Alcohol and Alcoholism Vol. 48, No. 1, pp. 9–14, 2013
Advance Access Publication 27 September 2012
doi: 10.1093/alcalc/ags104

ORIGINAL ARTICLE

Rare *ADH* Variant Constellations are Specific for Alcohol Dependence

Lingjun Zuo[1], Heping Zhang[2], Robert T. Malison[1], Chiang-Shan R. Li[1], Xiang-Yang Zhang[3], Fei Wang[1], Lingeng Lu[2], Lin Lu[4], Xiaoping Wang[5], John H. Krystal[1,6,7], Fengyu Zhang[8], Hong-Wen Deng[9] and Xingguang Luo[1,*]

[1]Department of Psychiatry, Yale University School of Medicine, West Haven, CT 06516, USA, [2]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA, [3]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA, [4]National Institute on Drug Dependence, Peking University, Beijing, China, [5]Department of Neurology, First People's Hospital, Shanghai Jiaotong University, Shanghai, China, [6]Psychiatry Services, Yale-New Haven Hospital New Haven, CT, USA, [7]VA Alcohol Research Center, VA Connecticut Healthcare System, West Haven, CT, USA, [8]Lieber Institute for Brain Development, Johns Hopkins University Medical Campus, Baltimore, MD, USA and [9]Department of Biostatistics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA

*Corresponding author: Tel.: +1-203-824-8262; Fax: +1-203-937-4741; E-mail: xingguang.luo@yale.edu

(*Received* 10 *April* 2012; *first review notified* 22 *May* 2012; *in revised form* 12 *July* 2012; *accepted* 31 *July* 2012)

Abstract — Aims: Some of the well-known functional alcohol dehydrogenase (ADH) gene variants (e.g. *ADH1B**2, *ADH1B**3 and *ADH1C**2) that significantly affect the risk of alcohol dependence are rare variants in most populations. In the present study, we comprehensively examined the associations between rare *ADH* variants [minor allele frequency (MAF) <0.05] and alcohol dependence, with several other neuropsychiatric and neurological disorders as reference. **Methods:** A total of 49,358 subjects in 22 independent cohorts with 11 different neuropsychiatric and neurological disorders were analyzed, including 3 cohorts with alcohol dependence. The entire *ADH* gene cluster (*ADH7–ADH1C–ADH1B–ADH1A–ADH6–ADH4–ADH5* at Chr4) was imputed in all samples using the same reference panels that included whole-genome sequencing data. We stringently cleaned the phenotype and genotype data to obtain a total of 870 single nucleotide polymorphisms with 0< MAF <0.05 for association analysis. **Results:** We found that a rare variant constellation across the entire *ADH* gene cluster was significantly associated with alcohol dependence in European-Americans (Fp1: simulated global $P = 0.045$), European-Australians (Fp5: global $P = 0.027$; collapsing: $P = 0.038$) and African-Americans (Fp5: global $P = 0.050$; collapsing: $P = 0.038$), but not with any other neuropsychiatric disease. Association signals in this region came principally from *ADH6*, *ADH7*, *ADH1B* and *ADH1C*. In particular, a rare *ADH6* variant constellation showed a replicable association with alcohol dependence across these three independent cohorts. No individual rare variants were statistically significantly associated with any disease examined after group- and region-wide correction for multiple comparisons. **Conclusion:** We conclude that rare *ADH* variants are specific for alcohol dependence. The *ADH* gene cluster may harbor a causal variant(s) for alcohol dependence.

INTRODUCTION

Alcohol dehydrogenases (ADHs) are largely distributed in the liver (e.g. ββADH encoded by *ADH1B*, γγADH encoded by *ADH1C* and ADH6 enzyme encoded by *ADH6*) and the upper digestive tract (e.g. σσADH encoded by *ADH7*) and partly in the central nervous system (e.g. σσADH and ADH6) (Shmueli *et al.*, 2003; Yanai *et al.*, 2005). They possess high activity in converting ethanol to toxic acetaldehyde (Yasunami *et al.*, 1991). Alterations of this activity may influence human drinking behavior and thus the risk of alcohol dependence. Additionally, these enzymes are also efficient in the oxidization of retinol, a vitamin A precursor (summarized in Satre *et al.*, 1994; Luo *et al.*, 2008). For example, σσADH, also called retinol dehydrogenase, is the most efficient enzyme among ADHs in catalyzing retinol formation (Satre *et al.*, 1994); the ADH6 enzyme is efficient in the oxidization of retinol as well (Km:15–40 μM) (Satre *et al.*, 1994). Specifically, they convert retinol to retinal, which in turn is synthesized to retinoic acid (RA), the active form of vitamin A. RA is a pleiotrophic regulator of gene expression in vertebrates and plays a role in regulating embryonic development (including development of the brain). Dopamine neurons contain all necessary enzymatic components for these regulations. Proper development and maintenance of a functional dopaminergic system may depend strongly upon the supply of RA. Functional alterations of these enzymes can thus influence the development and maintenance of physiological dopaminergic system functioning (Luo *et al.*, 2008). In addition to ethanol and retinol, ADH enzymes are also implicated in the metabolism of various dopamine-related neurotransmitters. These support the hypothesis that, in addition to alcohol dependence, there could be associations between *ADH* gene variants and more neuropsychiatric and neurological disorders, given that the dopaminergic system is well known to play an important role in the etiology of those disorders. Furthermore, alcohol dependence has high rates of co-morbidity with numerous psychiatric disorders including anxiety disorders, major depression, bipolar disorders, schizophrenia, post-traumatic stress disorder, etc. (Regier *et al.*, 1990; Kessler *et al.*, 1996; Grant *et al.*, 2004), which also supports the hypothesis that alcohol dependence and other neuropsychiatric disorders could have common susceptibility genes including *ADH* genes. So far, numerous studies have reported associations between *ADH* variants and alcohol dependence; *ADH* variants have also been associated with Parkinson's disease (*ADH1C* and *ADH7*) (Buervenich et al., 2000, 2005), cerebral infarction and lacunae (*ADH1B*) (Suzuki *et al.*, 2004).

