Rare *SERINC2* variants are specific for alcohol dependence in individuals of European descent

Lingjun Zuo[a], Ke-Sheng Wang[e], Xiang-Yang Zhang[f], Chiang-Shan R. Li[a], Fengyu Zhang[g], Xiaoping Wang[i], Wenan Chen[h], Guimin Gao[h], Heping Zhang[b], John H. Krystal[a,c,d] and Xingguang Luo[a]

Objectives **We have previously reported a top-ranked risk gene [i.e., serine incorporator 2 gene (*SERINC2*)] for alcohol dependence in individuals of European descent by analyzing the common variants in a genome-wide association study. In the present study, we comprehensively examined the rare variants [minor allele frequency (MAF)$<$0.05] in the *NKAIN1-SERINC2* region to confirm our previous finding.**

Materials and methods **A discovery sample (1409 European-American patients with alcohol dependence and 1518 European-American controls) and a replication sample (6438 European-Australian family participants with 1645 alcohol-dependent probands) were subjected to an association analysis. A total of 39 903 individuals from 19 other cohorts with 11 different neuropsychiatric and neurological disorders served as contrast groups. The entire *NKAIN1-SERINC2* region was imputed in all cohorts using the same reference panels of genotypes that included rare variants from the whole-genome sequencing data. We stringently cleaned the phenotype and genotype data, and obtained a total of about 220 single-nucleotide polymorphisms in individuals of European descent and about 450 single-nucleotide polymorphisms in the individuals of African descent with $0<MAF<0.05$ for an association analysis.**

Results **Using a weighted regression analysis implemented in the program SCORE-Seq, we found a rare variant constellation across the entire *NKAIN1-SERINC2* region that was associated with alcohol dependence in European-Americans (Fp: overall, $P=1.8\times10^{-4}$; VT: overall, $P=1.4\times10^{-4}$; Collapsing, $P=6.5\times10^{-5}$) and European-Australians (Fp: overall, $P=0.028$; Collapsing, $P=0.025$), but not in African-Americans, and not associated with any other disorder examined. Association signals in this region came mainly from *SERINC2*, a gene that codes for an activity-regulated protein expressed in the brain that incorporates serine into lipids. In addition, 26 individual rare variants were nominally associated with alcohol dependence in European-Americans ($P<0.05$). The associations of five of these rare variants that lay within *SERINC2* showed region-wide significance ($P<\alpha=0.0006$) and 25 associations survived correction for a false discovery rate ($q<0.05$). The associations of two rare variants at *SERINC2* were replicated in European-Australians ($P<0.05$).**

Conclusion **We concluded that *SERINC2* was a replicable and significant risk gene specific for alcohol dependence in individuals of European descent.** ***Pharmacogenetics and Genomics*** **23:395–402** © **2013 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

Pharmacogenetics and Genomics 2013, **23**:395–402

Keywords: alcohol dependence, association, European descent, rare variant constellations, *SERINC2*

Departments of [a]Psychiatry, Yale University School of Medicine, [b]Biostatistics, Yale University School of Epidemiology and Public Health, [c]Psychiatry Services, Yale-New Haven Hospital, New Haven, [d]VA Alcohol Research Center, VA Connecticut Healthcare System, West Haven, Connecticut, [e]Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, Johnson City, Tennessee, [f]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas, [g]Division of Clinical Sciences, Lieber Institute for Brain Development, Johns Hopkins University Medical Campus, Baltimore, Maryland, [h]Department of Biostatistics, Virginia Commonwealth University, Richmond, Virginia, USA and [i]Department of Neurology, First People's Hospital, Shanghai Jiaotong University, Shanghai, China

Correspondence to Xingguang Luo, MD, PhD, Department of Psychiatry, Yale University School of Medicine, West Haven, CT 06516, USA
Tel: +1 203 932 5711 x5745; fax: +1 203 937 4741;
e-mail: xingguang.luo@yale.edu

Received 1 October 2012 **Accepted** 17 April 2013

Introduction

Individuals with alcohol dependence continue to use alcohol despite adverse consequences on health, job, and family functions. Several lines of evidence have shown a major genetic component in the risk of developing alcohol dependence. Siblings of alcoholic probands had a three to eight-fold increase in the risk of developing alcohol dependence [1]. The heritability of risk for alcohol dependence was estimated to be ~39% by studies of the adopted-away offspring of affected and unaffected parents [2] and as high as 60% by twin studies [3]. These studies provided evidence that genetic factors constitute a significant cause of alcohol dependence.

