A New Genomewide Association Meta-Analysis of Alcohol Dependence

Lingjun Zuo, Yunlong Tan, Xiangyang Zhang, Xiaoping Wang, John Krystal, Boris Tabakoff, Chunlong Zhong, and Xingguang Luo

Background: Conventional meta-analysis based on genetic markers may be less powerful for heterogeneous samples. In this study, we introduced a new meta-analysis for 4 genomewide association studies on alcohol dependence that integrated the information of putative causal variants.

Methods: A total of 12,481 subjects in 4 independent cohorts were analyzed, including 1 European American cohort (1,409 cases with alcohol dependence and 1,518 controls), 1 European Australian cohort (a total of 6,438 family subjects with 1,645 probands), 1 African American cohort from SAGE + COGA (681 cases and 508 controls), and 1 African American cohort from Yale (1,429 cases and 498 controls). The genomewide association analysis was conducted for each cohort, and then, a new meta-analysis was performed to derive the combined *p*-values. *cis*-Acting expression of quantitative locus (*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 single nucleotide polymorphisms (SNPs) ($p < 10^{-6}$) that were associated with alcohol dependence. They included 6 at *SERINC2* ($3.1 \times 10^{-8} \le p \le 9.6 \times 10^{-8}$), 1 at *STK40* ($p = 1.3 \times 10^{-7}$), 2 at *KIAA0040* ($3.3 \times 10^{-7} \le p \le 5.2 \times 10^{-7}$), and 1 at *IPO11* ($p = 6.9 \times 10^{-7}$). In meta-analysis of 2 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 4 cohorts, we found 2 top-ranked SNPs in *PTP4A1-PHF3* locus ($6.0 \times 10^{-7} \le p \le 7.2 \times 10^{-7}$). In an African American cohort only, we found 1 top-ranked SNP at *PLD1* ($p = 8.3 \times 10^{-7}$; 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.

Key Words: Alcohol Dependence, Genomewide Association, Meta-Analysis.

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 nonsignificant 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 1 million single nucleotide polymorphisms (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., (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 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 2

From the Department of Psychiatry (LZ, JK, XL), Yale University School of Medicine, New Haven, Connecticut; Biological Psychiatry Research Center (YT, XL), Beijing Huilongguan Hospital, Beijing, China; Menninger Department of Psychiatry and Behavioral Sciences (XZ), Baylor College of Medicine, Houston, Texas; Department of Neurology (XW), First People's Hospital, Shanghai Jiaotong University, Shanghai, China; Department of Pharmacology (BT), University of Colorado School of Medicine, Aurora, Colorado; and Department of Neurosurgery (CZ), Renji Hospital, Shanghai Jiaotong University, Shanghai, China.

Received for publication July 30, 2014; accepted May 18, 2015.

Reprint requests: Xingguang Luo, MD, Yale University School of Medicine, West Haven, CT 06516; Tel.: 203-932-5711; Fax: 203-937-4741; E-mails: Xingguang.Luo@yale.edu or chunlongzhong@126.com

Copyright © 2015 by the Research Society on Alcoholism.

DOI: 10.1111/acer.12786

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 2 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 (MAF) difference between any 2 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 this study, we combined the effects of minor alleles (not necessary to be the same alleles) across 4 data sets 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 nonsignificant in another even though in the same direction or the gene effects of minor alleles with opposite directions between 2 samples will be averaged by weight and become less significant by meta-analysis. In other words, some of the combined effects across all data sets 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 4 independent cohorts with alcohol dependence (DSM-IV) from dbGaP were analyzed, including 1 European American Study of Addiction: Genetics and Environment (SAGE) + Collaborative Study on the Genetics of Alcoholism (COGA) cohort (1,409 cases with alcohol dependence and 1,518 controls), 1 European Australian Alcohol Research using Australian twins and their families (OZ-ALC) cohort (a total of 6,438 family subjects with 1,645 alcohol dependent probands), 1 African American SAGE + COGA cohort (681 cases and 508 controls), and 1 African American Yale cohort (1,429 cases and 498 controls). 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 (Bucholz et al., 1994), and subjects from the Yale cohort were interviewed using the Semi-Structured Assessment for Drug Dependence and Alcoholism (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., 2009, 2011, 2012, 2013a,d).

