Published in final edited form as:

Alcohol Clin Exp Res. 2015 August ; 39(8): 1388–1395. doi:10.1111/acer.12786.

A new genome-wide association meta-analysis of alcohol dependence

Lingjun Zuo¹, Yunlong Tan², Xiangyang Zhang³, Xiaoping Wang⁴, John Krystal¹, Boris Tabakoff⁵, Chunlong Zhong^{6,*}, and Xingguang Luo^{1,2,*}

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

²Biological Psychiatry Research Center, Beijing Huilongguan Hospital, Beijing, China

³Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas, USA

⁴Department of Neurology, First People's Hospital, Shanghai Jiaotong University, Shanghai, China

⁵Department of Pharmacology, University of Colorado School of Medicine, Aurora, CO, USA

⁶Department of Neurosurgery, Dongfang Hospital, Shanghai Jiaotong University, Shanghai, China

Abstract

Background—Conventional meta-analysis based on genetic markers may be less powerful for heterogeneous samples. In the present study, we introduced a new meta-analysis for four genome-wide association studies on alcohol dependence that integrated the information of putative causal variants.

Methods—A total of 12,481 subjects in four independent cohorts were analyzed, including one European-American cohort (1,409 cases with alcohol dependence and 1,518 controls), one European-Australian cohort (a total of 6,438 family subjects with 1,645 probands), one African-American cohort from SAGE+COGA (681 cases and 508 controls) and one African-American cohort from Yale (1429 cases and 498 controls). The genome-wide association analysis was conducted for each cohort, and then a new meta-analysis was performed to derive the combined *p* values. *cis*-eQTL analysis of each risk variant in human tissues and RNA expression analysis of each risk gene in rat brain served as functional validation.

Results—In meta-analysis of European-American and European-Australian cohorts, we found 10 top-ranked SNPs ($p < 10^{-6}$) that were associated with alcohol dependence. They included 6 at *SERINC2* (3.1×10^{-8} p 9.6×10^{-8}), 1 at *STK40* ($p = 1.3 \times 10^{-7}$), 2 at *KIAA0040* (3.3×10^{-7} p 5.2×10^{-7}) and 1 at *IPO11* ($p = 6.9 \times 10^{-7}$). In meta-analysis of two African-American cohorts, we found 2 top-ranked SNPs including 1 at *SLC6A11* ($p = 2.7 \times 10^{-7}$) and 1 at *CBLN2* ($p = 7.4 \times 10^{-7}$). In meta-analysis of all four cohorts, we found 2 top-ranked SNPs in *PTP4A1-PHF3*

*Corresponding authors and reprints: Chunlong Zhong, MD, PhD, Dongfang Hospital, Shanghai Jiaotong University, Shanghai 200000, China; chunlongzhong@126.com; Tel: +86-21-38804518. Xingguang Luo, MD, Yale University School of Medicine, West Haven, CT 06516, USA; Xingguang.Luo@yale.edu; Tel: +1-203-932-5711; Fax: 203-937-4741.

Declaration of interest: The authors declare not conflict.

locus (6.0×10^{-7} p 7.2×10^{-7}). In an African-American cohort only, we found 1 top-ranked SNP at *PLDI* ($p=8.3 \times 10^{-7}$; OR=1.56). Many risk SNPs had positive *cis*-eQTL signals and all these risk genes except *KIAA0040* were found to express in both rat and mouse brains.

Conclusions—We found multiple genes that were significantly or suggestively associated with alcohol dependence. They are among the most appropriate for follow-up as contributors to risk for alcohol dependence.

Keywords

Alcohol dependence; Genome-wide association; Meta-analysis

Introduction

A true association between a disease and a causal genetic variant usually is replicable across different (and even heterogeneous) populations, and the gene effects of this causal variant usually can be cumulated with increasing sample sizes. Conventionally, meta-analysis estimates the weighted average of effects of the same allele (“nucleotide-based”) across different samples. When these effects have the same direction across different (but usually homogeneous) samples, they can be additive in meta-analysis, which results in a statistical power increase. However, if the gene effects have opposite directions in different (and usually heterogeneous) samples, or the gene effects are significant in one sample but non-significant in another, they will be neutralized in meta-analysis, which results in a statistical power decrease.

Between genetically heterogeneous populations, the allele frequencies of the same marker may be different. For example, we reported that 754,259 (75%) among one million SNPs were significantly ($p < 10^{-8}$) different in allele frequency between European-Americans and African-Americans; and the minor alleles ($f < 0.5$) of 157,718 (16%) SNPs in European-Americans were the major alleles ($f > 0.5$) in African-Americans, and vice versa [see Supplemental Materials and Methods by Zuo et al (Zuo et al., 2012)]. Suppose the causal allele of a putative disease-causal variant is a minor allele, it is expected that the risk allele of a marker in complete linkage disequilibrium (LD) with this causal allele would be a minor allele too, although this minor allele may be in opposite phases between European-Americans and African-Americans if this marker is among those 157,718 SNPs (Pei et al., 2012, Zuo et al., 2012). The combined effects of the same allele of such a marker (i.e., rare in one population but common in another) across these two populations by meta-analysis will be neutralized, which results in information loss. However, the combined effects of the minor allele of such a marker between these two populations will be additive, although the minor alleles are different between them. Under this circumstance, meta-analysis based on minor alleles would be more powerful. The minor allele frequency difference between any two populations should not exceed 0.5, so that the potential inflation effects of heterogeneity on meta-analysis can be limited and thus the heterogeneous samples might be able to be meta-analyzed. In the present study, we combined the effects of minor alleles (not necessary to be the same alleles) across four datasets via meta-analysis, to search for risk markers for alcohol dependence. Although this new approach (“minor allele-based”) is as reasonable as the conventional one (“nucleotide-based”), it is not necessary to be more powerful in all

