encoded protein is unknown. *MAP3K9* on Chr14 encodes for a mixed lineage kinase including the serine/threonine kinase that acts as an essential component of the MAP kinase signal transduction pathway. It plays an important role in the cascades of cellular responses evoked by changes in the environment. It acts as an upstream activator of the MKK/JNK signal transduction cascade. The MKK/JNK signaling pathway regulates stress response. *MAP3K9* also plays a role in mitochondrial death signaling pathway, leading to apoptosis. A bidirectional LncRNA is located in the promoter of *MAP3K9*, 777bp from the TSS. *PCNX* on Chr14 encodes a pecanex homolog expressed in the blood plasma. All risk variants at these six genes (i.e., *PKNOX2*, *MREG*, *GPDIL*, *CMTM8*, *MAP3K9* and *PCNX*) are predicted to mildly to highly significantly alter the RNA secondary structure. Finally, *OPA3* is involved in sensory perception, specifically visual perception. *OPA3* is expressed consistently across human brain tissues. All of the five genes (*PKNOX2*, *MREG*, *GPDIL*, *PCNX* and *OPA3*) were expressed in rodent brain in the present study; however, their roles in alcohol dependence remain largely unclear.

Finally, we note that the four variants at *PKNOX2*, the four intergenic variants at *MREG-PECR*, *GPDIL-CMTM8* and *MAP3K9-PCNX*, and the one variant at *OPA3* had no significant eQTL signals in human brain and blood tissues in our analysis, and were not located at any TFBS/CNV/CpG/ESS, suggesting that they might play roles with less priority in risk for alcohol dependence than the preceding *ADH* cluster, *SERINC2*, *KIAA0040*, *NRD1* and *HTR7*. Their biological functions in human remained to be validated.

In summary, from the numerous GWAS findings, we reported the most promising risk genes that were significantly associated with alcohol dependence at genome-wide level across multiple independent samples and were potentially functional. This will significantly contribute to the post-GWAS follow-up studies on alcohol dependence.

Although we believe those non-replicable GWAS findings have lower possibility to be true than the replicable findings, we do not completely exclude the possibility that those non-replicable risk genes might harbor causal variants for alcohol dependence in some specific populations or for some specific endophenotypes of alcohol dependence. The non-replicability might be attributable to the heterogeneity of phenotype or population of different samples. More homogenous samples may be helpful with reducing the false negatives in this non-replicability.

Acknowledgments

This work was supported in part by National Institute on Drug Abuse (NIDA) Grant K01 DA029643, National Institute on Alcohol Abuse and Alcoholism (NIAAA) Grant R21 AA020319, R21 AA021380A and R24 AA013162, the National Alliance for Research on Schizophrenia and Depression (NARSAD) Award 17616 (to LZ), ABMRF/The Foundation for Alcohol Research (LZ) and the Banbury Fund. We thank NIH GWAS Data Repository (dbGaP), the Contributing Investigator(s) (Drs. Bierut, Edenberg, Gelernter and Kranzler) who contributed the phenotype and genotype data from his/her original study, and the funding support by U01 HG004422, U01HG004438, U01 HG004446, U10 AA008401, R01 DA013423, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA011330, R01 AA017535, HHSN268200782096C and HHSN268201100011I. We thank for the Center for Inherited Disease Research (CIDR) and the Genetics of Alcohol Dependence in American Populations (CIDR-Gelernter Study), and the National Center for Biotechnology Information.

References

1. Prescott CA, Kendler KS. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *Am J Psychiatry*. 1999; 156:34–40. [PubMed: 9892295]
2. Walters GD. The heritability of alcohol abuse and dependence: a meta-analysis of behavior genetic research. *Am J Drug Alcohol Abuse*. 2002; 28:557–84. [PubMed: 12211366]
3. Ystrom E, Reichborn-Kjennerud T, Aggen SH, Kendler KS. Alcohol dependence in men: reliability and heritability. *Alcohol Clin Exp Res*. 2011; 35:1716–22. [PubMed: 21676009]
4. Yang BZ, Kranzler HR, Zhao H, Gruen JR, Luo X, Gelernter J. Haplotypic variants in *DRD2*, *ANKK1*, *TTC12*, and *NCAM1* are associated with comorbid alcohol and drug dependence. *Alcohol Clin Exp Res*. 2008; 32:2117–27. [PubMed: 18828801]
5. Chen AC, Manz N, Tang Y, Rangaswamy M, Almasy L, Kuperman S, et al. Single-nucleotide polymorphisms in corticotropin releasing hormone receptor 1 gene (*CRHR1*) are associated with quantitative trait of event-related potential and alcohol dependence. *Alcohol Clin Exp Res*. 2010; 34:988–96. [PubMed: 20374216]
6. Li D, Zhao H, Gelernter J. Strong association of the alcohol dehydrogenase 1B gene (*ADH1B*) with alcohol dependence and alcohol-induced medical diseases. *Biol Psychiatry*. 2011; 70:504–12. [PubMed: 21497796]
7. Li D, Zhao H, Gelernter J. Further clarification of the contribution of the *ADH1C* gene to vulnerability of alcoholism and selected liver diseases. *Hum Genet*. 2012; 131:1361–74. [PubMed: 22476623]
8. Li D, Zhao H, Gelernter J. Strong protective effect of the aldehyde dehydrogenase gene (*ALDH2*) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians. *Hum Genet*. 2012; 131:725–37. [PubMed: 22102315]
9. Luo X, Kranzler HR, Zhao H, Gelernter J. Haplotypes at the *OPRM1* locus are associated with susceptibility to substance dependence in European-Americans. *Am J Med Genet B Neuropsychiatr Genet*. 2003; 120B:97–108. [PubMed: 12815747]