It is well known that at least four functional ADH gene variants including rs1229984 (*ADH1B**2; Arg48His), rs2066702 (*ADH1B**3; Arg370Cys), rs1693482 (*ADH1C**2; Arg272Gln) and rs698 (*ADH1C**2; Ile350Va) that significantly affect the risk of alcohol dependence are rare variants in most populations, e.g. in Asians [minor allele frequency (MAF) $f_{rs2066702} = 0.000$; $f_{rs1693482} = 0.023$; $f_{rs698} = 0.025$], Europeans ($f_{rs2066702} = 0.000$; $f_{rs1229984} = 0.008$) and/or Africans ($f_{rs1229984} = 0.000$; $f_{rs1693482} = 0.052$; $f_{rs698} = 0.042$) (Luo *et al.*, 2006). A recent genome-wide association study identified a common variant (rs1789891; $f = 0.192$) that was

significantly associated with alcohol dependence in people of German descent [$P = 1.3 \times 10^{-8}$; odds ratio (OR) = 1.46] (Frank *et al.*, 2012). Notably, this significant risk variant is located between the four functional *ADH* rare variants. These suggest to us that rare *ADH* variants may play important roles in human diseases.

The role of rare genetic variants in human diseases has not been well studied until recently. An important hypothesis in medical genetics research is that many genetically influenced human diseases may not result from a single common variant, but rather, from a constellation of more rare, regionally concentrated, disease-causing variants. The signals of association credited to common genetic variants may be synthetic associations resulting from the contributions of multiple rare variants within a given gene region (Dickson *et al.*, 2010). With the emergence of sequencing technology, it is now feasible to test this hypothesis by thoroughly investigating the rare variants across the genome (e.g. capitalizing on the vast array of rare variant data deposited in databases such as the 1000 Genome Project).

In this study, we aimed to comprehensively examine the associations between rare *ADH* variants (MAF < 0.05) and 11 different neuropsychiatric and neurological disorders in subjects of European or African descent, which included three independent cohorts with alcohol dependence in European-Americans, European-Australians and African-Americans. In these three cohorts, no significant common *ADH* variants for risk of alcohol dependence have been found before (Bierut *et al.*, 2010; Edenberg *et al.*, 2010; Heath *et al.*, 2011). This study would help us to know whether the rare *ADH* variants are specific for alcohol dependence or shared by susceptibility to other disorders.

MATERIALS AND METHODS

Subjects

A total of 49,358 subjects in 22 independent cohorts with 11 different neuropsychiatric and neurological disorders were analyzed (Tables 1 and 2). These 22 cohorts included case–control and family-based samples, genotyped on different microarray platforms. These 11 disorders included alcohol dependence, major depression, bipolar disorder, schizophrenia, autism, attention deficit hyperactivity disorder (ADHD), Alzheimer's disease, amyotrophic lateral sclerosis (ALS), early onset stroke, ischemic stroke and Parkinson's disease. These data were all of those with neuropsychiatric and neurological disorders available for our analysis from the database of Genotypes and Phenotypes (dbGaP). Detailed demographics data are shown in Table 1.

These subjects contained three cohorts with alcohol dependence, including 1409 European-American cases, 1518 European-American controls, 6410 European-Australian family subjects with 1633 alcohol-dependent probands, 681 African-American cases and 508 African-American controls. All subjects in these three cohorts were interviewed using the Semi-Structured Assessment for the Genetics of Alcoholism (Bucholz *et al.*, 1994). Affected subjects met DSM-IV criteria for alcohol dependence (American Psychiatric Association, 1994). Additionally, 65.9% of patients with major depression had alcohol-drinking behavior (data not shown), i.e. at least 12 alcoholic drinks in the past 12 months. The samples with alcohol dependence and major depression were identical to those used in the published work (Boomsma *et al.*, 2008; Zuo *et al.*, 2011a,b).