Genotyping and Imputation

The SAGE + COGA, OZ-ALC, and Yale cohorts were genotyped on the Illumina Human 1M beadchip (with 1 million SNPs; Illumina, Inc., San Diego, CA), the Illumina CNV370v1 beadchip (with 370,000 SNPs), and the Illumina HumanOmni1_Quad_v1-0_B beadchip (with 1 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. (i) 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 2 independent reference panels were selected for analysis. (ii) We used a Markov chain Monte Carlo 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. (iii) We set the imputation parameters at burn-in = 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 <1%. (iv) 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 1 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

Table 1. Demographic Data of All Cohorts

	Pedigrees	Subjects	Affected subjects					Unaffected subjects				
				Male		Female			Male		Female	
Cohort	Total *N*	Total *N*	Total *N*	*N*	Age (year)	*N*	Age (year)	Total *N*	*N*	Age (year)	*N*	Age (years)
European American SAGE + COGA	2,927	2,927	1,409	883	39.0 ± 10.4	526	36.7 ± 8.8	1,518	445	37.9 ± 10.1	1,073	39.0 ± 9.1
European Australian OZ-ALC	2,252	6,438	1,645	1,020	42.0 ± 8.4	625	39.2 ± 7.3	3,922	1,709	46.3 ± 9.8	2,213	45.6 ± 9.5
African American SAGE + COGA	1,189	1,189	681	428	41.0 ± 8.3	253	39.8 ± 6.8	508	169	40.2 ± 8.4	339	39.6 ± 6.8
African American Yale	1,927	1,927	1,429	858	42.4 ± 8.4	571	40.1 ± 8.2	498	145	38.5 ± 12.7	353	38.5 ± 12.9

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

BEAGLE were selected for analysis. And (v) we stringently cleaned the imputed genotype data after imputation. Furthermore, only the SNPs that had similar MAFs (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 ≥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 ≥2%, the monomorphic SNPs, and the SNPs with MAFs <0.01. The SNPs that deviated from Hardy-Weinberg equilibrium ($p < 10^{-4}$) within controls were also excluded.

Association Test

The genomewide association analysis was performed using logistic regression implemented in PLINK (Purcell et al., 2007) for each case–control cohort or using FBAT for the family-based cohort (Laird et al., 2000). 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 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 principal components 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 Fig. 1 or elsewhere (Zuo et al., 2011, 2012, 2013b). The genomewide Bonferroni-corrected α was set at 5×10^{-8} (by 1 million markers). The odds ratios (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 MAF among all control cohorts was 0.324 (=0.420-0.096 for rs11583322^C) in this study, much less than 0.518 (=0.720-0.202 for rs350660^C) before recoding (Table 2). An overall z-statistic and an overall *p*-value for each SNP were calculated from a weighted sum of the individual statistics. Weights were propor-