10. Zuo L, Wang K, Zhang X, J.H. K, Li CR, Zhang F, et al. NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk region specific for alcohol dependence in subjects of European descent. *Drug Alcohol Depend.* 2013; 129:254–64. [PubMed: 23455491]
11. Wang KS, Liu X, Zhang Q, Pan Y, Aragam N, Zeng M. A meta-analysis of two genome-wide association studies identifies 3 new loci for alcohol dependence. *J Psychiatr Res.* 2011; 45:1419–25. [PubMed: 21703634]
12. Zlojutro M, Manz N, Rangaswamy M, Xuei X, Flury-Wetherill L, Koller D, et al. Genome-wide association study of theta band event-related oscillations identifies serotonin receptor gene HTR7 influencing risk of alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet.* 2011; 156B: 44–58. [PubMed: 21184583]
13. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26:2190–1. [PubMed: 20616382]
14. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* 2005; 1:e78. [PubMed: 16362079]
15. Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.* 2008; 6:e1. [PubMed: 19222302]
16. Bhavé SV, Hornbaker C, Phang TL, Saba L, Lapadat R, Kechris K, et al. The PhenoGen informatics website: tools for analyses of complex traits. *BMC Genet.* 2007; 8:59. [PubMed: 17760997]
17. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013; 14:R36. [PubMed: 23618408]
18. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28:511–5. [PubMed: 20436464]
19. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013; Chapter 7(Unit7):20. [PubMed: 23315928]
20. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31:3406–15. [PubMed: 12824337]
21. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. fourth edition ed.. American Psychiatric Press; Washington, DC: 1994.
22. Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, et al. A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A.* 2010; 107:5082–7. [PubMed: 20202923]
23. Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, et al. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol Clin Exp Res.* 2010; 34:840–52. [PubMed: 20201924]
24. Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, et al. A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biological Psychiatry.* 2011; 70:513–8. [PubMed: 21529783]
25. Zuo L, Gelernter J, Zhang CK, Zhao H, Lu L, Kranzler HR, et al. Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. *Neuropsychopharmacology.* 2012; 37:557–66. [PubMed: 21956439]
26. Zuo L, Luo X, Listman JB, Kranzler HR, Wang S, Anton RF, et al. Population admixture modulates risk for alcohol dependence. *Hum Genet.* 2009; 125:605–13. [PubMed: 19306106]
27. Zuo L, Wang F, Zhang XY, Li CSR, Lu L, Ye L, et al. Genome-wide significant association signals in IPO11-HTR1A region specific for alcohol and nicotine co-dependence. *Alcohol Clin Exp Res.* 2013; 37:730–9. [PubMed: 23216389]
28. Zuo L, Zhang CK, Wang F, Li CS, Zhao H, Lu L, et al. A novel, functional and replicable risk gene region for alcohol dependence identified by genome-wide association study. *PLoS One.* 2011; 6:e26726. [PubMed: 22096494]

29. Zuo L, Zhang H, Malison RT, Li CSR, Zhang XY, Wang F, et al. Rare ADH variant constellations are specific for alcohol dependence. *Alcohol Alcohol*. 2013; 48:9–14. [PubMed: 23019235]
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–75. [PubMed: 17701901]
31. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–9. [PubMed: 16862161]
32. Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, et al. Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry*. 2009; 66:773–84. [PubMed: 19581569]
33. Frank J, Cichon S, Treutlein J, Ridinger M, Mattheisen M, Hoffmann P, et al. Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict Biol*. 2012; 17:171–80. [PubMed: 22004471]
34. Park BL, Kim JW, Cheong HS, Kim LH, Lee BC, Seo CH, et al. Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Hum Genet*. 2013; 132:657–68. [PubMed: 23456092]
35. Biernacka JM, Geske JR, Schneekloth TD, Frye MA, Cunningham JM, Choi DS, et al. Replication of genome wide association studies of alcohol dependence: support for association with variation in ADH1C. *PLoS One*. 2013; 8:e58798. [PubMed: 23516558]
36. Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, et al. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry*. 2014; 19:41–9. [PubMed: 24166409]
37. Chen X, Cho K, Singer BH, Zhang H. The nuclear transcription factor PKNOX2 is a candidate gene for substance dependence in European-origin women. *PLoS One*. 2011; 6:e16002. [PubMed: 21298047]
38. Biernacka JM, Geske J, Jenkins GD, Colby C, Rider DN, Karpyak VM, et al. Genome-wide gene-set analysis for identification of pathways associated with alcohol dependence. *Int J Neuropsychopharmacol*. 2013; 16:271–8. [PubMed: 22717047]
39. Edwards AC, Aliev F, Bierut LJ, Bucholz KK, Edenberg H, Hesselbrock V, et al. Genome-wide association study of comorbid depressive syndrome and alcohol dependence. *Psychiatr Genet*. 2012; 22:31–41. [PubMed: 22064162]
40. Lu L, Risch E, Deng Q, Biglia N, Picardo E, Katsaros D, et al. An insulin-like growth factor-II intronic variant affects local DNA conformation and ovarian cancer survival. *Carcinogenesis*. 2013; 34:2024–30. [PubMed: 23677070]
41. Lu L, Zhu G, Zhang C, Deng Q, Katsaros D, Mayne ST, et al. Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer. *Breast Cancer Res Treat*. 2012; 136:875–83. [PubMed: 23124417]
42. Lu L, Katsaros D, Mayne ST, Risch HA, Benedetto C, Canuto EM, et al. Functional study of risk loci of stem cell-associated gene lin-28B and associations with disease survival outcomes in epithelial ovarian cancer. *Carcinogenesis*. 2012; 33:2119–25. [PubMed: 22822098]
43. Luo X, Kranzler HR, Zuo L, Wang S, Schork NJ, Gelernter J. Diplotype trend regression analysis of the ADH gene cluster and the ALDH2 gene: multiple significant associations with alcohol dependence. *Am J Hum Genet*. 2006; 78:973–87. [PubMed: 16685648]
44. Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. *Science*. 2008; 319:1787–9. [PubMed: 18369136]
45. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009; 10:155–9. [PubMed: 19188922]
46. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*. 2008; 454:126–30. [PubMed: 18509338]
47. Sartor GC, St Laurent G 3rd, Wahlestedt C. The Emerging Role of Non-Coding RNAs in Drug Addiction. *Front Genet*. 2012; 3:106. [PubMed: 22737160]