Imputation

To make the genetic marker sets consistent across the different cohorts, we imputed the missing single nucleotide polymorphisms (SNPs) across the entire *ADH* gene cluster (*ADH7–ADH1C–ADH1B–ADH1A–ADH6–ADH4–ADH5* at Chr4: 100,204,900–100,631,900) in all samples using the same reference panels that included whole-genome sequencing data. To maximize the success rate and accuracy of imputation, we (a) used both 1000 Genome Project and HapMap 3 panels as the reference, and separated the European (CEU) and African (YRI) ethnicities during imputation; (b) used a Markov Chain Monte Carlo algorithm implemented in the program IMPUTE2 (Howie *et al.*, 2009) to derive full posterior probabilities, not the 'best-guess', of the genotypes of each SNP; (c) set the imputation parameters at burnin = 10,000, iteration = 10,000, $k = 100$, Ne = 11,500 and confidence level = 0.99 (Howie *et al.*, 2009); (d) merged, within the same ethnicity, the data sets as much as possible to increase sample sizes and marker density for imputation, being subject to the following criteria: cases and controls that were paired within the same study; different panels of array data in the same subjects; and separate samples that had the same phenotype and were genotyped on the same microarray platform and (e) stringently cleaned the imputed data before association analysis (see below). Additionally, because the imputation process 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 >0.5% Mendel errors (considering all SNPs tested) and the SNPs with >0.5% Mendel errors (considering all individuals tested) were excluded too. Finally, for SNPs that were directly genotyped, we used the direct genotypes rather than the imputed data.

Data cleaning

We stringently cleaned the phenotype data and the genotype data before association analysis (detailed previously; Zuo *et al.*, 2011a). Subjects with poor genotypic data and questionable diagnostic information, allele discordance, duplicated IDs, potential sample misidentification, sample relatedness, sample misspecification, gender anomalies, missing race, non-European and non-African ethnicity, population group outliers, a mismatch between self-identified and genetically inferred ethnicity, a missing genotype call rate ≥2% across all SNPs and subjects overlapped between two data sets [e.g. the Study of Addiction: Genetics and Environment (SAGE) data set and the Collaborative Study on the Genetics of Alcoholism (COGA) data set] were excluded (one copy). Furthermore, we excluded monomorphic SNPs and SNPs with allele discordance, Mendelian errors (in family samples) and an overall missing genotype call rate ≥2%. For those data sets merged from the separate samples (e.g. SAGE and COGA) that had the same phenotype and were genotyped on the same microarray platform, SNPs with allele frequency differences >2% between the original separate samples were excluded. For all merged data sets, SNPs with missing rate differences >2% between the original separate samples were