Recently, a gene at 6q21, that is Na^+/K^+ transporting ATPase interacting 2 gene (*NKAIN2*) (alias: *TCBA1*), was found to have replicable associations with alcohol dependence in both the COGA family-based Caucasian ($P<10^{-3}$) and the European-Australian samples ($P=5.1\times10^{-7}$) [4]. This gene is highly conserved among species, transcribed

1744-6872 © 2013 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/FPC.0b013e328362f9f2

in different splice variants, specific to the central nervous system [5], and critical for neuronal functions [6]. It has also been associated with neuroticism in a genome-wide associations study (GWAS) [7], a complex neurological phenotype [5], and a developmental delay with recurrent infections [8]. This gene encodes a member protein, that is NKAIN2, of a mammalian protein family that includes four members, that is NKAINs 1–4, with similar conservation, distributions, and functions. All four proteins interact with the beta subunit of Na, K-ATPase (ATP1B1) that belongs to the family of Na^+/K^+ and H^+/K^+ ATPases β chain proteins and to the subfamily of Na^+/K^+-ATPases. Na^+/K^+-ATPase is an integral membrane protein responsible for establishing and maintaining the electrochemical gradients of Na and K ions across the plasma membrane. These electrochemical gradients are essential for osmoregulation, sodium-coupled transport of a variety of organic and inorganic molecules, and electrical excitability of nerve and muscle.

Interestingly, the *NKAIN1* gene at 1p35.2, which encodes the member protein NKAIN1 from the same family as NKAIN2, is closely located to the serine incorporator 2 gene (*SERINC2*) that was a top-ranked risk gene ($P = 2.3 \times 10^{-7}$) for alcohol dependence reported by one of our recent GWASs [9]. In the present study, we comprehensively examined the rare variants [minor allele frequency (MAF) < 0.05] in the *NKAIN1-SERINC2* region to confirm our previous GWAS finding that was based on common variant analysis (MAF > 0.05).

The variants with low allele frequencies are proposed to be the key for 'missing' heritability. An increasing number of human diseases have been found to be caused by a constellation of multiple rare, regionally concentrated, variants. Some association signals credited to common variants may be synthetic associations resulting from the contributions of multiple rare variants within a genomic region [10]. In these cases, the synthetic effects of region-wide rare variant constellations may be more significant than individual rare variants. In this study, on the basis of our prior work, we comprehensively examined the associations between rare *NKAIN1-SERINC2* variants (MAF < 0.05) and alcohol dependence in a European-American discovery cohort and a European-Australian replication cohort.

It has been reported that alcohol dependence has a high rate of comorbidity with numerous neuropsychiatric conditions including anxiety disorders, major depression, bipolar disorders, schizophrenia, and post-traumatic stress disorder [11–13]. It has also been reported that many genes have pleiotropic effects on alcohol dependence and other neuropsychiatric conditions; for example, the autism susceptibility candidate gene 2 (*AUTS2*) was reported to be a risk gene for autism [14], alcohol consumption (by a GWAS) [15], mental retardation [16], and heroin dependence [17]. It is known that alcohol dependence and these other neuropsychiatric conditions shared etiologies that implicate monoaminergic, cholinergic, GABAergic, and glutamatergic neurotransmission. The *NKAIN1-SERINC2* region might be related to these neurotransmission systems [18]. Thus, in this study, we also examined the associations between the *NKAIN1-SERINC2* variants and several other neuropsychiatric and neurological disorders available from the dbGaP database to examine whether the *NKAIN1-SERINC2* variants are specific to alcohol dependence.

Materials and methods

Participants

The discovery cohort included 1409 European-American patients with alcohol dependence [*Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (DSM-IV)] (38.3 ± 10.2 years) and 1518 European-American controls (38.4 ± 10.4 years). The replication cohort included 2252 European-Australian small nuclear families with 2 generations or less. These families had a total of 6438 individuals including 4342 founders and 2096 nonfounders; the latter included 1645 alcohol-dependent probands. The average family size was 2.97 individuals. These families included 3818 sib-pairs, 20 half-sib-pairs, and 4192 parent–child pairs.

The discovery cohort came from the merged SAGE and COGA datasets (dbGaP access number: phs000092.v1.p1 and phs000125.v1.p1) and the replication cohort came from the OZ-ALC dataset (dbGaP access number: phs000181.v1.p1). Affected patients fulfilled the lifetime DSM-IV criteria for alcohol dependence [19]. The control participants were defined as individuals who had been exposed to alcohol (and possibly to other drugs), but never fulfilled the criteria for alcohol or substance use disorder (lifetime diagnosis). All participants were interviewed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) [20]. A total of 39 903 individuals of European or African descent from 19 other dbGaP cohorts with 11 different neuropsychiatric and neurological disorders served as the contrast groups. These different neuropsychiatric and neurological disorders included alcohol dependence (in African-Americans), major depression, bipolar disorder, schizophrenia, autism, attention deficit hyperactivity disorder, Alzheimer's disease, amyotrophic lateral sclerosis, early-onset stroke, ischemic stroke, and Parkinson's disease. Diagnoses, ethnicities, study designs, and dataset names for these cohorts are shown in Table 1. More detailed demographic information for these samples including dbGaP accession numbers, genotyping platforms, sample sizes, sex, and age structures, and mean ages has been published previously [26–29]. These 21 cohorts included case–control and family-based samples, genotyped on Illumina (Illumina Inc., San Diego, California, USA), Affymetrix (Affymetrix Inc., Santa Clara, California, USA), or Perlegen microarray platforms (Perlegen Sciences Inc., Mountain View, California, USA). The discovery