circumstances. It is also subject to the disadvantage of meta-analysis described above; that is, the gene effects significant in one sample but non-significant in another even though in the same direction or the gene effects of minor alleles with opposite directions between two samples will be averaged by weight and become less significant by meta-analysis. In other words, some of the combined effects across all datasets might not be more significant than those across some subsets. Thus, it is necessary to test within some subsets before meta-analyzing all cohorts.

Materials and Methods

Subjects

A total of 12,481 subjects in four independent cohorts with alcohol dependence (DSM-IV) from dbGaP were analyzed, including one European-American SAGE+COGA cohort (1,409 cases with alcohol dependence and 1,518 controls), one European-Australian OZ-ALC cohort (a total of 6,438 family subjects with 1,645 alcohol dependent probands), one African-American SAGE+COGA cohort (681 cases and 508 controls) and one African-American Yale cohort (1429 cases and 498 controls) (Gelernter et al., 2014). SAGE and COGA cohorts were merged because their samples overlapped and were genotyped on the same platform. Subjects from the SAGE, COGA and OZ-ALC cohorts were interviewed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al., 1994); and subjects from the Yale cohort (Gelernter et al., 2014) were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) (Pierucci-Lagha et al., 2005). Affected subjects met lifetime DSM-IV criteria for alcohol dependence (American Psychiatric Association, 1994). The control subjects were defined as individuals who had been exposed to alcohol (and possibly to other drugs), but had never become addicted to alcohol or other illicit substances (lifetime diagnoses). 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 shown in Table 1 or is available elsewhere (Bierut et al., 2010, Edenberg et al., 2010, Heath et al., 2011, Zuo et al., 2012, Zuo et al., 2009, Zuo et al., 2013a, Zuo et al., 2011, Zuo et al., 2013d).

Genotyping and Imputation

The SAGE+COGA, OZ-ALC and Yale cohorts were genotyped on the Illumina Human 1M beadchip (with one million SNPs), the Illumina CNV370v1 beadchip (with 370,000 SNPs) and the Illumina HumanOmni1_Quad_v1-0_B beadchip (with one million SNPs), respectively. To make the genetic marker sets highly consistent across the different cohorts, we imputed the untyped SNPs in all samples based on the same reference panels. We used the following strategies to maximize the success rate and accuracy of imputation. (1) We used both 1,000 Genome Project and HapMap 3 panels as the reference. The CEU (CEPH Europeans) and YRI (Yoruba Africans) panels from these panels were used to impute untyped SNPs in European-origin samples and African-origin samples, respectively. Only the genotypes that were consistently imputed from these two independent reference panels were selected for analysis. (2) We used a Markov Chain Monte Carlo (MCMC) algorithm implemented in the program IMPUTE2 (Howie et al., 2009) to derive full posterior

probabilities (i.e., not the “best-guess”) of the genotypes of each SNP to minimize the inference bias. (3) We set the imputation parameters at burnin=10,000, iteration=10,000, k=100, Ne=11,500 and confidence level=0.99 when using IMPUTE2 (Howie et al., 2009); that is, the uncertainty rate of inference was less than 1%. (4) Because the imputation process using IMPUTE2 did not incorporate the family relationship information, Mendelian errors might occur in the imputed data. Thus, the families with at least one individual who had more than 0.5% Mendel errors (considering all SNPs tested) and the SNPs with more than 0.5% Mendel errors (considering all individuals tested) were excluded. Meanwhile, we also used the program BEAGLE (Browning and Browning, 2009) to impute genotypes independently. The imputation process using BEAGLE does incorporate the family relationship information. Only the imputed genotypes that were consistently imputed by both IMPUTE2 and BEAGLE were selected for analysis. And (5) we stringently cleaned the imputed genotype data after imputation. Furthermore, only the SNPs that had similar minor allele frequencies (with frequency difference < 0.2%) in the healthy controls across different cohorts and HapMap database (within the same ethnicity) were selected for analysis. After this strict selection, we were highly confident with the quality of these imputed genotype data. Finally, for SNPs that were directly genotyped, we used the direct genotypes rather than the imputed.

Data cleaning

We stringently cleaned the phenotype and genotype data before the association analysis. Subjects with poor genotypic data, allele discordance, sample relatedness, gender anomalies, chromosome anomalies (such as aneuploidy and mosaic cell populations), missing race, population group outliers, a mismatch between self-identified and genetically-inferred ethnicity, or a missing genotype call rate $\geq 2\%$ across all SNPs were excluded. Furthermore, SNPs with allele discordance, chromosomal anomalies or batch effect were also excluded. We then filtered out the SNPs on all chromosomes with an overall missing genotype call rate $\geq 2\%$, the monomorphic SNPs, and the SNPs with minor allele frequencies (MAFs) < 0.01 . The SNPs that deviated from HWE ($p < 10^{-4}$) within controls were also excluded.