Table 1. Demographic data of all cohorts

Human disease	Ethnicity	Data set name	Pedigrees Total *n*	Subjects Total *n*	Affected subjects Male *n*	Affected subjects Male Age (yrs)	Affected subjects Female *n*	Affected subjects Female Age (yrs)	Unaffected subjects Male *n*	Unaffected subjects Male Age (yrs)	Unaffected subjects Female *n*	Unaffected subjects Female Age (yrs)
Alcohol dependence	EA (CC)	SAGE+COGA	2927	2927	883	39.0 ± 10.4	526	36.7 ± 8.8	445	37.9 ± 10.1	1073	39.0 ± 9.1
Alcohol dependence	AA (CC)	SAGE+COGA	1189	1189	428	41.0 ± 8.3	253	39.8 ± 6.8	169	40.2 ± 8.4	339	39.6 ± 6.8
Alcohol dependence	EAu (Fam)	OZ-ALC	1856	6410	1011	42.0 ± 8.4	622	39.2 ± 7.3	1709	46.3 ± 9.8	2213	45.6 ± 9.5
Major depression	CA (CC)	PRSC	3625	3625	548	44.2 ± 12.0	1257	41.2 ± 12.8	694	47.1 ± 14.4	1126	43.8 ± 13.7
Bipolar disorder	EA (CC)	BDO+GRU	1402	1402	190	43.1 ± 8.0	178	45.4 ± 10.0	532	54.7 ± 17.3	502	50.1 ± 17.6
Bipolar disorder	EA (CC)	BARD+GRU	1687	1687	322	42.1 ± 8.3	331	44.4 ± 9.7	532	54.7 ± 17.3	502	50.1 ± 17.6
Bipolar disorder	AA (CC)	BARD+GRU	812	812	39	42.4 ± 7.9	102	42.0 ± 7.8	272	46.0 ± 14.0	399	45.7 ± 13.5
Schizophrenia	AA (CC)	GAIN	2149	2149	746	41.9 ± 10.8	449	43.0 ± 9.8	362	46.2 ± 13.7	592	45.0 ± 12.9
Schizophrenia	EA (CC)	GAIN	2729	2729	947	42.5 ± 11.3	404	45.1 ± 11.2	634	53.5 ± 17.0	744	49.2 ± 16.7
Schizophrenia	EA (CC)	nonGAIN	2784	2784	996	42.3 ± 11.8	441	44.2 ± 12.4	669	51.8 ± 15.3	678	47.9 ± 16.1
Schizophrenia	AA (CC)	nonGAIN	118	118	60	41.5 ± 11.3	38	42.9 ± 10.5	20	49.7 ± 9.2	0	—
Autism	EA (Fam)	AGP	1366	4075	1121	7.2 ± 3.2	209	7.1 ± 3.0	0	—	0	—
ADHD	CA (Fam)	IMAGE	922	2757	802	10.9 ± 2.8	122	10.8 ± 3.0	0	—	0	—
Alzheimer's disease	CA (Fam)	LOAD × 4	2243	5219	788	84.1 ± 8.1	1510	86.3 ± 8.7	486	66.7 ± 10.7	773	66.1 ± 10.5
Alzheimer's disease	EA (CC)	GenADA	1588	1588	340	77.6 ± 8.3	466	78.3 ± 8.8	279	74.4 ± 7.7	503	72.8 ± 8.1
ALS	CA (CC)	GRU	507	507	138	56.5 ± 11.9	123	59.2 ± 11.6	136	69.6 ± 8.6	110	69.8 ± 8.9
Early onset stroke	EA (CC)	GEOS × 3	802	802	198	42.7 ± 6.1	174	38.7 ± 7.6	208	40.6 ± 6.4	222	37.8 ± 7.3
Early onset stroke	AA (CC)	GEOS × 3	599	599	144	42.3 ± 6.2	165	41.0 ± 7.3	129	40.9 ± 6.5	161	38.4 ± 7.5
Ischemic stroke	CA (CC)	ISGS	485	485	119	71.8 ± 8.3	100	71.6 ± 8.1	128	69.7 ± 8.6	138	69.7 ± 8.9
Parkinson's disease	CA (CC)	NGRC	3986	3986	1346	67.2 ± 10.5	654	67.3 ± 11.0	769	70.7 ± 13.9	1217	70.1 ± 14.2
Parkinson's disease	CA (CC)	PDRD+GRU	1767	1767	537	70.5 ± 9.3	363	70.2 ± 10.0	346	53.5 ± 15.7	521	55.2 ± 11.2
Parkinson's disease	CA (CC)	lng_coriell_pd	1741	1741	560	66.3 ± 10.9	380	65.9 ± 11.2	336	62.3 ± 14.4	465	56.0 ± 17.2

In the family data, only the affected and unaffected offspring are listed. Data set names refer to dbGaP.
n, sample size; yrs, years; CC, case–control sample; Fam, family sample. EA, European-American; AA, African-American; EAu, European-Australian; CA, Caucasian; ADHD, attention deficit hyperactivity disorder; ALS, amyotrophic lateral sclerosis. [GenADA: Li *et al*. Arch Neurol. 2008; 65(1):45–53; Filippini *et al*. Neuroimage. 2009;44(3):724–728. AGP: The AGP Consortium. Nature. 2010;466(7204):368–372; Human Molecular Genetics. 2010;19(20):4072–4082; Nature Genetics. 2007;39(3):319–328].

Table 2. Associations between ADH gene cluster and different neuropsychiatric or neurological disorders