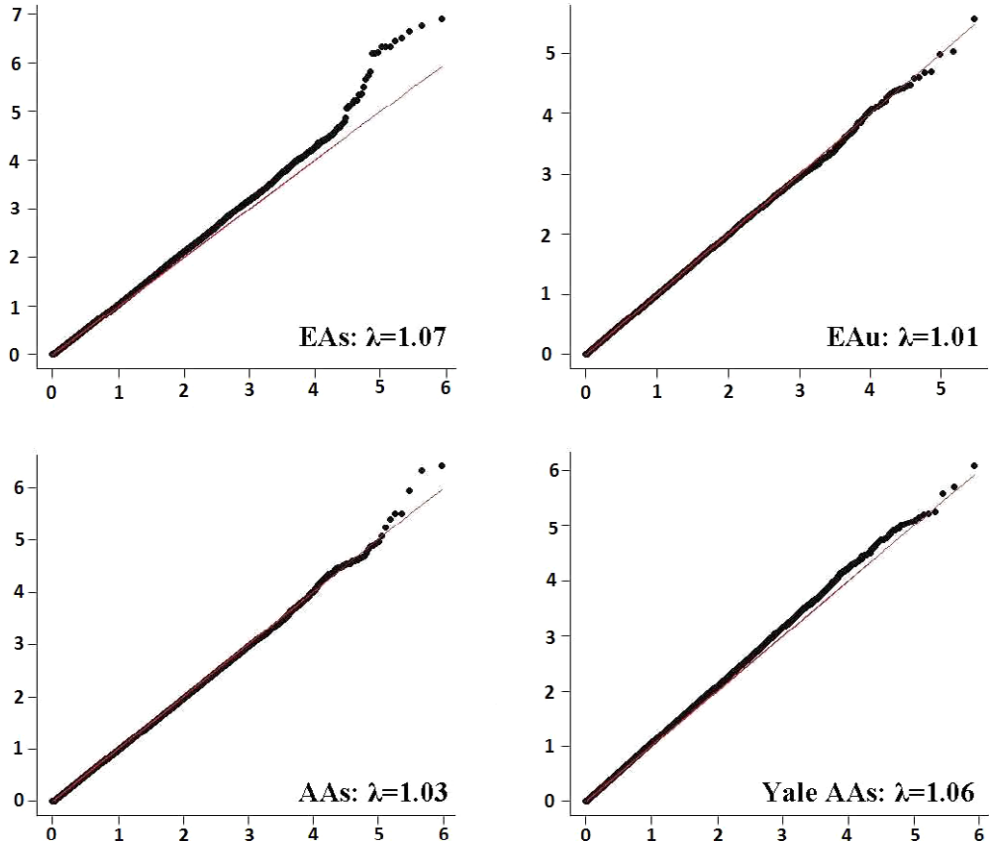

Fig. 1. QQ plot for the *p*-values in 4 cohorts.

Table 2. Top-Ranked Risk Markers for Alcohol Dependence in 4 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 SERINC2*	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	*SERINC2*	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	*SERINC2*	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	*SERINC2*	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	*SERINC2*	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	*SERINC2*	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	*STK40*	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	*KIAA0040*	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	*KIAA0040*	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	*SLC6A11*	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	*IPO11*	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 PTP4A1*	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 PHF3*	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	*CBLN2*	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}

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, odds ratios; *Z*, *Z* scores; *p*, *p*-value.

tional 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 2 cohorts with European ethnicity, 2 cohorts with African ethnicity, and globally for all 4 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 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*-acting expression of quantitative locus (*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 2 inbred strains of rats (BN-Lx/Cub-Prin 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 genomewide *p*-values in these 4 cohorts was 1.07, 1.03, 1.01, and 1.06, respectively (Zuo et al., 2011, 2012, 2013b) (Fig. 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). These risk markers were located at *SERINC2* ($2.3 \times 10^{-7} \leq p \leq 4.4 \times 10^{-7}$), *STK40* ($p = 4.0 \times 10^{-7}$), *KIAA0040* ($2.9 \times 10^{-7} \leq p \leq 4.0 \times 10^{-7}$), 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* ($3.1 \times 10^{-8} \leq p \leq 9.6 \times 10^{-8}$) (Zuo et al., 2013b), 1 at *STK40* ($p = 1.3 \times 10^{-7}$), 2 at *KIAA0040* ($3.3 \times 10^{-7} \leq p \leq 5.2 \times 10^{-7}$), 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 9 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 2 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 4 cohorts, we found 2 top-ranked SNPs ($p < 10^{-6}$) that were associated with alcohol dependence. They both were located in *PTP4A1-PHF3* locus ($6.0 \times 10^{-7} \leq p \leq 7.2 \times 10^{-7}$). These 2 SNPs were nominally associated with alcohol dependence both in European cohorts ($1.8 \times 10^{-3} \leq p \leq 2.4 \times 10^{-3}$) and in African cohorts ($1.6 \times 10^{-5} \leq p \leq 2.0 \times 10^{-5}$), 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 2 SNPs across all cohorts (data not shown). After correction for multiple testing ($\alpha = 5 \times 10^{-8}$), 3 associations at *SERINC2* remained significant and others remained suggestively significant (Table 3).