48. Bu Q, Hu Z, Chen F, Zhu R, Deng Y, Shao X, et al. Transcriptome analysis of long non-coding RNAs of the nucleus accumbens in cocaine-conditioned mice. *J Neurochem.* 2012; 123:790–9. [PubMed: 22957495]
49. Qureshi IA, Mattick JS, Mehler MF. Long non-coding RNAs in nervous system function and disease. *Brain Res.* 2010; 1338:20–35. [PubMed: 20380817]
50. Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, et al. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* 2010; 29:3082–93. [PubMed: 20729808]
51. Kryger R, Fan L, Wilce PA, Jaquet V. MALAT-1, a non protein-coding RNA is upregulated in the cerebellum, hippocampus and brain stem of human alcoholics. *Alcohol.* 2012; 46:629–34. [PubMed: 22560368]
52. Michelhaugh SK, Lipovich L, Blythe J, Jia H, Kapatos G, Bannon MJ. Mining Affymetrix microarray data for long non-coding RNAs: altered expression in the nucleus accumbens of heroin abusers. *J Neurochem.* 2011; 116:459–66. [PubMed: 21128942]
53. Kaplan R, Luettich K, Heguy A, Hackett NR, Harvey BG, Crystal RG. Monoallelic up-regulation of the imprinted H19 gene in airway epithelium of phenotypically normal cigarette smokers. *Cancer Res.* 2003; 63:1475–82. [PubMed: 12670893]
54. Ouko LA, Shantikumar K, Knezovich J, Haycock P, Schnugh DJ, Ramsay M. Effect of alcohol consumption on CpG methylation in the differentially methylated regions of H19 and IG-DMR in male gametes: implications for fetal alcohol spectrum disorders. *Alcohol Clin Exp Res.* 2009; 33:1615–27. [PubMed: 19519716]
55. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. The antisense transcriptomes of human cells. *Science.* 2008; 322:1855–7. [PubMed: 19056939]
56. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A.* 2009; 106:11667–72. [PubMed: 19571010]
57. Galter D, Carmine A, Buervenich S, Duester G, Olson L. Distribution of class I, III and IV alcohol dehydrogenase mRNAs in the adult rat, mouse and human brain. *Eur J Biochem.* 2003; 270:1316–26. [PubMed: 12631290]
58. Pares X, Farres J, Vallee BL. Organ specific alcohol metabolism: placental chi-ADH. *Biochem Biophys Res Commun.* 1984; 119:1047–55. [PubMed: 6370249]
59. Haseba T, Ohno Y. A new view of alcohol metabolism and alcoholism--role of the high-Km Class III alcohol dehydrogenase (ADH3). *Int J Environ Res Public Health.* 2010; 7:1076–92. [PubMed: 20617019]
60. Haseba T, Kameyama K, Mashimo K, Ohno Y. Dose-Dependent Change in Elimination Kinetics of Ethanol due to Shift of Dominant Metabolizing Enzyme from ADH 1 (Class I) to ADH 3 (Class III) in Mouse. *Int J Hepatol.* 2012; 2012:408190. [PubMed: 22164338]
61. Israel Y, Rivera-Meza M, Karahanian E, Quintanilla ME, Tampier L, Morales P, et al. Gene specific modifications unravel ethanol and acetaldehyde actions. *Front Behav Neurosci.* 2013; 7:80. [PubMed: 23847486]
62. Grossman TR, Luque JM, Nelson N. Identification of a ubiquitous family of membrane proteins and their expression in mouse brain. *J Exp Biol.* 2000; 203:447–57. [PubMed: 10637174]
63. Inuzuka M, Hayakawa M, Ingi T. Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *J Biol Chem.* 2005; 280:35776–83. [PubMed: 16120614]
64. Krystal JH, Petrakis IL, Mason G, Trevisan L, D'Souza DC. N-methyl-D-aspartate glutamate receptors and alcoholism: reward, dependence, treatment, and vulnerability. *Pharmacol Ther.* 2003; 99:79–94. [PubMed: 12804700]
65. Gratacos M, Costas J, de Cid R, Bayes M, Gonzalez JR, Baca-Garcia E, et al. Identification of new putative susceptibility genes for several psychiatric disorders by association analysis of regulatory and non-synonymous SNPs of 306 genes involved in neurotransmission and neurodevelopment. *Am J Med Genet B Neuropsychiatr Genet.* 2009; 150B:808–16. [PubMed: 19086053]
66. Hill SY, Weeks DE, Jones BL, Zezza N, Stiffler S. ASTN1 and alcohol dependence: family-based association analysis in multiplex alcohol dependence families. *Am J Med Genet B Neuropsychiatr Genet.* 2012; 159B:445–55. [PubMed: 22488871]