Association test

The genome-wide association analysis was performed using logistic regression implemented in PLINK for each case-control cohort or using FBAT for the family-based cohort. Diagnosis and alleles each served as the dependent and independent variables, respectively, with sex, age and the first 10 principal components of ancestries as covariates. The principal component scores of our samples were derived from all autosomal SNPs across the genome using principal component analysis (PCA) implemented in the software package EIGENSTRAT (Price et al., 2006). Each individual received scores on each principal component. These principal components reflected the population structure of our samples. The first 10 principal component scores accounted for $> 90\%$ of variance. These PCs when included as covariates in the model can control for the population stratification and admixture effects on association analysis. The p-values derived from these association analyses are illustrated in Figure 1 or elsewhere (Zuo et al., 2011, Zuo et al., 2012, Zuo et al., 2013b). The genome-wide Bonferroni-corrected α was set at 5×10^{-8} (by one million

markers). The ORs from regression analyses and the Z scores from FBAT analysis were also derived. The directions of ORs and Zs were incorporated into the meta-analysis.

Meta-analysis

Meta-analysis was performed to derive the combined p values between cohorts using the program METAL (Willer et al., 2010). We recoded the minor allele of each SNP to “minor” and the major allele to “major”, regardless of whether the minor alleles were the same ones in different samples or not, which is the major difference from the conventional meta-analysis. After this recoding, the maximal difference of minor allele frequency among all control cohorts was 0.324 ($=0.420-0.096$ for rs11583322^AC) in this study, much less than 0.518 ($=0.720-0.202$ for rs350660^AC) before recoding (Table 2a). An overall z-statistic and an overall p-value for each SNP were calculated from a weighted sum of the individual statistics. Weights were proportional to the square-root of the number of individuals examined in each cohort and selected such that the squared weights sum to 1.0 (Willer et al., 2010). We tested meta-analysis for two cohorts with European ethnicity, two cohorts with African ethnicity, and globally for all four cohorts. Heterogeneity Indexes (I^2) were calculated for all meta-analyses using the program PLINK (Purcell et al., 2007).

cis-acting expression of quantitative locus (*Cis*-eQTL) analysis

To examine relationships between all risk SNPs and local mRNA expression levels, we examined the expression data in human lymphoblastoid cell lines from 270 unrelated HapMap individuals (Stranger et al., 2005). Differences in the distribution of mRNA expression levels between SNP genotypes were compared using a Wilcoxon-type trend test. We also examined the expression data in 93 autopsy-collected frontal cortical brain tissue samples with no defined neuropsychiatric condition and 80 peripheral blood mononuclear cell (PBMC) samples collected from living healthy donors (Heinzen et al., 2008). Differences in the distribution of mRNA expression levels between SNP genotypes were analyzed using a linear regression model by correcting for age, sex, source of tissues, and principal component scores. This *cis*-eQTL analysis served as functional validation for our association findings.

Rat Brain Transcriptome analysis

We generated RNA-Seq data from total RNA (after ribosomal RNA depletion) in brains of two inbred strains of rats (BN-Lx/CubPrin and SHR/OlaPrin) (Bhave et al., 2007). After reads were trimmed for adapters and quality, they were aligned to the RGSC5.0/rn5 version of the rat genome using TopHat2 (Kim et al., 2013). RNA expression was quantitated for Ensembl genes using CuffLinks (Trapnell et al., 2010). Read fragments per kilobase of transcript per million reads (FPKM) for each strain are reported. This animal study also served as functional validation for our association findings.

Results

The inflation factor (λ) from QQ plots for genome-wide p values in these four cohorts was 1.07, 1.03, 1.01 and 1.06, respectively (Zuo et al., 2012, Zuo et al., 2011, Zuo et al., 2013b) (Figure 1), which indicated that the population stratification effects have been successfully

controlled. Before meta-analysis, we found 10 top-ranked ($p < 10^{-6}$) risk markers associated with alcohol dependence in European Americans (SAGE+COGA) and 1 in African-Americans (Yale) (Gelernter et al., 2014). These risk markers were located at *SERINC2* (2.3×10^{-7} p 4.4×10^{-7}), *STK40* ($p = 4.0 \times 10^{-7}$), *KIAA0040* (2.9×10^{-7} p 4.0×10^{-7}) and *IPO11* ($p = 2.8 \times 10^{-7}$) in European-Americans (Zuo et al., 2012), and *PLD1* (phosphatidylcholine-specific phospholipase D1) ($p = 8.3 \times 10^{-7}$; OR=1.56) in African-Americans (data not shown).