Table 1 **Associations between the *NKAIN1-SERINC2* region and different neuropsychiatric or neurological disorders**

Human diseases	Ethnicity	Dataset name	SNP # (total)	SNP # ($P<0.05$)	SNP # ($P<\alpha$)	SNP # ($q<0.05$)	Collapsing *P*-value	Minimal *P*-value	Most significant SNP	Gene	Minor allele frequency (*N*)	
											Affected	Unaffected
Alcoholism	EA (CC)	SAGE+COGA	196	26	5	25	6.5×10^{-5}	4.1×10^{-5}	rs35961897	*SERINC2*	0.059 (1409)	0.038 (1518)
Alcoholism	EAu (Fam)	OZ-ALC	185	9	0	0	0.025	0.0196	rs77840364	*SERINC2*	0.015 (1645)	0.012 (4793)
Alcoholism	AA (CC)	SAGE+COGA	450	4	0	0	0.534	0.0141	rs16834507	*SERINC2*	0.040 (681)	0.021 (508)
ADHD	CA (Fam)	IMAGE	163	12	0	0	0.429	0.0017	rs114467377	*NKAIN1*	0.015 (924)	0.014 (1833)
Autism	EA (Fam)	AGP	189	13	0	0	0.977	0.0013	rs114336824	*SERINC2*	0.010 (1330)	0.010 (2745)
Major depression	CA (CC)	PRSC	162	12	0	0	0.089	0.0010	rs116080631	*SERINC2*	0.028 (1805)	0.015 (1820)
Bipolar disorder	EA (CC)	BDO+GRU	136	0	0	0	0.735	0.0902	rs7541681	*SERINC2*	0.125 (368)	0.036 (1034)
Bipolar disorder	EA (CC)	BARD+GRU	138	5	0	0	0.714	0.0048	rs55781513	*NKAIN1*	0.026 (653)	0.044 (1034)
Bipolar disorder	AA (CC)	BARD+GRU	351	7	0	0	0.312	0.0033	rs114478713	*SERINC2*	0.019 (141)	0.001 (671)
Schizophrenia	EA (CC)	GAIN	180	1	0	0	0.692	0.0080	rs6659255	*SERINC2*	0.073 (1351)	0.046 (1378)
Schizophrenia	AA (CC)	GAIN	441	20	0	0	0.816	0.0147	rs80029070	*NKAIN1*	0.025 (1195)	0.038 (954)
Schizophrenia	EA (CC)	MGS_nonGAIN	144	12	0	0	0.145	0.0024	rs74872508	*SNRNP40*	0.009 (1437)	0.003 (1347)
Alzheimer's disease	CA (Fam)	LOAD × 4	191	16	0	0	0.556	0.0057	rs7417775	*SERINC2*	0.052 (2298)	0.037 (2921)
Alzheimer's disease	EA (CC)	GenADA	113	10	0	0	0.514	0.0064	rs76859788	*NKAIN1*	0.012 (806)	0.025 (782)
ALS	CA (CC)	GRU	125	9	0	0	0.111	0.0111	rs12024466	*ZCCHC17*	0.004 (261)	0.025 (246)
Early-onset stroke	EA (CC)	GEOS × 3	144	1	0	0	0.246	0.0447	rs13376139	*SNRNP40*	0.008 (372)	0.034 (430)
Early-onset stroke	AA (CC)	GEOS × 3	431	54	0	0	0.080	0.0008	rs56095638	*SERINC2*	0.087 (309)	0.038 (290)
Ischemic stroke	CA (CC)	ISGS	132	11	0	0	0.512	0.0041	rs116007405	*SERINC2*	0.028 (219)	0.004 (266)
Parkinson's disease	CA (CC)	NGRC	187	3	0	0	0.745	0.0348	rs114215404	*NKAIN1*	0.002 (2000)	0.005 (1986)
Parkinson's disease	CA (CC)	PDRD+GRU	142	0	0	0	0.088	0.0694	rs75239059	*NKAIN1*	0.003 (900)	0.008 (867)
Parkinson's disease	CA (CC)	lng_coriell_pd	171	5	0	0	0.079	0.0195	rs12564915	*SNRNP40*	0.015 (940)	0.039 (801)

Only the most significant risk markers with minimal *P* values are listed.

Dataset names refer to dbGaP and references (GenADA: Li *et al.* [21], Filippini *et al.* [22]; AGP: The AGP Consortium [23–25]). The significance level (α) is corrected for the numbers of effective genetic markers (calculated by SNPSpD). Collapsing *P* values for the entire *NKAIN1-SERINC2* region were calculated using the program ARIEL.

AA, African-American; ADHD, attention deficit hyperactivity disorder; ALS, amyotrophic lateral sclerosis; CA, Caucasian; CC, case–control design; EA, European-American; EAu, European-Australian; Fam, family-based design; *N*, sample size; SNP, single nucleotide polymorphism.

sample was genotyped on the Illumina Human 1M (Illumina Inc.) and the replication sample was genotyped on Illumina Human CNV370v1 (Illumina Inc.).