67. Hill SY, Shen S, Zezza N, Hoffman EK, Perlin M, Allan W. A genome wide search for alcoholism susceptibility genes. *Am J Med Genet B Neuropsychiatr Genet.* 2004; 128B:102–13. [PubMed: 15211641]
68. Nurnberger JI Jr, Foroud T, Flury L, Su J, Meyer ET, Hu K, et al. Evidence for a locus on chromosome 1 that influences vulnerability to alcoholism and affective disorder. *Am J Psychiatry.* 2001; 158:718–24. [PubMed: 11329392]
69. Schuckit MA, Edenberg HJ, Kalmijn J, Flury L, Smith TL, Reich T, et al. A genome-wide search for genes that relate to a low level of response to alcohol. *Alcohol Clin Exp Res.* 2001; 25:323–9. [PubMed: 11290841]
70. Begleiter H, Porjesz B, Reich T, Edenberg HJ, Goate A, Blangero J, et al. Quantitative trait loci analysis of human event-related brain potentials: P3 voltage. *Electroencephalogr Clin Neurophysiol.* 1998; 108:244–50. [PubMed: 9607513]
71. Porjesz B, Begleiter H, Wang K, Almasy L, Chorlian DB, Stimus AT, et al. Linkage and linkage disequilibrium mapping of ERP and EEG phenotypes. *Biol Psychol.* 2002; 61:229–48. [PubMed: 12385677]
72. Fumagalli P, Accarino M, Egeo A, Scartezzini P, Rappazzo G, Pizzuti A, et al. Human NRD convertase: a highly conserved metalloendopeptidase expressed at specific sites during development and in adult tissues. *Genomics.* 1998; 47:238–45. [PubMed: 9479496]
73. Bernstein HG, Stricker R, Dobrowolny H, Trubner K, Bogerts B, Reiser G. Histochemical evidence for wide expression of the metalloendopeptidase nardilysin in human brain neurons. *Neuroscience.* 2007; 146:1513–23. [PubMed: 17442499]
74. Ohno M, Hiraoka Y, Matsuoka T, Tomimoto H, Takao K, Miyakawa T, et al. Nardilysin regulates axonal maturation and myelination in the central and peripheral nervous system. *Nat Neurosci.* 2009; 12:1506–13. [PubMed: 19935654]
75. Naughton M, Mulrooney JB, Leonard BE. A review of the role of serotonin receptors in psychiatric disorders. *Hum Psychopharmacol.* 2000; 15:397–415. [PubMed: 12404302]
76. Hoyer D, Hannon JP, Martin GR. Molecular, pharmacological and functional diversity of 5-HT receptors. *Pharmacol Biochem Behav.* 2002; 71:533–54. [PubMed: 11888546]
77. Nichols DE, Nichols CD. Serotonin receptors. *Chem Rev.* 2008; 108:1614–41. [PubMed: 18476671]
78. Thomas DR, Melotto S, Massagrande M, Gribble AD, Jeffrey P, Stevens AJ, et al. SB-656104-A, a novel selective 5-HT₇ receptor antagonist, modulates REM sleep in rats. *Br J Pharmacol.* 2003; 139:705–14. [PubMed: 12812993]
79. Hedlund PB, Danielson PE, Thomas EA, Slanina K, Carson MJ, Sutcliffe JG. No hypothermic response to serotonin in 5-HT₇ receptor knockout mice. *Proc Natl Acad Sci U S A.* 2003; 100:1375–80. [PubMed: 12529502]
80. Roth BL, Craigo SC, Choudhary MS, Uluer A, Monsma FJ Jr, Shen Y, et al. Binding of typical and atypical antipsychotic agents to 5-hydroxytryptamine-6 and 5-hydroxytryptamine-7 receptors. *J Pharmacol Exp Ther.* 1994; 268:1403–10. [PubMed: 7908055]
81. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A.* 2008; 105:716–21. [PubMed: 18184812]
82. Imoto I, Sonoda I, Yuki Y, Inazawa J. Identification and characterization of human PKNOX2, a novel homeobox-containing gene. *Biochem Biophys Res Commun.* 2001; 287:270–6. [PubMed: 11549286]
83. Mulligan MK, Ponomarev I, Hitzemann RJ, Belknap JK, Tabakoff B, Harris RA, et al. Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proc Natl Acad Sci U S A.* 2006; 103:6368–73. [PubMed: 16618939]
84. Makiyama T, Akao M, Haruna Y, Tsuji K, Doi T, Ohno S, et al. Mutation analysis of the glycerol-3 phosphate dehydrogenase-1 like (GPD1L) gene in Japanese patients with Brugada syndrome. *Circ J.* 2008; 72:1705–6. [PubMed: 18762705]
85. Palmer AA, Lessov-Schlaggar CN, Ponder CA, McKinnon CS, Phillips TJ. Sensitivity to the locomotor-stimulant effects of ethanol and allopregnanolone: a quantitative trait locus study of common genetic influence. *Genes Brain Behav.* 2006; 5:506–17. [PubMed: 17010097]