In meta-analysis of European-American and European-Australian cohorts, we found 10 top-ranked SNPs ($p < 10^{-6}$) that were associated with alcohol dependence. They included 6 at *SERINC2* (3.1×10^{-8} p 9.6×10^{-8}) (Zuo et al., 2013b), 1 at *STK40* ($p = 1.3 \times 10^{-7}$), 2 at *KIAA0040* (3.3×10^{-7} p 5.2×10^{-7}) and 1 at *IPO11* ($p = 6.9 \times 10^{-7}$). Except for the one at *IPO11* that was nominally significant in meta-analysis of African cohorts ($p = 2.2 \times 10^{-2}$) but had opposite effect to that in Europeans, all other nine associations were not significant in African-American cohorts, so that adding African-American cohorts into the meta-analysis did not increase the power. In meta-analysis of two African-American cohorts, we found 2 top-ranked SNPs ($p < 10^{-6}$) that were associated with alcohol dependence. They included 1 at *SLC6A11* ($p = 2.7 \times 10^{-7}$) and 1 at *CBLN2* ($p = 7.4 \times 10^{-7}$). These associations were not significant in European cohorts, so that adding European cohorts into the meta-analysis did not increase the power either. In meta-analysis of all four cohorts, we found 2 top-ranked SNPs ($p < 10^{-6}$) that were associated with alcohol dependence. They both were located in *PTP4A1-PHF3* locus (6.0×10^{-7} p 7.2×10^{-7}). These two SNPs were nominally associated with alcohol dependence both in European cohorts (1.8×10^{-3} p 2.4×10^{-3}) and African cohorts (1.6×10^{-5} p 2.0×10^{-5}), but their risk alleles were in opposite phases between Europeans and Africans. Compared with the conventional approach based on risk alleles, our new approach based on minor alleles reduced about 25% of the maximal frequency difference of the target alleles among all control cohorts and about 10% of the heterogeneity index (I^2) when meta-analyzing these two SNPs across all cohorts (data not shown). After correction for multiple testing ($\alpha = 5 \times 10^{-8}$), three associations at *SERINC2* remained significant and others remained suggestively significant (Table 2b).

Significant *cis*-eQTL signals were presented in Table 2b. *cis*-eQTL analysis showed that rs4478858 at *SERINC2* had significant regulatory effect on the expression of *SERINC2* transcript both in human brain ($p = 0.024$) and PBMC sample ($p = 0.026$) (Zuo et al., 2013b). Rs350660 at *SLC6A11* had significant regulatory effect on the expression of *SLC6A11* transcript in the PBMC sample ($p = 0.016$). Rs1057239 ($p = 0.011$) and rs1894709 at *KIAA0040* ($p = 0.008$), rs6942342 at *PTP4A1* ($p = 0.008$), rs7652788 at *PLD1* ($p = 0.023$) and rs9294269 at *PHF3* ($p = 0.002$) had significant regulatory effects on the expression of local transcripts in HapMap samples (Table 2b).

All of the aforementioned risk genes except for *KIAA0040* have orthologous genes in rat. They were expressed above background in brain (FPKM > 1) of both inbred rat strains assessed (Table 3). The transcript for *KIAA0040* which is not thoroughly annotated in the human genome was not distinguishable among the RNAs expressed in rat brain.

Discussion

In the present study, we found multiple genes that were significantly or suggestively associated with alcohol dependence. Ten associations at four genes (*SERINC2*, *STK40*, *KIAA0040* and *IPO11*) were European-specific. Two associations at two genes (*SLC6A11* and *CBLN2*) were African-specific. Two associations at *PTP4A1-PHF3* locus were suggestively significant in both Europeans and Africans. The gene products of *SERINC2*, *CBLN2*, *SLC6A11*, *IPO11* and *PTP4A1* are all components of cellular membrane that can be enriched by David Functional Annotation Clustering Tool (Huang da et al., 2007). Additionally, one association at *PLD1* was cohort-specific that needed further validation in more cohorts in the future. These genes are among the most appropriate for follow-up as contributors to risk for alcohol dependence.

Most genetic risk markers are not the disease-causal variants *per se*, so that the allele frequencies of genetic risk markers may not be completely consistent with the causal variants. These allele frequencies could vary among different populations due to the different histories of evolution. It is not necessary that the same alleles are associated with causal alleles in different populations when these alleles have different frequencies among them. Taking the alleles with similar frequencies in different populations (called the putative causal alleles in the context), instead of the same allele, as the risk allele could be reasonable. In this case, combining the effects of alleles of the putative causal variants, instead of those of the markers *per se*, could be an alternative reasonable and powerful strategy for meta-analysis. Our findings from this strategy on *SERINC2*, *KIAA0040*, *STK40*, *IPO11* and *PTP4A1-PHF3* were generally consistent with previous reports that used genome-wide association analysis (Zuo et al., 2012, Zuo et al., 2013a, Zuo et al., 2013b, Zuo et al., 2011) and conventional strategies for meta-analysis (Wang et al., 2011). We previously reported that rare variant constellation across *SERINC2* was specific to alcohol dependence in European-origin populations (Zuo et al., 2013c), which is consistent with the current findings using common variants. Specifically pointing out here that the minor alleles of *PTP4A1-PHF3* SNPs were opposite between European-Americans and African-Americans. Only our meta-analysis using this new strategy detected association signals with $p < 10^{-6}$ in *PTP4A1-PHF3* region that a previous report ignored (Zuo et al., 2011). Furthermore, two novel top-ranked risk genes including *SLC6A11* (GABA transporter) and *CBLN2* (cerebellin 2 precursor) were also identified. Bioinformatics analysis showed that all of these top-ranked risk SNPs were located in the transcription factor binding sites (<http://brainarray.mbnj.med.umich.edu/Brainarray/Database/SearchSNP/snpfunc.aspx>). Many of them were located within methylated CpG islands, within copy number variations and/or had significant *cis*-eQTL signals in brain, PBMC or lymphoblastoid cell tissue (Table 2b), suggesting that they may be functional.