Imputation

The *NKAIN1-SERINC2* region includes *NKAIN1*, *SNRNP40*, *ZCCHC17*, *FABP3*, and *SERINC2*. We imputed the missing single-nucleotide polymorphisms (SNPs) across the entire *NKAIN1-SERINC2* region from Chr1:31,425,179 to Chr1:31,732,987 using the same reference panels to render the genetic marker sets highly consistent across different cohorts. The reference CEU panel YRI panels from two databases (i.e., 1000 genome project and HapMap 3) were adopted in the imputation for the samples of European descent and African descent, respectively. All cohorts were imputed using both programs IMPUTE2 [30] and BEAGLE [31]. We maximized the success rate and accuracy of imputation and minimized the false-positives during imputation. Only the genotypes that were imputed consistently between the two independent reference databases and consistently both by IMPUTE2 and BEAGLE were selected for analysis. The uncertainty rate of inference for missing genotypes was controlled at less than 1%. 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. On the basis of these selection criteria, we were highly confident in the quality of these imputed genotype data. On checking the imputed genotypes in all of our four family-based cohorts, we did not find any one individual (considering all SNPs tested) or any one SNP (considering all individuals tested) with more than 0.1% Mendelian inconsistency.

Data cleaning

We stringently cleaned the phenotype and genotype data within each ethnicity before association analysis. Individuals with poor genotypic data, individuals with allele discordance, sample relatedness, individuals with a mismatch between self-identified and genetically inferred ethnicity, and individuals with a missing genotype call rate of at least 2% across all SNPs were excluded. Furthermore, we filtered out the monomorphic SNPs and the SNPs with allele discordance, Mendelian errors (in family samples), an overall missing genotype call rate of at least 2%, and MAFs greater than 0.05. We also filtered out the SNPs with missing rate differences greater than 2% between two samples that had the same phenotype and microarray platform. As a result, a total of $\sim$220 (in the individuals of European descent) and $\sim$450 (in the individuals of African descent) SNPs with $0 < \text{MAF} < 0.05$ in either cases or controls were extracted for association analysis. The cleaned sample sizes and SNP numbers of all cohorts are shown in Table 1.

Association tests for region-wide rare variant constellations

We initially tested associations between rare variant constellations and alcohol dependence using a score-type program, SCORE-Seq [32]. The mutation information was aggregated by a weighted linear combination across all rare variants of the entire *NKAIN1-SERINC2* region or across each gene within the *NKAIN1-SERINC2* region, and then related to alcohol dependence using regression models. Sex, age, and the first 10 principal components served as the covariates in the regression models. The principal component scores of our samples were derived from all common autosomal SNPs across the genome using principal component analysis implemented in the software package EIGENSOFT [33]. Each individual received scores on each principal component. These principal components reflected the population structure of our samples, with the first 10 principal component scores accounting for more than 95% of variance. As covariates in the regression model, these principal components controlled for population stratification and admixture effects on association analysis. The same analytic procedures were used for the other 10 neuropsychiatric and neurological disorders. For the association analyses on major depression, bipolar disorder, schizophrenia, and Alzheimer's disease, 'alcohol drinking' was also included as an additional covariate in the regression to control for the potential confounding effects of 'alcohol drinking behaviors'. This covariate was not assessed in attention deficit hyperactivity disorder, autism, amyotrophic lateral sclerosis, early-onset stroke, ischemic stroke, and Parkinson's disease.

Two tests, Fp and VT, were used to derive the overall P values. (a) In the Fp tests, the MAF upper bound threshold was fixed at 0.05, but the weight was $1/\text{sqrt}(P(1-P))$, where P was the estimated MAF with pseudo counts in the pooled sample. (b) In the VT test, the weight was fixed at 1, whereas the threshold varied between 0 and 0.05. Statistical significance was assessed using a bootstrap procedure with one million times of resampling. Finally, these tests were confirmed using another program ARIEL [34] that used a regression-based collapsing approach.

Association tests for individual rare variants (exploratory)

For case–control samples in the discovery cohort, the allele frequencies of each SNP were compared between cases and controls using logistic regression as implemented in PLINK [35]. Diagnosis and alleles each served as the dependent and independent variables, with sex, age, and the first 10 principal components as covariates. For family samples in the replication cohort, we tested the allele–disease associations using the program FBAT [36], assuming an additive genetic model under the null hypothesis of no linkage and association, biallelic mode,

minimum number of informative families of 10 for each analysis, and offset of zero. The same analytic procedures were used for the other 10 neuropsychiatric and neurological disorders, with 'alcohol drinking', if available, as an additional covariate in the models. Different cohorts were analyzed independently. The MAFs and the minimal P values of the most significant risk SNPs and the numbers of the nominally significant risk SNPs ($P < 0.05$) in all cohorts are shown in Table 1. Finally, the cumulative positive predictive values (PV^+) and the cumulative positive likelihood ratios (LR^+) of the significant ($P < \alpha$) and independent ($r^2 < 0.2$) risk alleles across the *NKAIN1-SERINC2* region were calculated using the Bayesian formula.

