Published in final edited form as:

Alcohol Clin Exp Res. 2017 October ; 41(10): 1695–1704. doi:10.1111/acer.13480.

Associations between Genomic Variants in Alcohol Dehydrogenase (ADH) Genes and Alcohol Symptomatology in American Indians and European Americans: Distinctions and Convergence

Q. Peng, Ph.D.^{1,*}, Ian R. Gizer, Ph.D.², K.C. Wilhelmsen, M.D., Ph.D.³, and C.L. Ehlers, Ph.D.^{1,*}

¹Department of Neuroscience, The Scripps Research Institute, La Jolla, CA 92037 USA

²Department of Psychological Sciences, University of Missouri-Columbia, Columbia, MO 65211 USA

³Department of Genetics and Neurology, University of North Carolina, Chapel Hill, NC 27599 USA

Abstract

Background—Higher rates of alcohol use disorders (AUD) have been observed in some Native American (NA) populations than other ethnic groups such as Euro-Americans (EA) in the U.S. Previous studies have shown that variation in the alcohol dehydrogenase (*ADH*) genes may affect the risk for development of AUD, and that the prevalence of these variants differs depending on the ancestral origins of a population.

Methods—In the present study, we assessed sequencing variants in the *ADH* genomic region (*ADH1-7*) and tested for their associations with AUD phenotypes in two independent populations: an American Indian (AI) community sample and an EA cohort from the San Francisco Family Alcohol Study. Association tests were conducted for both common and rare variants using sequencing data for two phenotypes: the number of alcohol-related life events, and the count of alcohol dependence drinking symptoms. A regularized regression method was used to select the best set of *ADH* variants associated with phenotypes. Variance component model was incorporated in all analyses to leverage the admixture and relatedness.

Results—Two variants near *ADH4* and two near *ADH1C* exhibited significant associations with AUD in AI; no variant was significant in EA. Common risk variants in AI were either absent from or much less frequent in EA. The feature selection method selected mostly distinct yet often co-located subsets of *ADH* variants to be associated with AUD phenotypes between the two cohorts.

*Corresponding Authors: Cindy L. Ehlers, 10550 North Torrey Pines Road, La Jolla, CA 92037, Mail SP30-1501, Tel: 858-784-7055; Fax: 858-784-7475; cindy@scripps.edu, Qian Peng, 3366 North Torrey Pines Court, Suite 200, La Jolla, CA 92037, Tel: 858-784-7215; qpeng@scripps.edu.
DR. QIAN PENG (Orcid ID : 0000-0002-2662-3412)

Authors Contribution

Conceived and designed the study: QP CLE. Performed the analysis: QP. Contributed materials/data: IRG KCW. Wrote the paper: QP CLE. Discussed the interpretation of the data: QP IRG KCW CLE. Critically reviewed, revised and approved the manuscript: QP IRG KCW CLE.

In the rare variant analyses, the only association was observed between the whole region and the alcohol-related life events in AI.

Conclusions—Our results suggest that *ADH* variants, both common and rare, are more likely to impact risk for alcohol-related symptomatology in this AI population than in this EA sample; and *ADH* variants that might affect AUD are likely different but convergent on similar regions between the two populations.

Keywords

ADH Sequencing Variants; Alcohol Use Disorder; Genomic Variant Selections; Ethnicity

Introduction

Large-scale epidemiological studies have shown that alcohol use disorders (AUD), as defined by DSM-5 (American Psychiatric Association, 2013), are highly prevalent in the U.S. (Grant et al., 2015). The rates of AUD vary across ethnic groups and populations. Compared to other ethnic groups such as European Americans (EA) in the U.S., the rates of AUD and other drug dependence are among the highest (4–5 times of the national average) in some American Indian/Alaska Native (AI/AN) populations and the lowest among some Asian population (Grant et al., 2015, Compton et al., 2007). Ethnic differences in the rates of alcohol and other substance dependence are thought to be due to both genetic and environmental factors (Ehlers and Gizer, 2013).

Alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) are the primary enzymes involved in alcohol (ethanol) metabolism. The main pathway of ethanol metabolism consists of its conversion to acetaldehyde mediated by ADH, and subsequent conversion from acetaldehyde to acetate by ALDH. Liver is the primary site of ethanol oxidation. Both enzymes occur in multiple forms that differ in their kinetic properties, and are encoded by different genes. Further, coding variants within these genes have been shown to alter the properties of the enzymes they encode. For example, previous studies have suggested that variants within the *ADH* and the *ALDH* genes may affect the risk of development of AUD (Bosron et al., 1993, Edenberg, 2007, Chen et al., 2009). Indeed one of the most consistent linkage findings for alcohol dependence contains the *ADH* gene cluster on chromosome 4q: the region was initially reported in the Collaborative Studies on the Genetics of Alcoholism (COGA) study (Reich et al., 1998, Edenberg and Foroud, 2006) and subsequently replicated in different populations including European Americans, African Americans and American Indians (Ehlers et al., 2004a, Prescott et al., 2006, Gizer et al., 2011b).

Variants within these genes, however, may differentially affect the risk for the disorder depending on the ethnic groups being evaluated. For instance, the *ADH1B*2* and *ADH1B*3* alleles—each resulting in an amino acid change—produce more efficient enzymes that allow rapid acetaldehyde accumulation (Carr et al., 1989, Hurley et al., 1990). As a result, *ADH1B*2* has shown a protective effect against alcohol dependence in several populations, particularly in East Asians where the allele is found at high frequency (Macgregor et al., 2009, Thomasson et al., 1991, Thomasson et al., 1994, Gelernter et al., 2014, Park et al.,

2013). Similarly, the *ADH1B*3* allele has shown a significant protective effect on risk for alcoholism in certain African-American populations (Edenberg and Foroud, 2006, Gelernter et al., 2014). The allele also showed a protective effect in an admixed American Indian population (Wall et al., 2003). Other prominent associations between *ADH* genes and alcohol dependence have been reported for *ADH1C* in various populations such as East Asian (Higuchi et al., 1995), European (Kuo et al., 2008), and American Indian (Mulligan et al., 2003), and for *ADH4* in Euro-American (Edenberg et al., 2006), Australian (Macgregor et al., 2009) and American Indian (Gizer et al., 2011a) cohorts. The majority of these studies were based on linkage analysis or candidate genes; a small number resulted from genome-wide association studies (Gelernter et al., 2014).

The advent of genome sequencing technology and analysis methodology has enabled us to investigate common as well as rare genomic variants at high resolution in special populations enriched for AUD phenotypes. The present report is part of a larger study exploring risk factors for substance dependence in an American Indian community (Ehlers et al., 2004b, Gilder et al., 2004, Gilder et al., 2008) and a nationwide study on genetics of alcoholism and other substance dependence in a primarily European American population (Seaton et al., 2004, Vieten et al., 2004, Gizer et al., 2011b). We previously evaluated effects of variants in the *ADH* and *ALDH* genes on alcohol dependence phenotypes in a subset of the AI cohort, and found significant evidence for association with a severity level of alcohol drinking-related dependence symptoms phenotype across the *ADH* gene cluster (Peng et al., 2014). Our study also suggested that a higher degree of AI ancestry was associated with higher frequencies of potential risk variants. Following that study, we sequenced the EA cohort and additional AI samples using the same pipeline to allow for direct comparisons of findings across the two cohorts.

The aim of the present study was twofold. The first aim of the study was to assess the influence of common variation in the *ADH* genomic region (*ADH1-7*) on AUD phenotypes for the two independent populations: AI and EA. We focused on the *ADH* region in the present work given the previously reported significant results on the *ADH* genes but little evidence of effects of *ALDH* variants in the American Indians. The AUD phenotypes we investigated include the count of alcohol dependence drinking symptoms and the weighted alcohol-related life events, both quantifying the level of severity or progression of AUD. To accomplish this aim, a mixed effect model was used to test for single variant associations to incorporate admixture and familial relatedness. And a feature selection method was used to select multiple variants in the region that were collectively associated with the phenotypes. The second aim was to assess the influence of rare variation in the *ADH* genomic region (*ADH1-7*) on AUD phenotypes for the two independent populations. The sequence kernel association test (SKAT) was used for rare-variant analysis.

Materials and Methods

Participants and phenotypes

The protocol of the study of the American Indian cohort was approved by the Scripps Research Institute Institutional Review Board (IRB) and Indian Health Council, a tribal review group overseeing health issues for the reservations where recruitments took place.

The protocol for collection of participants in the San Francisco Family Alcohol Study (SFFS) was approved by the UCSF Committee for the Protection of the Rights of Human Subject while the recruitment took place. Subsequently, the University of North Carolina, at Chapel Hill IRB approved the data analysis plan. Written informed consent was obtained from each participant after study procedures had been fully explained. Participants were compensated for their time spent in the study.