Since the propensity to develop alcohol dependence or other addictions is considered to be a component of brain function, we assessed the expression of RNA from each of the risk genes in brain of two strains of rats using RNA-Seq technology for measuring total RNA. The results from all of our measurements are available on <http://phenogen.ucdenver.edu>. Keeping in mind the caveat that expression in rat brain may not wholly mimic the expression of transcripts in brain of humans, we found that eight of the nine risk genes were expressed

in rat brain. We also verified that the same eight genes were also expressed above background in mouse brain (data not shown; ERP000614).

The first observation generated from our gene expression studies in rat brain was the large difference in expression of the *Serinc2* gene product between the two strains of rats (one of which consumes 2× the alcohol as the other in a 2 bottle choice paradigm) (Tabakoff et al., 2009). In studies with humans, several SNPs across the length of this gene were significantly associated with alcohol dependence in European Americans and European Australians (Zuo et al., 2013b). One could speculate that the involvement of *SERINC2* in the etiology of alcohol dependence may be related to its expression levels. The Serinc proteins are critical in the regulation of lipid biosynthesis in brain, particularly phosphatidylserine and spingolipids. Serinc2 is localized primarily to neurons and the activity of Serinc in neurons may be coupled to glutamatergic transmission (Inuzuka et al., 2005) (i.e., phosphatidylserine levels in neuron membranes modulate glutamate release (Yang and Wang, 2009)). *Slc6a11* was also highly expressed in rat brain, and although one cannot ascertain from our meta-analysis whether the product of this gene is differentially expressed in humans, or structural variants are possible, it is important to note that the *SLC6A11* gene product is a GABA transporter located in astrocytes, as well as neurons (Itouji et al., 1996) and is of importance in reducing synaptic and extrasynaptic levels of GABA (Song et al., 2013). It is clear in studies of humans (Edenberg and Foroud, 2006, Covault et al., 2004) and animals (Tabakoff et al., 2009) that GABA neurotransmitter function is intimately related to alcohol dependence, but studies differ on which aspect of the GABA signal may vary with regard to alcohol dependence. It should also be noted that the product of the *PTP4A1* gene was previously assessed by in-situ hybridization in brains of humans. The *PTP4A1* transcript showed a disparate distribution in the two tested brain areas (Dumaual et al., 2006). Little or no *PTP4A1* RNA was found in the “cerebral cortex”, but moderate expression was noted in the cerebellum granule cell layer. Our measures in the whole brain of the rat, however, indicated robust expression of this dual specificity phosphatase, the protein product of which localizes to neuronal membranes, but signals through cAMP-dependent transcription factors (Rios et al., 2013). *Pld1* and *Cbln2* were also expressed in both strains of rat brains. *PLD1* encodes the phosphatidylcholine-specific phospholipase D (PLD) which catalyzes the hydrolysis of phosphatidylcholine in order to yield phosphatidic acid and choline. PLD has a high affinity for short chain alcohols (100–1000-fold higher than for water). In the presence of ethanol, it promotes a transphosphatidyl reaction, with the production of phosphatidylethanol (PEth). The expression of PEth in blood is a direct marker of chronic alcohol use and abuse (Viel et al., 2012). *CBLN2* encodes a cerebellin 2 precursor. Cerebellin-2 is a secreted glycoprotein and may serve as a transneuronal cytokine involved in the regulation of synapse development and synaptic plasticity in various brain regions. *CBLN2* has been associated with alcohol dependence in a genome-wide association study (GWAS) (Lydall et al., 2011). These genes are more likely to be the contributors to risk for alcohol dependence. Their mechanisms underlying these contributions are worth of more follow-up investigation.

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 AA021380, R21 AA020319 and R24 AA013162, the National Alliance for Research on Schizophrenia and Depression (NARSAD) Award 17616 (to LZ) and ABMRF/The Foundation for Alcohol Research (LZ). We thank NIH GWAS Data Repository (dbGaP), the Contributing Investigator(s) (Drs. Bierut, Edenberg, Heath, Gelernter and Kranzler) who contributed the phenotype and genotype data (SAGE: phs000092.v1.p1, COGA: phs000125.v1.p1, OZ-ALC: phs000181.v1.p1 and Yale CIDR-Gelernter Study: phs000425.v1.p1) 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