Correction for multiple testing in single-point association tests

The experiment-wide significance levels (α) were corrected for the numbers of effective markers that were calculated using the Bonferroni-type program SNPSpD [37], accounting for linkage disequilibrium. Approximately 80 and 120 effective SNPs captured most of the information content of all rare variants across the entire *NKAIN1-SERINC2* region in individuals of European and African descent, respectively. Thus, the corrected significance levels (α) for single-point association tests were set at 0.0006 in individuals of European descent and 0.0004 in individuals of African descent, respectively. The numbers of the statistically significant (i.e., $P < \alpha$) risk SNPs in all cohorts are shown in Table 1. The false discovery rate (q-value) for each SNP was estimated from the P values within each disease group using the R package QVALUE [38]. Finally, for those associations replicated in the European-Australian cohort, the α was set at 0.05.

Results

A rare variant constellation across the entire *NKAIN1-SERINC2* region was associated with alcohol dependence in European-Americans (Fp: overall, $P = 1.8 \times 10^{-4}$; VT: overall, $P = 1.4 \times 10^{-4}$; Collapsing, $P = 6.5 \times 10^{-5}$) and European-Australians (Fp: overall, $P = 0.028$; Collapsing, $P = 0.025$), but not in African-Americans, and not associated with any other disease examined (Collapsing, $P > 0.05$). When the rare variant constellation within each gene region was tested, the *SERINC2* variant constellation was significantly associated with alcohol dependence in European-Americans (Fp: $P = 2.7 \times 10^{-4}$; VT: $P = 1.6 \times 10^{-4}$; Collapsing, $P = 8.0 \times 10^{-5}$) and suggestively in European-Australians (VT: $P = 0.028$; Collapsing, $P = 0.030$) (corrected $\alpha = 0.01$ for five genes within *NKAIN1-SERINC2*). The other four genes were suggestively (i.e., P close to 0.05) or modestly ($0.01 < P < 0.05$) associated with alcohol dependence in European-Americans (Tables 1 and 2).

Single-point association analysis showed that, among a total of 196 individual rare variants in European-Americans, 26 SNPs were nominally associated with alcohol dependence ($P < 0.05$). Twenty-five associations survived correction for false discovery rate ($q < 0.05$) and the associations of five *SERINC2* variants survived Bonferroni's correction ($P < \alpha = 0.0006$) (Table 1). If further corrected by the number of cohorts examined (i.e., $n = 21$), two variants (i.e., rs35961897 and rs4949405) of these five SNPs remained suggestively significant ($\alpha = 2.9 \times 10^{-5}$). Two independent SNPs, that is rs34278290 and rs7417775, can tag these five variants (Table 3). The cumulative PV^+ of these two markers was 0.0791 when we used 3.81% as the 1-year prevalence rate of alcohol dependence; it was 0.2366 when we used 12.5% as the lifetime prevalence rate of alcohol dependence. Furthermore, the cumulative LR^+ of these

Table 2 ***P*-values for associations between rare variant constellations and alcohol dependence**

	European-Americans						Associations	
Tests	Whole region	*NKAIN1*	*SNRNP40*	*ZCCHC17*	*FABP3*	*SERINC2*	Whole region	*SERINC2*
Fp	1.8×10^{-4}	0.039	0.020	0.025	0.041	2.7×10^{-4}	0.028	0.028
VT	1.4×10^{-4}	–	0.022	0.022	0.066	1.6×10^{-4}	–	–
Collapsing	6.5×10^{-5}	0.088	4.7×10^{-3}	8.0×10^{-3}	0.045	8.0×10^{-5}	0.025	0.030

Collapsing, association test using ARIEL; Fp and VT, association tests using SCORE-Seq.

Table 3 Top-ranked risk SNPs ($P < \alpha$) for alcohol dependence in *SERINC2* in European-Americans

			MAF		European-Americans		
SNP	Gene	Location	Cases	Controls	OR	*P*-value	*q*-value
rs34278290▲	*SERINC2*	Intron 2	0.061	0.044	1.52	5.9×10^{-4}	0.0043
rs2275437	*SERINC2*	Exon 7	0.057	0.039	1.60	2.0×10^{-4}	0.0025
rs35961897	*SERINC2*	Intron 10	0.059	0.038	1.69	4.1×10^{-5}	0.0008
rs4949405	*SERINC2*	Intron 12	0.050	0.030	1.79	4.2×10^{-5}	0.0008
rs7417775●#	*SERINC2*	3′ UTR	0.033	0.021	1.79	5.3×10^{-4}	0.0043

MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.
▲, located in transcription factor-binding sites; ●, affecting miRNA-binding site activity; #, with significant *cis*-acting regulatory effects on *SERINC2* mRNA expression ($P = 0.024$).