Table 1

Significant and/or replicable GWAS findings for alcohol dependence

SNP	Gene	Chr	Location	Independent Sample 1			Independent Sample 2			Meta/combined		
				Race	P	Ref.	Race	P	Ref.	n	Z/OR	p-value
rs1789891	ADH cluster	4	Intergenic	German male	1.3×10^{-8}	(33)	EA	1.5×10^{-2}	(33)	2234	6.0	2.3×10^{-9}
rs2851300	ADH cluster	4	Intergenic	German male	1.0×10^{-7}	(33)	Korean	1.5×10^{-3}	(34)			
rs469748	ADH cluster	4	Intergenic	Asian	9.9×10^{-10}	(34)				2008	8.3	1.5×10^{-16}
				German male	2.5×10^{-7}	(33)	EA	4.7×10^{-2}	(33)			
rs1372679	ADH cluster	4	Intergenic	Korean	2.4×10^{-3}	(34)	Asian	1.7×10^{-7}	(34)	2855	7.5	6.9×10^{-14}
				German male	1.6×10^{-6}	(33)	EA	3.5×10^{-2}	(33)			
				Korean	3.4×10^{-3}	(34)	Asian	2.2×10^{-7}	(34)	2855	7.3	3.8×10^{-13}
rs28470942	ADH cluster	4	Intergenic	AA	3.4×10^{-8}	(36)	AA	1.3×10^{-4}	(22)	2415	6.7	1.9×10^{-11}
rs28864441	ADH cluster	4	Intergenic	AA	6.6×10^{-9}	(36)	AA	3.2×10^{-5}	(22)	2415	7.1	1.1×10^{-12}
rs2066702	ADH cluster	4	Exonic	AA	2.2×10^{-9}	(36)	AA	2.0×10^{-5}	(22)	2415	7.3	2.2×10^{-13}
rs1229984	ADH cluster	4	Exonic	EA	6.8×10^{-14}	(36)	EA	8.1×10^{-7}	(22)			
				Asian	2.6×10^{-21}	(34)				3173	11.9	1.4×10^{-32}
rs2173201	ADH cluster	4	Intergenic	AA	4.5×10^{-7}	(36)	AA	1.4×10^{-3}	(22)			
				Korean	1.5×10^{-5}	(34)				2532	6.8	1.4×10^{-11}
rs1614972	ADH cluster	4	Intronic	German male	2.8×10^{-4}	(32)	German male	3.6×10^{-2}	(32)			
				EA	1.7×10^{-3}	(35)	Korean	1.0×10^{-5}	(34)			
				AA	1.8×10^{-7}	(36)	AA	3.2×10^{-2}	(22)	4124	7.8	7.2×10^{-15}
rs4147541	ADH cluster	4	Intergenic	AA	1.5×10^{-6}	(36)	AA	2.0×10^{-3}	(22)			
				Korean	1.5×10^{-5}	(34)				2532	6.5	7.3×10^{-11}
rs1372680	ADH cluster	4	Intergenic	German male	2.7×10^{-6}	(33)	EA	6.8×10^{-2}	(33)	2234	4.8	1.4×10^{-6}
20 SNPs	ADH cluster	4		Korean	6.3×10^{-8}		Asian	2.0×10^{-16}		621	9.8	1.7×10^{-22}
					-1.5×10^{-3}	(34)		-3.9×10^{-8}	(34)		-6.5	-6.1×10^{-11}
12 SNPs	ADH cluster	4		Korean	5.0×10^{-5}		Asian	7.0×10^{-8}		621	6.3	3.7×10^{-10}
					-1.2×10^{-3}	(34)		-3.4×10^{-6}	(34)		-5.9	-2.8×10^{-9}
20 SNPs	ADH cluster	4		AA	1.4×10^{-7}		AA	4.8×10^{-5}		2415	6.7	3.6×10^{-11}
					-3.3×10^{-6}	(36)		-2.3×10^{-2}	(22)		-5.4	-3.4×10^{-8}