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. fourth. American Psychiatric Press; Washington, DC: 1994.
- Bhave SV, Hornbaker C, Phang TL, Saba L, Lapadat R, Kechris K, Gaydos J, McGoldrick D, Dolbey A, Leach S, Soriano B, Ellington A, Ellington E, Jones K, Mangion J, Belknap JK, Williams RW, Hunter LE, Hoffman PL, Tabakoff B. The PhenoGen informatics website: tools for analyses of complex traits. *BMC Genet.* 2007; 8:59. [PubMed: 17760997]
- Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Gruzsa RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nothen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP. Gene Environment Association Studies Consortium. A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences of the United States of America.* 2010; 107:5082–5087. [PubMed: 20202923]
- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American journal of human genetics.* 2009; 84:210–223. [PubMed: 19200528]
- Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol.* 1994; 55:149–158. [PubMed: 8189735]
- Covault J, Gelernter J, Hesselbrock V, Nellissery M, Kranzler HR. Allelic and haplotypic association of GABRA2 with alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet.* 2004; 129B:104–109. [PubMed: 15274050]
- Dumaual CM, Sandusky GE, Crowell PL, Randall SK. Cellular localization of PRL-1 and PRL-2 gene expression in normal adult human tissues. *J Histochem Cytochem.* 2006; 54:1401–1412. [PubMed: 16957164]
- Edenberg HJ, Foroud T. The genetics of alcoholism: identifying specific genes through family studies. *Addict Biol.* 2006; 11:386–396. [PubMed: 16961766]
- Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, Bierut LJ, Bucholz KK, Goate A, Aliev F, Dick D, Hesselbrock V, Hinrichs A, Kramer J, Kuperman S, Nurnberger JI Jr, Rice JP, Schuckit MA, Taylor R, Todd Webb B, Tischfield JA, Porjesz B, Foroud T. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcoholism, clinical and experimental research.* 2010; 34:840–852.
- Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, Anton R, Preuss UW, Ridinger M, Rujescu D, Wodarz N, Zill P, Zhao H, Farrer LA. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry.* 2014; 19:41–49. [PubMed: 24166409]
- Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, McEvoy BP, Schrage AJ, Grant JD, Chou YL, Zhu R, Henders AK, Medland SE, Gordon SD, Nelson EC, Agrawal A,

- Nyholt DR, Bucholz KK, Madden PA, Montgomery GW. A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biological psychiatry*. 2011; 70:513–518. [PubMed: 21529783]
- Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, Welsh-Bohmer KA, Hulette CM, Denny TN, Goldstein DB. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol*. 2008; 6:e1.
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; 5:e1000529. [PubMed: 19543373]
- Huang da W, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 2007; 8:R183. [PubMed: 17784955]
- Inuzuka M, Hayakawa M, Ingi T. Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *The Journal of biological chemistry*. 2005; 280:35776–35783. [PubMed: 16120614]
- Itouji A, Sakai N, Tanaka C, Saito N. Neuronal and glial localization of two GABA transporters (GAT1 and GAT3) in the rat cerebellum. *Brain Res Mol Brain Res*. 1996; 37:309–316. [PubMed: 8738166]
- 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]
- Lydall GJ, Bass NJ, McQuillin A, Lawrence J, Anjorin A, Kandaswamy R, Pereira A, Guerrini I, Curtis D, Vine AE, Sklar P, Purcell SM, Gurling HM. Confirmation of prior evidence of genetic susceptibility to alcoholism in a genome-wide association study of comorbid alcoholism and bipolar disorder. *Psychiatric genetics*. 2011; 21:294–306. [PubMed: 21876473]
- Pei YF, Zhang L, Yang TL, Han Y, Hai R, Ran S, Tian Q, Shen H, Li J, Zhu XZ, Luo X, Deng HW. Genome-Wide Association Study of Copy Number Variants Suggests LTBP1 and FGD4 Are Important for Alcohol Drinking. *PLoS One*. 2012; 7:e30860. [PubMed: 22295116]
- Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, Farrer L, Kranzler HR. Diagnostic reliability of the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug and alcohol dependence*. 2005; 80:303–312. [PubMed: 15896927]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*. 2006; 38:904–909. [PubMed: 16862161]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007; 81:559–575. [PubMed: 17701901]
- Rios P, Li X, Kohn M. Molecular mechanisms of the PRL phosphatases. *FEBS J*. 2013; 280:505–524. [PubMed: 22413991]
- Song I, Volynski K, Brenner T, Ushkaryov Y, Walker M, Semyanov A. Different transporter systems regulate extracellular GABA from vesicular and non-vesicular sources. *Frontiers in cellular neuroscience*. 2013; 7:23. [PubMed: 23494150]
- Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavaré S, Deloukas P, Dermitzakis ET. Genome-wide associations of gene expression variation in humans. *PLoS Genet*. 2005; 1:e78. [PubMed: 16362079]
- Tabakoff B, Saba L, Printz M, Flodman P, Hodgkinson C, Goldman D, Koob G, Richardson HN, Kechris K, Bell RL, Hubner N, Heinig M, Pravenec M, Mangion J, Legault L, Dongier M, Conigrave KM, Whitfield JB, Saunders J, Grant B, Hoffman PL. Genetical genomic determinants of alcohol consumption in rats and humans. *BMC Biol*. 2009; 7:70. [PubMed: 19874574]
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28:511–515. [PubMed: 20436464]