Table 4 Replicable risk SNPs ($P<0.05$) for alcohol dependence in *SERINC2* in individuals of European descent

			MAF		European-Americans			Australians		Meta-analysis	
SNP	Gene	Location	Cases	Controls	OR	*P*-value	*q*-value	OR	*P*-value	*Z*-score	*P*-value
rs77840364	*SERINC2*	Intron 10	0.030	0.022	1.45	0.029	0.0433	8.00	0.020	2.348	0.0189
rs115360541	*SERINC2*	Intron 12	0.065	0.043	1.50	0.008	0.0256	2.00	0.019	2.815	0.0049

MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.

two markers was 2.169. Two other associations at *SERINC2* were replicable ($P < 0.05$) between European-Americans and European-Australians (Table 4). Among these SNPs, rs7417775 at the 3′-UTR of *SERINC2* had significant *cis*-acting regulatory effects on *SERINC2* mRNA expression ($P = 0.024$; $n = 45$ unrelated HapMap individuals) [39], and was predicted to affect miRNA-binding site activity; this SNP was also the most significant risk variant for Alzheimer's disease in one Caucasian sample (Tables 1 and 3). In addition, rs34278290 at intron 2 of *SERINC2* was located in a transcript factor-binding site. Finally, no significant individual rare variant was associated with any other disease including alcohol dependence in African-Americans ($P > \alpha$), although this African-American cohort had a 66% power to detect the most significant risk variant, that is rs35961897.

Furthermore, we carried out transcriptome-wide mRNA expression correlation analysis in 93 European brain tissues and 80 European peripheral blood mononuclear cell samples [40]. The expression of *NKAIN1-SERINC2* transcripts was correlated significantly with the expression of numerous alcoholism-related genes, mostly from the dopaminergic, serotoninergic, cholinergic, GABAergic, glutamatergic, histaminergic, endocannabinoid, metabolic, neuropeptide, and opioidergic systems [18].

Discussion

Mainly from the association tests for the region-wide rare variant constellations, we drew the conclusion that *SERINC2* was a replicable and significant risk gene specific for alcohol dependence in individuals of European descent. Results from individual variant analysis supported this conclusion. Rare *SERINC2* variants may lead to a small increase in the risk for alcohol dependence on the basis of their cumulative PV^+ and LR^+. On the basis of these results, we postulate that *SERINC2* may harbor a causal variant(s) for alcohol dependence. Our study provides an additional example to support the hypothesis that region-wide rare variant constellations could have significant synthetic effects on disease phenotypes, even though the effects of individual variants might be not significant. Rare variant constellation analysis is an important tool in genetic association studies.

SERINC2 encodes serine incorporator 2 (Serinc2). Serinc2 is highly expressed in neurons of the hippocampus and cerebral cortex [41]. It is an effector in endoplasmic reticulum membranes that incorporates serine into membranes and facilitates the synthesis of phosphatidylserine and sphingolipids [42]. Phosphatidylserine is specifically distributed in the brain. Consumption of phosphatidylserine supplement has been reported to reduce the risk of dementia and cognitive dysfunction in the elderly [43,44], and thus has been used to treat memory deficit disorders such as Alzheimer's disease and other forms of dementia, to support cognitive functions during aging, and to remediate cognitive deficits as a result of heavy drinking and cigarette smoking. Also specifically expressed in the brain [42], sphingolipids play a functional role in neural plasticity, signaling, and axonal guidance [45–47]. Activity of the sphingolipid metabolism enzyme, that is acid sphingomyelinase, has been reported to be increased in alcohol-dependent patients [48]. Alcohol consumption can increase sphingosine levels in the rat brains [49]. In addition, there are numerous functional variants in *SERINC2* including rare variants (Table 3), common variants, and frameshift variants such as Indels and CNVs (see NCBI dbSNPs). The function of Serinc2 altered by the alleles of these functional *SERINC2* variants may be implicated in the synthesis of phosphatidylserine and sphingolipids and may thus be relevant for the development of alcohol dependence. Alternatively, correlation between the expression of *NKAIN1-SERINC2* transcripts and other genes suggested that *NKAIN1-SERINC2* may contribute toward alcohol dependence through other neurotransmitter or metabolic pathways [18]. For example, the glutaminergic pathway is known to play important roles in alcohol intoxication and withdrawal [50]. Within the hippocampus, Serinc2 expression is increased following seizures induced by kainite, a glutamate agonist [42]. A drug that blocks kainite glutamate receptor function appears to decrease drinking [51]. This evidence supports the glutaminergic pathway hypothesis underlying the connection between Serinc2 and alcohol dependence.

A few limitations need to be considered in the current study. The imputed genotypes were not observed directly from molecular experiments, even though their error rates and uncertainty were very low. Future work is warranted to verify these results by directly sequencing the samples. In addition, because not all neuropsychiatric and neurological disorders comorbid with alcohol dependence were exhaustively examined in the present study, we could not completely exclude the possibility that the other neuropsychiatric and neurological disorders not examined might share this *SERINC2* risk gene with

alcohol dependence. Furthermore, 'alcohol drinking behaviors' were not assessed in some of the neurological disorders in the present study. Their potential confounding effects on the association analysis of these disorders need to be assessed. Finally, more independent cohorts with alcohol dependence to replicate our findings in the future are also warranted.

