**ORIGINAL ARTICLE**

Rare ADH Variant Constellations are Specific for Alcohol Dependence

Lingjun Zuo1, Heping Zhang2, Robert T. Malison1, Chiang-Shan R. Li1, Xiang-Yang Zhang3, Fei Wang1, Lingeng Lu2, Lin Lu4, Xiaoping Wang5, John H. Krystal1,6,7, Fengyu Zhang6, Hong-Wen Deng6 and Xingguang Luo1,*

1Department of Psychiatry, University School of Medicine, West Haven, CT 06516, USA, 2Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA, 3Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA, 4National Institute on Drug Dependence, Peking University, Beijing, China, 5Department of Neurology, First People’s Hospital, Shanghai Jiaotong University, Shanghai, China, 6Psychiatry Services, Yale-New Haven Hospital New Haven, CT, USA, 7VA Alcohol Research Center, VA Connecticut Healthcare System, West Haven, CT, USA, 8Psychiatry Institute for Brain Development, Johns Hopkins University Medical Campus, Baltimore, MD, USA and 9Department 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. β1ADH encoded by ADH1B, γγADH encoded by ADH1C and ADH6 enzyme encoded by ADH6) and the upper digestive tract (e.g. σαADH encoded by ADH7) and partly in the central nervous system (e.g. σαADH 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 oxidation of retinol, a vitamin A precursor (summarized in Satre et al., 1994; Luo et al., 2008). For example, σαADH, 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; Ile350Val) that significantly affect the risk of alcohol dependence are rare variants in most populations, e.g. in Asians [minor allele frequency (MAF) \( f_{\text{rs2066702}} = 0.000 \); \( f_{\text{rs1693482}} = 0.023 \); \( f_{\text{rs698}} = 0.025 \)]; Europeans \( f_{\text{rs2066702}} = 0.000 \); \( f_{\text{rs1229984}} = 0.008 \) and/or Africans \( f_{\text{rs1229984}} = 0.000 \); \( f_{\text{rs1693482}} = 0.052 \); \( f_{\text{rs698}} = 0.042 \) (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 \[ P = 1.3 \times 10^{-8}; \text{ odds ratio (OR)} = 1.46 \] (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 demographic 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 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, \( k = 100 \), \( N_e = 11,500 \) 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...
if they were missed during the imputation process or truly excluded. The SNPs with MAF = 0 in either cases or controls were also excluded. The SNPs with MAF = 0 in either cases or controls were also 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 MAF<0.05 in either cases or controls were extracted for association analysis.
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: 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 alcohol dependence (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 > α), 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 R01DA029643, K24DA017899 and R01DA016750, National Institute on Alcohol Abuse and Alcoholism (NIAAA) grants R01 AA16015 and R21 AA202319 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, Somoza-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) (U01HG004424, U01HG004436 and U01HG004438); the GENEVA Coordinating Center (U01HG004446); the NIAAA (U01AA088401, R01AA013520, 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) (R01INS4012); 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, R01MH63803, R01MH67257, R01MH681800, U01MH66276, U01MH66282, U01MH66289, U01MH66318, U01MH70469, U01MH70470 and R01MH672377); 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,
Conflict of interest statement. J.H.K. has been a paid consultant for Aisling Capital, LLC, AstraZeneca Pharmaceuticals, Brinmnall & 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., Loholca Research Corporation, Mmesome Pharmaceutical, Inc., Naurex, Inc., Pfizer Pharmaceuticals and Shire Pharmaceuticals. He is the Editor for Biological Psychiatry, a member of Board of Directors for 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


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.