- Viel G, Boscolo-Berto R, Cecchetto G, Fais P, Nalesso A, Ferrara SD. Phosphatidylethanol in blood as a marker of chronic alcohol use: a systematic review and meta-analysis. *International journal of molecular sciences*. 2012; 13:14788–14812. [PubMed: 23203094]
- 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. *Journal of psychiatric research*. 2011; 45:1419–1425. [PubMed: 21703634]
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–2191. [PubMed: 20616382]
- Yang TT, Wang SJ. Facilitation of glutamate release from rat cerebrocortical glutamatergic nerve terminals (synaptosomes) by phosphatidylserine and phosphatidylcholine. *Synapse*. 2009; 63:215–223. [PubMed: 19072841]
- Zuo L, Gelernter J, Zhang CK, Zhao H, Lu L, Kranzler HR, Malison RT, Li CR, Wang F, Zhang XY, Deng HW, Krystal JH, Zhang F, Luo X. Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2012; 37:557–566. [PubMed: 21956439]
- Zuo L, Luo X, Listman JB, Kranzler HR, Wang S, Anton RF, Blumberg HP, Stein MB, Pearlson GD, Covault J, Charney DS, van Kammen DP, Price LH, Lappalainen J, Cramer J, Krystal JH, Gelernter J. Population admixture modulates risk for alcohol dependence. *Hum Genet*. 2009; 125:605–613. [PubMed: 19306106]
- Zuo L, Wang F, Zhang XY, Li CSR, Lu L, Ye L, Zhang H, Krystal JH, Deng HW, Luo X. Genome-wide significant association signals in IPO11-HTR1A region specific for alcohol and nicotine co-dependence. *Alcoholism, clinical and experimental research*. 2013a; 37:730–739.
- Zuo L, Wang K, Zhang X JHK, Li CR, Zhang F, Zhang H, Luo X. NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk region specific for alcohol dependence in subjects of European descent. *Drug and alcohol dependence*. 2013b; 129:254–264. [PubMed: 23455491]
- Zuo L, Wang KS, Zhang XY, Li CS, Zhang F, Wang X, Chen W, Gao G, Zhang H, Krystal JH, Luo X. Rare SERINC2 variants are specific for alcohol dependence in individuals of European descent. *Pharmacogenet Genomics*. 2013c; 23:395–402. [PubMed: 23778322]
- Zuo L, Zhang CK, Wang F, Li CS, Zhao H, Lu L, Zhang XY, Lu L, Zhang H, Zhang F, Krystal JH, Luo X. A novel, functional and replicable risk gene region for alcohol dependence identified by genome-wide association study. *PLoS One*. 2011; 6:e26726. [PubMed: 22096494]
- Zuo L, Zhang H, Malison RT, Li CSR, Zhang XY, Wang F, Lu L, Lu L, Wang X, Krystal JH, Zhang F, Deng H, Luo X. Rare ADH variant constellations are specific for alcohol dependence *Alcohol*. 2013d; 48:9–14. [PubMed: 23019235]

Figure 1. QQ-plot for the p-values in four cohorts

[Left top: in European-Americans; left bottom: in African-Americans; right top: in European-Australians; right bottom: in Yale African-Americans; X-axis: expected $-\log(p)$; Y-axis: observed $-\log(p)$; λ , inflation factor]

Table 1

Demographic data of all cohorts

Cohort	Pedigrees		Subjects		Affected subjects				Unaffected subjects			
	Total		Total		Total		Female		Total		Male	
	N		N		N		N	Age (yrs)	N		N	Age (yrs)
European-American SAGE+COGA	2927		2927		1409	883	39.0±10.4	526	36.7±8.8	1518	445	37.9±10.1
European-American SAGE+COGA	2927		2927		1409	883	39.0±10.4	526	36.7±8.8	1518	445	37.9±10.1
European-Australian OZ-ALC	2252		6438		1645	1020	42.0±8.4	625	39.2±7.3	3922	1709	46.3±9.8
African-American SAGE+COGA	1189		1189		681	428	41.0±8.3	253	39.8±6.8	508	169	40.2±8.4
African-American Yale	1927		1927		1429	858	42.4±8.4	571	40.1±8.2	498	145	38.5±12.7
												38.5±12.9

In the European-Australian family data, only the affected and unaffected offspring are listed. N, sample size; yrs, years.