Acknowledgements

The authors thank Dr Malison for helpful comments. This work was supported in part by National Institute on Drug Abuse (NIDA) Grants K01 DA029643 and R01DA016750, National Institute on Alcohol Abuse and Alcoholism (NIAAA) Grants R01 AA016015 and R21 AA020319, and the National Alliance for Research on Schizophrenia and Depression (NARSAD) Award 17616 (L.Z.). They thank NIH GWAS Data Repository, the Contributing Investigator(s) who contributed the phenotype and genotype data from his/her original study (e.g. Drs Bierut, Edenberg, Heath, Singleton, Hardy, Foroud, Myers, Gejman, Faraone, Sonuga-Barke, Sullivan, Nurnberger, Devlin, Monaco, etc.), and the primary funding organization that supported the contributing study. Funding and other supports for phenotype and genotype data were provided through the National Institutes of Health (NIH) Genes, Environment and Health Initiative (GEI) (U01HG004422, U01HG004436, and U01HG004438); the GENEVA Coordinating Center (U01HG004446); the NIAAA (U10AA008401, R01AA0-13320, and P60AA011998); the NIDA (R01DA013423); the National Cancer Institute (P01 CA089392); the Division of Neuroscience, the NIA National Institute of Neurological Disorders and Stroke (NINDS); the NINDS Human Genetics Resource Center DNA and Cell Line Repository; the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease' (HHSN268200782096C); the Center for Inherited Disease Research (CIDR); a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and Prevention; the NIH Office of Research on Women's Health (ORWH) (R01NS45012); the Department of Veterans Affairs; the University of Maryland General Clinical Research Center (M01RR165001), the National Center for Research Resources, NIH; the National Institute of Mental Health (R01MH059160, R01MH59565, R01MH59566, R01MH-59571, R01MH59586, R01MH59587, R01MH59588, R01MH60870, R01MH60879, R01MH61675, R01MH-62873, R01MH081803, R01MH67257, R01MH81800, U01MH46276, U01MH46282, U01MH46289, U01MH-46318, U01MH79469, U01MH79470, and R01MH67257); the NIMH Genetics Initiative for Bipolar Disorder; the Genetic Association Information Network (GAIN); the Genetic Consortium for Late Onset Alzheimer's Disease; the Autism Genome Project, the MARC: Risk Mechanisms in Alcoholism and Comorbidity; the Molecular Genetics of Schizophrenia Collaboration; the Medical Research Council (G0601030) and the Wellcome Trust (075491/Z/04), University of Oxford; the Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, NWO Genomics, SPI 56-464-1419) the Centre for Neurogenomics and Cognitive Research (CNCR-VU); Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR); and the European Union (EU/WLRT-2001-01254), ZonMW (geestkracht program, 10-000-1002). Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Genetic Consortium for Late Onset Alzheimer's Disease, the GENEVA Coordinating Center (U01 HG004446), and the National Center for Biotechnology Information. Genotyping was performed at the Johns Hopkins University Center for Inherited Disease Research (CIDR), and GlaxoSmithKline, R&D Limited. The datasets used for the analyses described in this manuscript were obtained from dbGaP (*http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap*). The dbGaP accession numbers include phs000125.v1.p1, phs000021.v3.p2, phs000021.v3.p2, phs000167.v1.p1, phs000167.v1.p1, phs000267.v1.p1, phs000016.v2.p2, phs000092.v1.p1, phs000092.v1.p1, phs000181.v1.p1, phs000020.v2.p1, phs000017.v3.p1, phs000017.v3.p1, phs000017.v3.p1, phs000168.v1.p1, phs000219.v1.p1, phs000101.v3.p1, phs000292.v1.p1, phs000292.v1.p1, phs000102.v1.p1, phs000196.v2.p1, phs000126.v1.p1, phs000089.v3.p2, phs000089.v3.p2, phs000089.v3.p2, and phs000089.v3.p2.

Conflicts of interest

There are no conflicts of interest.

References

1 Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, *et al.* Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet* 1998; **81**:207–215.

2 Cloninger CR, Bohman M, Sigvardsson S. Inheritance of alcohol abuse. Cross-fostering analysis of adopted men. *Arch Gen Psychiatry* 1981; **38**:861–868.

3 Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, *et al.* Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychol Med* 1997; **27**:1381–1396.

4 Wang KS, Liu X, Aragam N, Jian X, Mullersman JE, Liu Y, *et al.* Family-based association analysis of alcohol dependence in the COGA sample and replication in the Australian twin-family study. *J Neural Transm* 2011; **118**:1293–1299.

5 Bocciardi R, Giorda R, Marigo V, Zordan P, Montanaro D, Gimelli S, *et al.* Molecular characterization of a t(2;6) balanced translocation that is associated with a complex phenotype and leads to truncation of the TCBA1 gene. *Hum Mutat* 2005; **26**:426–436.