Significant *cis*-eQTL signals were presented in Table 3. *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 in 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 3).

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 4). 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 4 genes (*SERINC2*, *STK40, KIAA0040,* and *IPO11*) were European specific. Two associations at 2 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 membranes that can be enriched by David Functional Annotation Clustering Tool (Huang da et al., 2007). Additionally, 1 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

Table 3. Top-Ranked Risk Markers for Alcohol Dependence in Meta-Analysis ($p < 10^{-6}$)

				Meta-analysis					
				EA and EAu		AA and AA (Y)		All	
SNP	Chr	Position (Build 36)	Bioinformatics Analysis	P	Z	P	Z	P	Z
rs1039630	1	31654059	TFBS	4.3×10^{-8}	5.48				
rs4478858	1	31656512	TFBS/eQTL	3.1×10^{-8}	5.53				
rs4949400	1	31670719	TFBS	9.3×10^{-8}	5.34				
rs4949402	1	31670821	TFBS/ESS	7.3×10^{-8}	5.39				
rs2275436	1	31671050	TFBS/CpG/CNV	3.8×10^{-8}	5.5				
rs2275435	1	31671123	TFBS/CpG	9.6×10^{-8}	5.33				
rs11583322	1	36594899	TFBS	1.3×10^{-7}	5.28				
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		

EA, European Americans; EAu, European Australians; AA, African American SAGE + COGA cohorts; AA (Yale), African American Yale cohort; P, *p*-value of meta-analysis; Z, *Z* scores; TFBS, these single nucleotide polymorphisms (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 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.

Meta-analysis with $p > 10^{-6}$ is not shown in this table.

Table 4. Measures of Risk Gene Expression Levels in Brain of 2 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.

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 genomewide association analysis (Zuo et al., 2011, 2012, 2013a,b) 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, 2 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.mbni.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 3), suggesting that they may be functional.

As 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 2 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 8 of the 9 risk genes were expressed in rat brain. We also verified that the same 8 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 the expression of the *Serinc2* gene product between the 2 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 sphingolipids. 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 (Covault et al., 2004; Edenberg and Foroud, 2006) 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 2 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 to yield phosphatidic acid and choline. PLD has a high affinity for short chain alcohols (100- to 1,000-fold higher than for water). In the presence of ethanol, it promotes a transphosphatidylation 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 genomewide 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, R21 AA023237, and R24 AA013162, the National Alliance for Research on Schizophrenia and Depression (NARSAD) Award 17616 (to LZ), and ABMRF/The Foundation for Alcohol Research (to 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 HHSN2682011000011I. 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.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

REFERENCES

American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders. 4th ed. American Psychiatric Press, Washington, DC.

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 (2007) The PhenoGen informatics website: tools for analyses of complex traits. BMC Genet 8:59.

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, Grucza 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 (2010) A genome-wide association study of alcohol dependence. Proc Natl Acad Sci USA 107:5082–5087.

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA (1994) A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol 55:149–158.

Covault J, Gelernter J, Hesselbrock V, Nellissery M, Kranzler HR (2004) Allelic and haplotypic association of GABRA2 with alcohol dependence. Am J Med Genet B Neuropsychiatr Genet 129B:104–109.

Dumaual CM, Sandusky GE, Crowell PL, Randall SK (2006) Cellular localization of PRL-1 and PRL-2 gene expression in normal adult human tissues. J Histochem Cytochem 54:1401–1412.

Edenberg HJ, Foroud T (2006) The genetics of alcoholism: identifying specific genes through family studies. Addict Biol 11:386–396.

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 (2010) Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol Clin Exp Res 34:840–852.

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 (2011) A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. Biol Psychiatry 70:513–518.

Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, Welsh-Bohmer KA, Hulette CM, Denny TN, Goldstein DB (2008) Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS Biol 6:e1.