Table 2

a. Top-ranked risk markers for alcohol dependence in four cohorts																					
European-American							European-Australian					African-American				African-American (Y)					
SNP	Gene	Allele	f(case)	f(con)	OR	p	Allele	f(case)	f(con)	OR	p	Allele	f(case)	f(con)	OR	p	Allele	f(case)	f(con)	OR	p
rs1039630	5' to <i>SERINC2</i>	C	0.451	0.385	1.32	2.6×10 ⁻⁷	C	0.431	0.404	1.25	0.049	T	0.426	0.433	0.96	0.665	T	0.425	0.437	0.94	0.415
rs4478858	<i>SERINC2</i>	G	0.453	0.389	1.31	4.4×10 ⁻⁷	G	0.448	0.426	1.26	0.021	A	0.445	0.458	0.95	0.505	A	0.458	0.458	0.99	0.891
rs4949400	<i>SERINC2</i>	T	0.455	0.390	1.32	2.3×10 ⁻⁷	T	0.444	0.416	1.12	0.116	C	0.208	0.214	0.93	0.476	C	0.210	0.187	1.11	0.285
rs4949402	<i>SERINC2</i>	T	0.455	0.390	1.32	2.6×10 ⁻⁷	T	0.441	0.414	1.12	0.109	C	0.211	0.216	0.94	0.532	C	0.212	0.188	1.11	0.249
rs2275436	<i>SERINC2</i>	C	0.463	0.396	1.32	2.4×10 ⁻⁷	C	0.475	0.461	1.34	0.043	T	0.206	0.208	0.96	0.660	T	0.184	0.160	1.13	0.257
rs2275435	<i>SERINC2</i>	G	0.456	0.391	1.31	3.0×10 ⁻⁷	G	0.445	0.416	1.13	0.097	A	0.208	0.215	0.93	0.465	A	0.191	0.174	1.06	0.554
rs11583322	<i>STK40</i>	C	0.359	0.420	0.76	4.0×10 ⁻⁷	C	0.387	0.411	0.90	0.451	C	0.097	0.096	1.02	0.892	C	0.101	0.115	0.80	0.069
rs1057239	<i>KIAA0040</i>	T	0.468	0.400	1.32	1.3×10 ⁻⁷	T	0.385	0.399	0.95	0.195	T	0.379	0.411	0.87	0.105	T	0.387	0.393	0.96	0.626
rs1894709	<i>KIAA0040</i>	A	0.468	0.401	1.31	1.7×10 ⁻⁷	A	0.385	0.399	0.94	0.195	A	0.379	0.412	0.87	0.095	A	0.388	0.395	0.96	0.569
rs350660	<i>SLC6A11</i>	C	0.207	0.213	0.99	0.835	C	0.205	0.202	1.02	0.211	T	0.290	0.322	0.83	0.039	T	0.280	0.346	0.69	3.4×10 ⁻⁶
rs7445832	<i>IPO11</i>	A	0.272	0.213	1.38	2.8×10 ⁻⁷	A	0.233	0.228	1.03	0.452	A	0.360	0.389	0.89	0.162	A	0.355	0.392	0.86	0.068
rs6942342	5' to <i>PTP4A1</i>	T	0.466	0.429	1.18	1.8×10 ⁻³	T	0.460	0.455	1.23	0.062	C	0.254	0.175	1.56	2.0×10 ⁻⁵	C	0.220	0.208	1.04	0.682
rs9294269	5' to <i>PHF3</i>	A	0.467	0.429	1.17	2.4×10 ⁻³	A	0.462	0.453	1.24	0.059	C	0.259	0.179	1.56	1.6×10 ⁻⁵	C	0.217	0.202	1.06	0.548
rs12969601	<i>CBLN2</i>	T	0.260	0.280	0.89	0.057	T	0.261	0.275	0.93	0.725	C	0.442	0.506	0.75	8.9×10 ⁻⁴	C	0.455	0.519	0.76	2.3×10 ⁻⁴

b. Top-ranked risk markers for alcohol dependence in meta-analysis (p<10 ⁻⁶)											
SNP	Chr	Position (Build 36)	Bioinformatics analysis	EA and EAu				Meta-analysis AA and AA (Y)			
				P	Z	P	Z	P	Z		
										P	Z
rs1057239	1	173396827	TFBS/CpG/eQTL	3.3×10 ⁻⁷	5.11						
rs1894709	1	173398994	TFBS/eQTL	5.2×10 ⁻⁷	5.02						
rs350660	3	10772131	TFBS/CpG/eQTL			2.7×10 ⁻⁷	5.14				
rs7445832	5	62622057	TFBS/CpG	6.9×10 ⁻⁷	4.97	2.2×10 ⁻²	-2.29				
rs6942342	6	64234406	TFBS/eQTL	1.8×10 ⁻³	2.95	2.0×10 ⁻⁵	2.98	6.0×10 ⁻⁷	4.99		
rs9294269	6	64395256	TFBS/eQTL	2.4×10 ⁻³	2.97	1.6×10 ⁻⁵	3.23	7.2×10 ⁻⁷	4.96		
rs12969601	18	67486279	TFBS			7.4×10 ⁻⁷	-4.95				

African-American, African-American SAGE+COGA cohorts; African-American (Y), African-American Yale cohort; f(case) and f(con), minor allele frequencies in cases and controls, respectively; OR, odd ratios; Z, Z scores; p, p-value.

Meta-analysis with $p>10^{-6}$ are not shown in this table. EA, European-Americans; EAu, European-Americans; AA, African-American SAGE+COGA cohorts; AA (Yale), African-American Yale cohort. P, p-value of meta-analysis; Z, z scores; TFBS, these SNPs are located in the transcription factor binding sites; ESS, this marker is located in an exonic splicing silencer or enhancer; CpG, these SNPs are located within CpG islands; CNV, these SNPs are located within methylated CNVs (see UCSC Genome Browser); eQTL, this SNP has significant transcript-level cis-eQTL signal ($p<0.05$) in brain, peripheral blood mononuclear cell (PBMC) or lymphoblastoid cell tissue.

Table 3

Measures of risk gene expression levels in brain of two strains of rats

Gene Symbol	BN-Lx/CubPrin (FPKM)	SHR/OlaPrin (FPKM)
Serinc2	1.61	14.93
Stk40	3.65	4.60
Slc6a11	10.59	11.67
Pld1	2.90	2.64
Ipo11	5.33	5.76
Ptp4a1	16.51	14.67
Phf3	8.53	8.23
bln2	7.32	7.89

FPKM - read fragments per kilobase of transcript per million read fragments. RNA-seq was used for measuring the presence of the gene products in rat brain