Human disease	Ethnicity	Data set name	dbGaP#	Most sig. SNP	Gene	Affected *n*	Affected MAF	Unaffected *n*	Unaffected MAF	Minimal *P*-value	SNP # (total)	SNP # ($P < 0.05$)
Alcohol dependence	EA (CC)	SAGE+COGA	phs000092.v1.p1	rs1596180	*ADH7*	1409	0.018	1518	0.007	0.0009	343	9
Alcohol dependence	AA (CC)	SAGE+COGA	phs000092.v1.p1	rs114618736	*ADH1C*	681	0.018	508	0.006	0.0108	486	6
Alcohol dependence	EAu (Fam)	OZ-ALC	phs000181.v1.p1	rs11733695	*ADH6*	1633	0.042	1633	0.105	0.0353	385	2
Major depression	CA (CC)	PRSC	phs000020.v2.p1	rs7690269	*ADH7*	1805	0.019	1820	0.009	0.0004	341	16
Bipolar disorder	EA (CC)	BDO+GRU	phs000017.v3.p1	rs6532797	*ADH4*	368	0.043	1034	0.021	0.0072	215	10
Bipolar disorder	EA (CC)	BARD+GRU	phs000017.v3.p1	rs1391088	*ADH1C*	653	0.060	1034	0.038	0.0224	250	7
Bipolar disorder	AA (CC)	BARD+GRU	phs000017.v3.p1	rs283417	*ADH1C*	141	0.029	671	0.008	0.0094	193	2
Schizophrenia	AA (CC)	GAIN	phs000021.v3.p2	rs4699743	*ADH1C*	1195	0.013	954	0.030	0.0086	276	11
Schizophrenia	EA (CC)	GAIN	phs000021.v3.p2	rs60652198	*ADH4*	1351	0.041	1378	0.027	0.0165	277	4
Schizophrenia	EA (CC)	nonGAIN	phs000167.v1.p1	rs71612689	*ADH7*	1437	0.004	1347	0.000	0.0220	354	5
Schizophrenia	AA (CC)	nonGAIN	phs000167.v1.p1	rs76919634	*ADH6*	98	0.008	20	0.045	0.2905	35	0
Autism	EA (Fam)	AGP	phs000267.v1.p1	rs62325239	*ADH5*	1330	0.003	1330	0.002	0.0141	361	44
ADHD	CA (Fam)	IMAGE	phs000016.v2.p2	rs1442483	*ADH7*	924	0.017	924	0.059	0.0039	356	9
Alzheimer's Disease	CA (Fam)	LOAD × 4	phs000168.v1.p1	rs116192122	*ADH4*	2298	0.012	2298	0.002	0.0030	356	10
Alzheimer's disease	EA (CC)	GenADA	phs000219.v1.p1	rs35361391	*ADH4*	806	0.034	782	0.009	0.0162	267	18
ALS	CA (CC)	GRU	phs000101.v3.p1	rs115081066	*ADH4*	261	0.006	246	0.033	0.0017	334	7
Early Onset Stroke	EA (CC)	GEOS × 3	phs000292.v1.p1	rs1596180	*ADH1B*	372	0.019	430	0.003	0.0048	301	8
Early onset stroke	AA (CC)	GEOS × 3	phs000292.v1.p1	rs114188790	*ADH1C*	309	0.081	290	0.036	0.0026	451	48
Ischemic stroke	CA (CC)	ISGS	phs000102.v1.p1	rs72681936	*ADH7*	219	0.078	266	0.034	0.0202	348	6
Parkinson's disease	CA (CC)	NGRC	phs000196.v2.p1	rs78304974	*ADH1B*	2000	0.007	1986	0.002	0.0089	341	7
Parkinson's disease	CA (CC)	PDRD+GRU	phs000126.v1.p1	rs1693457	*ADH7*	900	0.029	867	0.048	0.0129	360	25
Parkinson's disease	CA (CC)	lng_coriell_pd	phs000089.v3.p2	rs28472487	*ADH7*	940	0.026	801	0.043	0.0260	354	11

Only the most significant risk markers are listed; in family-based cohorts, N = sample size of affected offspring; "affected MAF" = "transmitted MAF", "unaffected MAF" = "untransmitted MAF" in offspring. Data set names refer to dbGaP. COGA data set access number is phs000125.v1.p1. MAF, minor allele frequency; *n*, CC, Fam, EA, AA, EAu, CA, ADHD and ALS: also see Table 1. The corrected α was 2.1×10^{-5} (European) and 1.5×10^{-5} (African), respectively.

also excluded. The SNPs with MAF = 0 in either cases or controls were excluded, because it could not be determined if they were missed during the imputation process or truly non-polymorphic in nature in some phenotype groups. Finally, only a total of 870 SNPs with 0<MAF<0.05 in either cases or controls were extracted for association analysis

Table 3. *P*-values for associations between rare variant constellations and diseases

Diseases	Ethnicity	Data set name	Tests	MAF upper bound	ADH cluster	ADH1A	ADH1B	ADH1C	ADH4	ADH5	ADH6	ADH7
Alcohol dependence	EA	SAGE+COGA	Fp	0.01	**0.045**	0.231	0.979	0.591	0.085	0.725	**0.008**	0.076
Alcohol dependence	EA	SAGE+COGA	Fp	0.05	0.707	0.929	0.965	0.210	0.893	0.965	0.655	0.650
Alcohol dependence	EA	SAGE+COGA	VT	Variable	0.089	0.239	0.887	0.877	0.329	0.986	**0.010**	0.138
Alcohol dependence	EA	SAGE+COGA	Collapsing	0.05	0.542	0.827	0.943	0.197	0.693	0.847	0.479	0.568
Alcohol dependence	EAu	OZ-ALC	Fp	0.01	0.206	0.092	0.509	0.056	0.645	0.765	0.397	0.195
Alcohol dependence	EAu	OZ-ALC	Fp	0.05	**0.027**	0.635	**0.025**	**0.034**	0.315	0.480	0.449	**0.009**
Alcohol dependence	EAu	OZ-ALC	VT	Variable	0.055	0.396	0.103	**0.023**	0.344	0.797	**0.030**	**0.047**
Alcohol dependence	EAu	OZ-ALC	Collapsing	0.05	**0.038**	0.851	**0.016**	**0.038**	0.335	0.482	0.472	**0.005**
Alcohol dependence	AA	SAGE+COGA	Fp	0.01	0.543	0.941	0.795	0.452	0.492	0.447	0.496	0.889
Alcohol dependence	AA	SAGE+COGA	Fp	0.05	**0.050**	0.784	0.723	0.151	0.077	0.121	**0.051**	0.491
Alcohol dependence	AA	SAGE+COGA	VT	Variable	0.226	0.822	0.854	0.563	0.273	0.173	0.099	0.981
Alcohol dependence	AA	SAGE+COGA	Collapsing	0.05	**0.038**	0.555	0.861	0.139	0.051	0.069	0.056	0.581
Major depression	CA	PRSC	Fp	0.01	0.307	0.643	0.282	0.339	0.765	0.646	**0.040**	0.697
Major depression	CA	PRSC	Fp	0.05	0.107	0.349	0.319	0.856	0.123	0.071	0.294	0.675
Major depression	CA	PRSC	VT	Variable	0.557	0.198	0.678	0.496	0.607	0.314	0.302	0.133
Major depression	CA	PRSC	Collapsing	0.05	0.108	0.336	0.359	0.770	0.106	0.072	0.361	0.621