SNP	Gene	Chr	Location	Independent Sample 1			Independent Sample 2			Meta/combined		
				Race	P	Ref.	Race	P	Ref.	n	Z/OR	p-value
rs4478858	SERINC2	1	Intronic	EA	4.4×10 ⁻⁷	(10)	Australian ^F	2.1×10 ⁻²	(10)	3054	5.5	3.1×10 ⁻⁸
rs2275436	SERINC2	1	Intronic	EA	2.4×10 ⁻⁷	(10)	Australian ^F	4.3×10 ⁻²	(10)	3054	5.5	3.8×10 ⁻⁸
rs1039630	(SERINC2)	1	Intergenic	EA	2.6×10 ⁻⁷	(10)	Australian ^F	0.049	(10)	3054	5.5	4.7×10 ⁻⁸
82 SNPs	(SERINC2)	1	Intergenic	EA	2.0×10 ⁻⁶		Australian ^F	<0.05	(10)	3054	5.5	2.8×10 ⁻⁷
					-7.9×10 ⁻⁴	(10)					-3.9	-9.8×10 ⁻⁵
rs6701037	(KIAA0040)	1	Intergenic	EA	3.6×10 ⁻⁵	(11)	EA	1.4×10 ⁻³	(11)	2069	5.8	6.7×10 ⁻⁹
rs6425323	(KIAA0040)	1	Intergenic	EA	1.8×10 ⁻⁶	(22)	EA	1.3×10 ⁻⁶	(25)			
				EA	5.1×10 ⁻⁵	(11)	EA	1.7×10 ⁻³	(11)	2069	5.7	1.2×10 ⁻⁸
rs1057302	KIAA0040	1	3' UTR	EA	2.3×10 ⁻⁶	(22)	EA	1.9×10 ⁻⁶	(25)			
				EA	7.9×10 ⁻⁵	(11)	EA	1.8×10 ⁻³	(11)	2069	5.6	1.6×10 ⁻⁸
rs1057239	KIAA0040	1	3' UTR	EA	1.9×10 ⁻⁶	(22)	EA	2.5×10 ⁻⁶	(25)			
rs1894709	KIAA0040	1	Intronic	EA	6.9×10 ⁻⁵	(11)	EA	8.8×10 ⁻³	(11)	2069	5.7	1.1×10 ⁻⁸
				EA	2.0×10 ⁻⁶	(22)	EA	2.8×10 ⁻⁷	(25)			
rs1344694	(MREG-PECR)	2	Intergenic	EA	7.0×10 ⁻⁵	(11)	EA	9.7×10 ⁻³	(11)	2069	5.6	1.6×10 ⁻⁸
				EA	2.8×10 ⁻⁶	(22)	EA	3.9×10 ⁻⁷	(25)			
rs7590720	(MREG-PECR)	2	Intergenic	German male	5.6×10 ⁻⁶	(32)	German male	3.4×10 ⁻³	(32)	1511	1.3	1.7×10 ⁻⁸
				German male	4.6×10 ⁻⁶	(33)						
rs10893366	PKNOX2	11	Intronic	German male	5.7×10 ⁻⁶	(32)	German male	6.7×10 ⁻⁴	(32)	1511	1.4	9.7×10 ⁻⁹
				EA	3.4×10 ⁻⁵	(22)	AA	1.2×10 ⁻³	(22)			
rs750338	PKNOX2	11	Intronic	EA female	6.9×10 ⁻⁷	(37)	Australian ^F	1.2×10 ⁻⁵	(11)	3547	6.8	8.5×10 ⁻¹²
				EA	6.5×10 ⁻⁶	(11)	EA	5.0×10 ⁻²	(11)			
rs10893365	PKNOX2	11	Intronic	EA	2.4×10 ⁻⁶	(22)	Australian ^F	1.4×10 ⁻²	(11)	3719	5.4	8.1×10 ⁻⁸
				EA female	4.2×10 ⁻⁷	(37)						
rs12284594	PKNOX2	11	Intronic	EA	4.0×10 ⁻⁵	(11)	AA	2.0×10 ⁻²	(22)			
				EA	2.5×10 ⁻⁵	(22)	EA female	2.3×10 ⁻⁷	(37)	3829	1.3	7.2×10 ⁻⁶
	PKNOX2	11	Intronic	EA	2.4×10 ⁻⁵	(11)	EA	8.1×10 ⁻⁶	(22)	2668	1.4	8.1×10 ⁻⁶
				EA female	7.1×10 ⁻⁸	(37)						

SNP	Gene	Chr	Location	Independent Sample 1			Independent Sample 2			Meta/combined		
				Race	P	Ref.	Race	P	Ref.	n	Z/OR	p-value
rs2842576	<i>NRD1</i>	1	<i>Intronic</i>	<u>E</u> A	3.4×10^{-5}	(11)	<u>E</u> A	4.9×10^{-2}	(11)	1895	<u>4.8</u>	1.5×10^{-6}
				EA	6.1×10^{-6}	(22)						
rs9825310	<i>(GPD1L-CMTM8)</i>	3	<i>Intergenic</i>	<u>German male</u>	7.6×10^{-6}	(33)	<u>E</u> A	3.0×10^{-3}	(33)	2234	5.4	8.6×10^{-8}
rs36563	<i>(MAP3K9-PCNX)</i>	14	<i>Intergenic</i>	German male	6.2×10^{-6}	(32)	German male	3.8×10^{-2}	(32)	1511	1.3	4.6×10^{-6}
rs7916403	<i>HTR7</i>	10	<i>Intronic</i>	<u>E</u> A+ <u>A</u> A	8.0×10^{-3}	(12)	<u>E</u> A+ <u>A</u> A ^F	4.2×10^{-2}	(12)	3452	4.5	6.3×10^{-6}
				<u>E</u> A	1.8×10^{-3}	(38)						
rs8111589	<i>OPA3</i>	19	<i>Intronic</i>	<u>E</u> A	1.5×10^{-4}	(23)	<u>E</u> A	2.8×10^{-2}	(23)	1978	<u>4.9</u>	1.1×10^{-6}
				<u>A</u> A	1.6×10^{-2}	(23)	E <u>A</u> ♣	4.0×10^{-6}	(39)			

Significant associations with $p<5\times10^{-5}$ are bold. Replicable associations: $p<10^{-5}$ in one sample and $p<0.05$ in more than 3 samples. The samples included in meta-analysis are underlined. The p values not underlined correspond to the pooled samples reported in the original literatures. The genes in parentheses denote the proximate genes. Except for alcohol dependence comorbid major depression (♣), all phenotypes for associations are alcohol dependence. EA, European-American; AA, African-American; n, sample size of cases.

^FThese studies are family-based, otherwise, case-control designed.