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529.

Huang da W, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8:R183.

Inuzuka M, Hayakawa M, Ingi T (2005) Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. J Biol Chem 280:35776–35783.

Itouji A, Sakai N, Tanaka C, Saito N (1996) Neuronal and glial localization of two GABA transporters (GAT1 and GAT3) in the rat cerebellum. Brain Res 37:309–316.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36.

Laird N, Horvath S, Xu X (2000) Implementing a unified approach to family based tests of association. Genet Epidemiol 19(Suppl 1):S36–S42.

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 (2011) Confirmation of prior evidence of genetic susceptibility to alcoholism in a genome-wide association study of comorbid alcoholism and bipolar disorder. Psychiatr Genet 21:294–306.

Pei YF, Zhang L, Yang TL, Han Y, Hai R, Ran S, Tian Q, Shen H, Li J, Zhu XZ, Luo X, Deng HW (2012) Genome-wide association study of copy number variants suggests LTBP1 and FGD4 are important for alcohol drinking. PLoS One 7:e30860.

Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, Farrer L, Kranzler HR (2005) Diagnostic reliability of the semi-structured assessment for drug dependence and alcoholism (SSADDA). Drug Alcohol Depend 80:303–312.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.

Rios P, Li X, Kohn M (2013) Molecular mechanisms of the PRL phosphatases. FEBS J 280:505–524.

Song I, Volynski K, Brenner T, Ushkaryov Y, Walker M, Semyanov A (2013) Different transporter systems regulate extracellular GABA from vesicular and non-vesicular sources. Front Cell Neurosci 7:23.

Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, Deloukas P, Dermitzakis ET (2005) Genome-wide associations of gene expression variation in humans. PLoS Genet 1:e78.

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 (2009) Genetical genomic determinants of alcohol consumption in rats and humans. BMC Biol 7:70.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515.

Viel G, Boscolo-Berto R, Cecchetto G, Fais P, Nalesso A, Ferrara SD (2012) Phosphatidylethanol in blood as a marker of chronic alcohol use: a systematic review and meta-analysis. Int J Mol Sci 13:14788–14812.

Wang KS, Liu X, Zhang Q, Pan Y, Aragam N, Zeng M (2011) A meta-analysis of two genome-wide association studies identifies 3 new loci for alcohol dependence. J Psychiatr Res 45:1419–1425.

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191.

Yang TT, Wang SJ (2009) Facilitation of glutamate release from rat cerebrocortical glutamatergic nerve terminals (synaptosomes) by phosphatidylserine and phosphatidylcholine. Synapse 63:215–223.

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 (2012) Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. Neuropsychopharmacology 37:557–566.

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 (2009) Population admixture modulates risk for alcohol dependence. Hum Genet 125:605–613.

Zuo L, Wang F, Zhang XY, Li CSR, Lu L, Ye L, Zhang H, Krystal JH, Deng HW, Luo X (2013a) Genome-wide significant association signals in IPO11-HTR1A region specific for alcohol and nicotine co-dependence. Alcohol Clin Exp Res 37:730–739.

Zuo L, Wang K, Zhang X, Krystal JH, Li CR, Zhang F, Zhang H, Luo X (2013b) NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk region specific for alcohol dependence in subjects of European descent. Drug Alcohol Depend 129:254–264.

Zuo L, Wang KS, Zhang XY, Li CS, Zhang F, Wang X, Chen W, Gao G, Zhang H, Krystal JH, Luo X (2013c) Rare SERINC2 variants are specific for alcohol dependence in individuals of European descent. Pharmacogenet Genomics 23:395–402.

Zuo L, Zhang CK, Wang F, Li CS, Zhao H, Lu L, Zhang XY, Lu L, Zhang H, Zhang F, Krystal JH, Luo X (2011) A novel, functional and replicable risk gene region for alcohol dependence identified by genome-wide association study. PLoS One 6:e26726.

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 (2013d) Rare ADH variant constellations are specific for alcohol dependence. Alcohol Alcohol (Oxford, Oxfordshire) 48:9–14.