MAF, minor allele frequency; Fp1, Fp5 and VT, association tests using SCORE-Seq; Collapsing, association test using ARIEL; EA, EAu, AA, CA and data set names refer to Table 1. Significant *P*-values are bold.

(Supplementary data, Table A1). The cleaned sample sizes, cleaned SNP numbers, ethnicity, diagnosis, dbGaP access numbers and data set name abbreviations of these samples are shown in Tables 1 and 2.

Association tests for region-wide rare variant constellations

Synthetic effects of region-wide rare variant constellations may be more significant than individual rare variants in some specific gene regions on disease phenotypes. These effects were tested using a score-type program, SCORE-Seq (Lin and Tang, 2011). The mutation information was aggregated by virtue of a weighted linear combination across all rare variants of the entire *ADH* gene cluster or across each *ADH* gene region, and then related to disease phenotypes using appropriate regression models. Sex, age, alcohol drinking and the first 10 principal components served as the covariates in the regression models. Principal component scores for each individual were estimated using the program EIGENSTRAT (Price *et al.*, 2006). The first 10 principal components explained >95% of variance in our samples. Two fixed MAF threshold with flexible weight tests (Fp1: MAF <0.01; Fp5: MAF <0.05) and one variable threshold with fixed weight test (VT test: MAF <0.05) were performed to derive the global *P*-values from these regression models (Table 3). In Fp tests, the weight was 1/sqrt(p(1-p)) where *P* was the estimated MAF with pseudo counts in the pooled sample. In VT test, the weight was 1 when MAF <threshold and 0 otherwise, where the threshold varied between 0 and 0.05. Statistical significance was assessed by resampling 1 million times (Lin and Tang, 2011). Additionally, we used ARIEL (Asimit *et al.*, 2012), a regression-based collapsing approach that incorporates variant quality scores, to confirm the tests by SCORE-Seq. All association analyses were performed within the same ethnicity.

Association tests for individual rare variants

For case–control samples, the allele frequencies of each SNP were compared between cases and controls using logistic regression analysis as implemented in PLINK (Purcell *et al.*, 2007). Diagnosis served as the dependent variable, alleles served as the independent variables and sex, age, alcohol drinking and the first 10 principal components served as the covariates. For family samples, we tested the allele-disease associations using the program Family-Based Association Test (Horvath *et al.*, 2001). The MAFs and *P*-values of the most significant risk SNPs and the numbers of the nominally significant risk SNPs ($P < 0.05$) in all samples are shown in Table 2.

Correction for multiple testing in single-point association tests

The experiment-wide significance levels (α) were corrected for the numbers of cohorts (i.e. 22) and the numbers of effective markers that were calculated by the program SNPSpD (Li and Ji, 2005), which is an adjusted Bonferroni correction taking the linkage disequilibrium structure into account. Approximately, 110 and 150 effective SNPs captured most of the information of all rare variants across the entire *ADH* gene cluster in cohorts of European and African descent, respectively. Thus, the corrected significance levels (α) for single-point association tests were set at 2.1×10^{-5} in cohorts of European descent and 1.5×10^{-5} in cohorts of African descent, respectively.

RESULTS

The rare variant constellation across the entire *ADH* gene cluster was specifically associated with alcohol dependence in European-Americans [Fp1: global $P = 0.045$; 108 variants (SNPs) with 2067 minor alleles], European-Australians (Fp5: global $P = 0.027$; Collapsing $P = 0.038$; 388 variants with 92,429 minor alleles) and African-Americans (Fp5: global $P = 0.050$; Collapsing $P = 0.038$; 486 variants with 20,513 minor alleles), but not with any other neuropsychiatric disease ($P > 0.10$). In testing the rare variant constellations within each individual gene region, several results were obtained. First, the *ADH6* variant constellation was significantly associated with alcohol dependence in European-Americans (Fp1: $P = 0.008$; VT: $P = 0.010$; 10 variants with 155 minor alleles), European-Australians (VT:

$P = 0.030$; 49 variants with 10,546 minor alleles) and African-Americans (Fp5: $P = 0.051$; Collapsing $P = 0.056$; 85 variants with 4529 minor alleles). Second, the *ADH7* variant constellation was significantly associated with alcohol dependence in European-Australians (Fp5: $P = 0.009$; VT: $P = 0.047$; Collapsing $P = 0.005$; 98 variants with 20,280 minor alleles), and suggestively in European-Americans (Fp1: $P = 0.076$; 22 variants with 348 minor alleles). Third, the *ADH1B* and *ADH1C* variant constellations were modestly associated with alcohol dependence in European-Australians (for *ADH1B*: Fp5: $P = 0.025$ and collapsing: $P = 0.016$; for *ADH1C*: Fp1: $P = 0.056$, Fp5: $P = 0.034$, VT: $P = 0.023$ and collapsing: $P = 0.038$), but not in European-Americans and African-Americans ($P > 0.10$; Table 3). Additionally, single-point association analysis showed that, of a total of 343 individual rare variants in European-Americans, 9 SNPs were nominally associated with alcohol dependence ($P < 0.05$), the most significant of which (rs1596180, at 5 of *ADH7*) was suggestively associated with alcohol dependence ($P = 0.0009$; Table 2).

The rare variant constellation across the *ADH6* gene region was also modestly associated with major depression in Caucasians (Fp1: $P = 0.040$; 10 variants with 307 minor alleles). This association turned out to be non-significant after correction for multiple testing. Furthermore, among a total of 341 individual rare variants in Caucasians, 16 SNPs were nominally associated with major depression ($P < 0.05$), the most significant of which (rs7690269, at 5 of *ADH7*) was suggestively associated with major depression (OR = 2.16; $P = 0.0004$). This rs7690269 was also the most significant one among all 22 cohorts. Finally, no individual variants were statistically significantly associated with any disease examined after group- and region-wide correction ($P > \alpha$), including alcohol dependence and major depression (Table 2).

DISCUSSION

We found that rare *ADH* variant constellations were specific for alcohol dependence. In particular, a rare *ADH6* variant constellation showed replicable association with alcohol dependence across three independent cohorts of European or African descent. Additionally, *ADH7*, *ADH1B* and *ADH1C* variant constellations might also be implicated in the risk for alcohol dependence. We speculate that the *ADH* gene cluster may harbor a causal variant(s) for alcohol dependence.

Searching the entire *ADH* cluster, we found no individual rare variants which were statistically significantly associated with any disease examined (including alcohol dependence) after group- and region-wide correction for multiple comparisons. Our study provides an additional example to support the hypothesis that the synthetic effects of region-wide rare variant constellations may be more significant than individual rare variants on disease phenotypes. Using multiple cohorts with large sample sizes, we found that rare *ADH* variant constellations were specific for alcohol dependence, but not associated with any other disease, which was consistent with previous reports (Luo *et al*., 2006) and with the fact that the ADH enzymes are mainly distributed in the liver, but only partly distributed in the central nervous system. Although the synthetic effects of rare *ADH* variants on alcohol dependence seemed to be modest in the present study, these effects appeared to be highly significant when compared with those on other 'non-alcohol dependence' neuropsychiatric disorders.

When testing each gene region, we detected modest associations between rare *ADH1B* and *ADH1C* variant constellations and alcohol dependence in European-Australians. The variants in these two genes may influence the risk of alcohol dependence via ethanol metabolism pathways, which is well-known by numerous studies. However, these associations were not strong and not replicated in other populations in the present study. They remained to be confirmed in the future.

More robust associations were detected between *ADH6* variants and alcohol dependence, which was replicable in three cohorts. Alteration of ADH6 enzyme activity caused by *ADH6* variants may influence the ethanol metabolism as introduced above, and thus may influence the human drinking behavior and the risk for alcohol dependence. Alternatively, the retinol metabolism pathway or other non-ethanol metabolism pathways introduced above may be other possible mechanisms underlying the associations between *ADH6* variant constellation and alcohol dependence, and possibly the suggestive association between *ADH6* variant constellation and major depression as well. Similarly, these mechanisms might also underlie the suggestive associations between the rare *ADH7* variant constellation and alcohol dependence and between individual *ADH7* variants and major depression.

SUPPLEMENTARY DATA

Supplementary data are available at *Alcohol and Alcoholism* online.

Funding — This work was supported in part by National Institute on Drug Abuse (NIDA) grants K01DA029643, K24DA017899 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.). We 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, R01AA013320, 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, R01MH59571, R01MH59586, R01MH59587, R01MH59588, R01MH60870, R01MH60879, R01MH61675, R01MH62873, R01MH081803, R01MH67257, R01MH81800, U01MH46276, U01MH46282, U01MH46289, U01MH46318, 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 Co-morbidity; 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, and GlaxoSmithKline, R&D Limited. The data sets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap through dbGaP accession numbers listed in Table 2.