Table 2

The potential functions of risk variants

SNP	Gene	Structure	Bioinfo	Alteration of 2 nd RNA						cis-eQTL					
				Duke samples		HapMap samples				Duke samples		HapMap samples			
				Brain	PBMC	CEU children	CEU parent	JPT	YRI children	YRI parent	Brain	PBMC	CEU children	CEU parent	JPT
rs1789891	<i>ADH cluster</i>	Mild change	-	-	-	-	9.7×10 ⁻⁵	-	-	-	-	-	-	-	-
rs2851300	<i>ADH cluster</i>	No change	-	-	-	-	0.024	-	-	-	-	-	-	-	-
rs4699748	<i>ADH cluster</i>	Mild change	-	-	-	-	5.5×10 ⁻⁵	-	-	-	-	-	-	-	-
rs1372679	<i>ADH cluster</i>	No change	-	-	-	-	2.7×10 ⁻⁶	-	-	-	-	-	-	-	-
rs28470942	<i>ADH cluster</i>	No change	-	-	-	-	-	-	-	-	-	-	-	-	-
rs28864441	<i>ADH cluster</i>	Highly Sig.	-	-	-	-	-	-	-	-	-	-	-	-	-
rs2066702	<i>ADH cluster</i>	No change	damaging	-	-	-	-	-	-	-	-	-	-	-	-
rs1229984	<i>ADH cluster</i>	No change	ESS	-	-	-	-	-	-	-	-	-	-	-	-
rs2173201	<i>ADH cluster</i>	Significant	-	0.044	-	-	-	-	-	-	-	-	-	-	-
rs1614972	<i>ADH cluster</i>	No change	-	-	-	-	0.012	-	-	-	-	-	-	-	-
rs4147541	<i>ADH cluster</i>	Significant	TFBS	-	-	-	-	-	-	-	-	-	-	-	-
rs1372680	<i>ADH cluster</i>	No change	-	-	-	-	2.7×10 ⁻⁶	-	-	-	-	-	-	-	-
rs4478858	<i>SERINC2</i>	Mild change	TFBS/CNV	0.024	0.026	-	-	-	-	-	-	-	-	-	-
rs2275436	<i>SERINC2</i>	Highly Sig.	CpG/CNV	-	-	-	-	-	-	-	-	-	-	-	-
rs1039630	<i>SERINC2</i>	No change	CNV	-	-	-	-	-	-	-	-	-	-	-	-
82 SNPs (<i>SERINC2</i>)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
rs6701037	<i>KIAA0040</i>	Mild change	-	-	-	0.027	0.007	0.025	0.037	6.0×10 ⁻⁴	-	-	-	-	-
rs6425323	<i>KIAA0040</i>	No change	CpG	-	-	0.010	8.9×10 ⁻³	0.025	0.038	5.0×10 ⁻⁴	-	-	-	-	-
rs1057302	<i>KIAA0040</i>	No change	-	-	-	0.013	4.8×10 ⁻³	0.025	-	2.0×10 ⁻⁴	-	-	-	-	-
rs1057239	<i>KIAA0040</i>	No change	TFBS	-	-	-	0.028	-	0.011	-	-	-	-	-	-
rs1894709	<i>KIAA0040</i>	Mild change	TFBS	-	-	-	0.008	-	0.016	-	-	-	-	-	-
rs1344694	(<i>MREG-PECR</i>)	Mild change	-	-	-	-	-	-	-	-	-	-	-	-	-
rs7590720	(<i>MREG-PECR</i>)	No change	-	-	-	-	-	-	-	-	-	-	-	-	-
rs10893366	<i>PKNOX2</i>	No change	-	-	-	-	-	-	-	-	-	-	-	-	-
rs750338	<i>PKNOX2</i>	Significant	-	-	-	-	-	-	-	-	-	-	-	-	-
rs10893365	<i>PKNOX2</i>	No change	-	-	-	-	-	-	-	-	-	-	-	-	-

SNP	Alteration of 2 nd RNA		cis-eQTL							
			Duke samples			HapMap samples				
			Brain	PBMC	Bioinfo	CEU children	CEU parent	JPT	YRI children	YRI parent
rs12284594	<i>PKNOX2</i>	No change	-	-	-	-	-	-	-	-
rs2842576	<i>NRD1</i>	Highly Sig.	-	-	-	-	-	-	-	0.028
rs9825310	<i>(GPD1L-CMTM8)</i>	Highly Sig.	-	-	-	-	-	-	-	-
rs36563	<i>(MAP3K9-PCNX)</i>	Mild change	-	-	-	-	-	-	-	-
rs7916403	<i>HTR7</i>	Highly Sig.	-	-	-	0.017	9.7×10 ⁻³	-	-	-
rs8111589	<i>OPA3</i>	Mild change	-	-	-	-	-	-	-	-

cis-eQTL, the cis-acting regulatory effects of the markers on the mRNA transcript expression of proximate genes. TFBS, these SNPs are located in the transcription factor binding sites; ESS, this marker is located in an exonic splicing silencer or enhance; CpG, these SNPs are located within methylated CpG islands; CNV, these SNPs are located within CNV s; damaging, this SNP is predicted to be with high confidence supposed to affect protein function or structure (“probably damaging”) by Polyphen. The alteration of secondary RNA structure is predicted using MFOLD.

Table 3

Genes of interest expression levels in rodent brain

Gene Name	Gene Symbol	Rat Brain Total RNAseq
		Evidence of Expression (FPKM)
alcohol dehydrogenase 1 *	<i>ADH1</i> *	Not Expressed
alcohol dehydrogenase 4 (class II), pi polypeptide	<i>ADH4</i>	Not Expressed
alcohol dehydrogenase 5 (class III), chi polypeptide	<i>ADH5</i>	Present (1.3)
alcohol dehydrogenase 6 (class V)	<i>ADH6</i>	Not Expressed
alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	<i>ADH7</i>	Not Expressed
serine incorporator 2 gene	<i>SERINC2</i>	Present (14.9)
Uncharacterized Protein KIAA0040	<i>KIAA0040</i> **	Not Expressed
melanoregulin	<i>MREG</i>	Present (1.4)
peroxisomal trans-2-enoyl-CoA reductase	<i>PECR</i>	Present (1.8)
Pbx/knotted 1 homeobox 2	<i>PKN0X2</i>	Present (1.3)
nardilysin 1	<i>NRD1</i>	Present (4.8)
glycerol-3-phosphate dehydrogenase 1-like	<i>GPD1L</i>	Present (2.7)
CKLF-like MARVEL transmembrane domain containing 8	<i>CMTM8</i>	Not Expressed
mitogen-activated protein kinase kinase kinase 9	<i>MAP3K9</i>	Present (3.2)
pecanex homolog (Drosophila)	<i>PCNX</i>	Present (6.6)
5-hydroxytryptamine (serotonin) receptor 7, adenylate cyclase-coupled	<i>HTR7</i>	Present (1.2)
optic atrophy 3	<i>OPA3</i>	Present (3.0)

FPKM values were calculated for high confidence genes identified in de novo genome-guided transcriptome reconstruction. High confidence genes had an FPKM value greater than 1 and were longer than 350 nucleotides and noted as "Present". "Not expressed" signifies no high confidence gene was identified in the transcriptome reconstruction.