6 Gorokhova S, Bibert S, Geering K, Heintz N. A novel family of transmembrane proteins interacting with beta subunits of the Na,K-ATPase. *Hum Mol Genet* 2007; **16**:2394–2410.

7 Calboli FC, Tozzi F, Galwey NW, Antoniades A, Mooser V, Preisig M, *et al.* A genome-wide association study of neuroticism in a population-based sample. *PLoS One* 2010; **5**:e11504.

8 Yue Y, Stout K, Grossmann B, Zechner U, Brinckmann A, White C, *et al.* Disruption of TCBA1 associated with a de novo t(1;6)(q32.2;q22.3) presenting in a child with developmental delay and recurrent infections. *J Med Genet* 2006; **43**:143–147.

9 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**:557–566.

10 Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010; **8**:e1000294.

11 Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, *et al*. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA* 1990; **264**:2511–2518.

12 Kessler RC, Nelson CB, McGonagle KA, Edlund MJ, Frank RG, Leaf PJ. The epidemiology of co-occurring addictive and mental disorders: implications for prevention and service utilization. *Am J Orthopsychiatry* 1996; **66**:17–31.

13 Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, *et al*. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry* 2004; **61**:807–816.

14 Sultana R, Yu CE, Yu J, Munson J, Chen D, Hua W, *et al*. Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. *Genomics* 2002; **80**:129–134.

15 Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, *et al*. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc Natl Acad Sci USA* 2011; **108**:7119–7124.

16 Kalscheuer VM, FitzPatrick D, Tommerup N, Bugge M, Niebuhr E, Neumann LM, *et al*. Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation. *Hum Genet* 2007; **121**:501–509.

17 Chen YH, Liao DL, Lai CH, Chen CH. Genetic analysis of AUTS2 as a susceptibility gene of heroin dependence. *Drug Alcohol Depend* 2012; **128**:238–242.

18 Zuo L, Wang K, Zhang XY, Krystal JH, Li CS, 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**:254–264.

19 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. Washington, DC: American Psychiatric Press; 1994.

20 Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, *et al*. 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.

21 Li H, Wetten S, Li L St, Jean PL, Upmanyu R, Surh L, *et al*. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 2008; **65**:45–53.

22 Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, Guzman D, *et al*. Anatomically-distinct genetic associations of APOE epsilon4 allele load with regional cortical atrophy in Alzheimer's disease. *Neuroimage* 2009; **44**:724–728.

23 AGP: The AGP Consortium. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007; **39**: 319–328.

24 AGP: The AGP Consortium. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**:368–372.

25 AGP: The AGP Consortium. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010; **19**:4072–4082.

26 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.

27 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. *Biol Psychiatry* 2011; **70**:513–518.

28 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 USA* 2010; **107**:5082–5087.

29 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**:840–852.

30 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.

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

32 Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 2011; **89**: 354–367.

33 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**:904–909.

34 Asimit JL, Day-Williams AG, Morris AP, Zeggini E. ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. *Hum Hered* 2012; **73**:84–94.

35 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**:559–575.

36 Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype–phenotype associations. *Eur J Hum Genet* 2001; **9**:301–306.

37 Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005; **95**:221–227.

38 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**:9440–9445.

39 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.

40 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.

41 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**:447–457.

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

43 Schreiber S, Kampf-Sherf O, Gorfine M, Kelly D, Oppenheim Y, Lerer B. An open trial of plant-source derived phosphatydilserine for treatment of age-related cognitive decline. *Isr J Psychiatry Relat Sci* 2000; **37**:302–307.

44 Jorissen BL, Brouns F, Van Boxtel MP, Ponds RW, Verhey FR, Jolles J, *et al*. The influence of soy-derived phosphatidylserine on cognition in age-associated memory impairment. *Nutr Neurosci* 2001; **4**:121–134.

45 Nguyen PV, Abel T, Kandel ER. Requirement of a critical period of transcription for induction of a late phase of LTP. *Science* 1994; **265**: 1104–1107.

46 Tsui-Pierchala BA, Encinas M, Milbrandt J, Johnson EM Jr. Lipid rafts in neuronal signaling and function. *Trends Neurosci* 2002; **25**:412–417.

47 Guirland C, Suzuki S, Kojima M, Lu B, Zheng JQ. Lipid rafts mediate chemotropic guidance of nerve growth cones. *Neuron* 2004; **42**: 51–62.

48 Reichel M, Beck J, Muhle C, Rotter A, Bleich S, Gulbins E, *et al*. Activity of secretory sphingomyelinase is increased in plasma of alcohol-dependent patients. *Alcohol Clin Exp Res* 2011; **35**:1852–1859.

49 Dasgupta S, Adams JA, Hogan EL. Maternal alcohol consumption increases sphingosine levels in the brains of progeny mice. *Neurochem Res* 2007; **32**:2217–2224.

50 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.

51 Johnson BA, Rosenthal N, Capece JA, Wiegand F, Mao L, Beyers K, *et al*. Topiramate for treating alcohol dependence: a randomized controlled trial. *JAMA* 2007; **298**:1641–1651.