Conflict of interest statement. J.H.K. has been a paid consultant for Aisling Capital, LLC, AstraZeneca Pharmaceuticals, Brintnall & Nicolini, Inc., Easton Associates, Gilead Sciences, Inc., GlaxoSmithKline, Janssen Pharmaceuticals, Lundbeck Research USA, Medivation, Inc., Merz Pharmaceuticals, MK Medical Communications, F. Hoffmann-La Roche Ltd, SK Holdings Co., Ltd, Sunovion Pharmaceuticals, Inc., Takeda Industries and Teva Pharmaceutical Industries, Ltd. He serves as a member of Scientific Advisory Boards for Abbott Laboratories, Bristol-Myers Squibb, Eisai, Inc., Eli Lilly and Co., Forest Laboratories, Inc., Lohocla Research Corporation, Mnemosyne Pharmaceuticals, Inc., Naurex, Inc., Pfizer Pharmaceuticals and Shire Pharmaceuticals. He is the Editor for Biological Psychiatry, a member of Board of Directors of Coalition for Translational Research in Alcohol and Substance Use Disorders, and the President Elect for American College of Neuropsychopharmacology. He also gets support from Tetragenex Pharmaceuticals. Additionally, H.Z. has been a paid consultant for Eisai, Inc. Other authors have no conflict of interest.

See Letters to the Editor (p.129) for a response to this article.

REFERENCES

American Psychiatric Association. (1994) *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Press.

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

Bierut LJ, Agrawal A, Bucholz KK *et al*. (2010) A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* **107**:5082–7.

Boomsma DI, Willemsen G, Sullivan PF *et al*. (2008) Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet* **16**:335–42.

Bucholz KK, Cadoret R, Cloninger CR *et al*. (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**:149–58.

Buervenich S, Sydow O, Carmine A *et al*. (2000) Alcohol dehydrogenase alleles in Parkinson's disease. *Mov Disord* **15**:813–8.

Buervenich S, Carmine A, Galter D *et al*. (2005) A rare truncating mutation in ADH1C (G78Stop) shows significant association with Parkinson disease in a large international sample. *Arch Neurol* **62**:74–8.

Dickson SP, Wang K, Krantz I *et al*. (2010) Rare variants create synthetic genome-wide associations. *PLoS Biol* **8**:e1000294.

Edenberg HJ, Koller DL, Xuei X *et al*. (2010) Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol Clin Exp Res* **34**:840–52.

Frank J, Cichon S, Treutlein J *et al*. (2012) Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict Biol* **17**:171–80.

Grant BF, Stinson FS, Dawson DA *et al*. (2004) 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* **61**:807–16.

Heath AC, Whitfield JB, Martin NG *et al*. (2011) A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biol Psychiatry* **70**:513–8.

Horvath S, Xu X, Laird NM. (2001) The family based association test method: strategies for studying general genotype–phenotype associations. *Eur J Hum Genet* **9**:301–6.

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.

Kessler RC, Nelson CB, McGonagle KA *et al*. (1996) The epidemiology of co-occurring addictive and mental disorders: implications for prevention and service utilization. *Am J Orthopsychiatry* **66**:17–31.

Li J, Ji L. (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**:221–7.

Lin DY, Tang ZZ. (2011) A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* **89**:354–67.

Luo X, Kranzler HR, Zuo L *et al*. (2006) Diplotype trend regression analysis of the ADH gene cluster and the ALDH2 gene: multiple significant associations with alcohol dependence. *Am J Hum Genet* **78**:973–87.

Luo X, Kranzler HR, Zuo L *et al*. (2008) ADH7 variation modulates extraversion and conscientiousness in substance-dependent subjects. *Am J Med Genet B Neuropsychiatr Genet* **147B**:179–86.

Price AL, Patterson NJ, Plenge RM *et al*. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**:904–9.

Purcell S, Neale B, Todd-Brown K *et al*. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**:559–75.

Regier DA, Farmer ME, Rae DS *et al*. (1990) Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *J Am Med Assoc* **264**:2511–18.

Satre MA, Zgombic-Knight M, Duester G. (1994) The complete structure of human class IV alcohol dehydrogenase (retinol dehydrogenase) determined from the ADH7 gene. *J Biol Chem* **269**:15606–12.

Shmueli O, Horn-Saban S, Chalifa-Caspi V *et al*. (2003) GeneNote: whole genome expression profiles in normal human tissues. *C R Biol* **326**:1067–72.

Suzuki Y, Fujisawa M, Ando F *et al*. (2004) Alcohol dehydrogenase 2 variant is associated with cerebral infarction and lacunae. *Neurology* **63**:1711–13.

Yanai I, Benjamin H, Shmoish M *et al*. (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**:650–9.

Yasunami M, Chen CS, Yoshida A. (1991) A human alcohol dehydrogenase gene (ADH6) encoding an additional class of isozyme. *Proc Natl Acad Sci USA* **88**:7610–4.

Zuo L, Gelernter J, Zhang CK *et al*. (2011a) Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. *Neuropsychopharmacology* **37**:557–66.

Zuo L, Zhang CK, Wang F *et al*. (2011b) A novel, functional and replicable risk gene region for alcohol dependence identified by genome-wide association study. *PLoS ONE* **6**:e26726.