* Rodents do not have multiple subunits for ADH1, therefore we are comparing ADH1 in rats to ADH1a, ADH1b and ADH1c in humans.

** KIAA0040 is not thoroughly annotated in the human genome and is not distinguishable among the RNAs expressed in rodent brain.

Table 4

Long non-coding RNAs (LncRNAs) proximate to the risk genes

LncRNA name (NCBI)	Alias	Full name	Length (nt)	Distance to risk gene	Category
NR_037884.1	LOC100507053	-	212,506	Covering <i>ADHs</i> 5, 4, 6 and 1A; 5,014bp to <i>ADH1B</i>	antisense LncRNA
NR_028270.1	PCNAP1	Proliferating cell nuclear antigen pseudogene 1	1,055	16,301 bp to <i>ADH4</i> and 40,991 bp to <i>ADH6</i>	intergenic LincRNA
TCONS_00001448	-	-	11,184	25,641 bp to <i>SERINC2</i>	intergenic LincRNA
NR_034112.1	LOC149086	Ornithine decarboxylase 1 pseudogene	2,332	64,309bp to <i>SERINC2</i>	intergenic LincRNA
NR_027085.1	LOC284551/TCONS_00000005	-	5,811	76,509bp to <i>SERINC2</i>	intergenic LincRNA
TCONS_00000180	-	-	1,227	80,437bp to <i>SERINC2</i>	intergenic LincRNA
NR_015350.2	KIAA0040	KIAA0040	36,107	8,921 bp to <i>TNN</i>	intergenic LincRNA
TCONS_00000690	-	-	1,622	10,135bp to <i>KIAA0040</i>	intergenic LincRNA
NR_037195.1	LINC00607	-	231,974	99,055bp to <i>MREG</i>	intergenic LincRNA
NR_037701.1	PKI55/TCONS_00021483	DKFZp434H1419	3,304	135,073p to <i>PECR</i>	intergenic LincRNA
TCONS_00019792	-	-	1,308	157,758bp to <i>PKN0X2</i>	intergenic LincRNA
TCONS_12_00001233	-	-	4,428	220,405bp to <i>NRD1</i>	intergenic LincRNA
XR_110219.3	LOC100506411	-	6,107	777bp to <i>MAP3K9</i>	bidirectional LncRNA
TCONS_00022778	-	-	1,144	65,580bp to <i>PCNX</i>	intergenic LncRNA
TCONS_00018282	-	-	3,295	456bp to <i>HTR7</i>	bidirectional LncRNA

Intergenic, located between two protein-coding genes and at least 1 kb away from these genes; Sense, LncRNAs are transcribed from the same genomic strand as the protein-coding mRNAs; Antisense, LncRNAs are transcribed from the antisense strand; Bidirectional, LncRNAs are oriented head to head with a protein-coding gene within 1kb.

Table 5

Significant risk SNPs within LncRNAs for alcohol dependence

LncRNA	Chr	Position (Build 36.3)	European-Americans (SAGE+COGA)		African-Americans (SAGE+COGA)	
			SNPs within LncRNA	P-value	SNPs within LncRNA	P-value
NR_037884.1	4	100,229,031–100,441,536	rs17028615, rs6812015	0.016 p 0.030	rs7690613	0.014
TCONS_12_00022340	5	62,633,106–62,633,639	rs2169520		rs13148577, rs1800759, rs4148884, rs2000864, rs4147547, rs4147545, rs3857224, rs2187483, rs3819197, rs1893883, rs1566141 (Yale)	2.6×10 ⁻³ p 0.049
TCONS_00001448	1	31,705,755–31,716,938	rs10798854	4.0×10 ⁻⁶	rs346425 (Yale)	0.021
NR_034112.1	1	31,744,423–31,746,754	rs1316509, rs1574104	1.4×10 ⁻⁴	rs10798854 (Yale)	0.025
NR_027085.1	1	31,756,623–31,762,433	rs10914430, rs2153275, rs6657902	1.9×10 ⁻⁴ p 2.2×10 ⁻³		
TCONS_00000180	1	31,760,551–31,761,777	rs2153275	4.0×10 ⁻³ p 0.038		
TCONS_00000690	1	173,438,987–173,440,608	rs760489, rs1012876, rs1012877	4.0×10 ⁻³		
NR_037195.1	2	216,184,531–216,416,504	rs16854689, rs6716449, rs12469571, rs2888365, rs1403327, rs12473821	0.016 p 0.048	rs11677106, rs7589090, rs6726418, rs2204167, rs10932627, rs4467263	0.017 p 0.045
XR_110219.3	14	70,346,418–70,352,524	rs11623147	3.8×10 ⁻³ p 0.047	rs2678272, rs10199888, rs16854581 (Yale)	6.3×10 ⁻⁴ p 4.6×10 ⁻³
TCONS_00018282	10	92,608,107–92,611,401		0.017	rs1891311 (Yale)	0.032

(Yale), this sample is another independent sample different from SAGE+COGA sample.