American Indian Cohort (AI)—903 American Indians from extended pedigrees participated in the study. The population characteristics and the recruitment procedures have been previously described (Ehlers et al., 2011). 750 individuals had their whole genome sequenced.

San Francisco Family Alcohol Study (SFFS)—The SFFS was a nationwide genetic study on alcoholism and other substance dependence. The recruitment process and the population characteristics have been previously described (Seaton et al., 2004, Vieten et al., 2004, Gizer et al., 2011b). 2524 individuals participated in the study. 1889 subjects had their whole genome sequenced.

All participants were assessed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al., 1994, Hesselbrock et al., 1999, Wall et al., 2003) in order to collect demographic information and to make DSM-IV and DSM-5 diagnoses (American Psychiatric Association Task Force on DSM-IV, 1994, American Psychiatric Association, 2013). After quality control and missing data removal, 742 participants from the American Indian cohort and 1215 participants from the SFFS cohort remained for further analysis in the present study. Their demographics are listed in Table 1.

Two phenotypes were studied, a count variable corresponding to the presence or absence of a subset of 4 DSM-IV alcohol dependence symptoms related to drinking severity, and the sum of weighted alcohol-related life events derived from 36 items measuring the clinical course of alcohol dependence. The number of DSM-IV alcohol dependence symptoms related to drinking included: 1) drank more than intended/more days in a row or when promised self wouldn't for three or more times; 2) drunk when didn't want to three or more times; 3) had little time for anything else during drinking or recovering from the effects of drinking; and 4) given up or greatly reduced important activities to drink (Ehlers et al., 2004c). Scores on this DSM-IV alcohol dependence-based drinking severity index ranged from 0 to 4. The weighted alcohol-related life events phenotype was derived from 36 alcohol clinical course items of DSM-5 alcohol use disorders (Ehlers et al., 2015). The 36 items are listed in Table 2 and ordered by the mean age of occurrence for each event. The alcohol life events were given a severity weight of 1 for items 1–12, 2 for 13–24 and 3 for 25–36 such that events that typically happen later in life are weighted higher. The phenotype is defined as the sum of the severity weights of the 36 alcohol-related life events. It is a new metric reflecting the severity and progression of AUD. The statistical characteristics of the two phenotypes are listed in Table S1 for the two cohorts.

Sequencing and variants calling

The same methods and pipeline were used to sequence the AI and the SFFS cohorts, and has been previously published (Bizon et al., 2014). In brief, blood derived DNA was sequenced using Illumina low-coverage whole genome sequencing. Reads from WGS were aligned to GRCh37/hg19 human reference genome using BWA, and realigned near indels with GATK (DePristo et al., 2011). Variants were called using both GATK Unified Genotyper following the best practices for low-coverage samples (Van der Auwera et al., 2002) and imputation was carried out using the LD-aware variant caller Thunder (Li et al., 2011). The variants for each cohort were called separately.

Candidate genomic region

We tested for associations between AUD phenotypes and variants across the genomic regions around the *ADH* genes. The sequence map coordinates for the analyzed region on chromosome 4 include the *ADH* gene cluster region *ADH7-ADH1C-ADH1B-ADH1A-ADH6-ADH4-ADH5* and the flanking 10,000 basepairs (bp) were 99,982,130–100,366,894. The region had 3529 single nucleotide variations (SNVs) called for the AI cohort, and 5020 SNVs for the EA cohort. There were 1657 and 1640 variants with at least 1% minor allele frequency (MAF) for AI and EA respectively.

ADH region single variant association analysis

While SFFS cohort is primarily composed of European Americans, most AI participants have an admixed ancestral background (Norden-Krichmar et al., 2014, Peng et al., 2014). Additionally, both cohorts have family structures. A linear mixed model as implemented in EMMAX (Kang et al., 2010) was used for single variant association analysis to incorporate both population structures and familial relatedness. Age, age-squared and sex were included as covariates in all analyses. We also tested differential effects of the resulting associations between the two cohorts using a two-sample Z-test (Cohen et al., 2003). Full details are given in Supplemental Information.

Multiple testing adjustments

Many variants in the candidate *ADH* regions are in linkage disequilibrium (LD), or in other words, correlated, especially given that genomic sequencing data were used. In addition, the AUD phenotypes are also correlated. A simple Bonferroni correction would thus be overly conservative. To correct for multiple comparisons, we first computed the effective number of independent variants (M_{eff}) using a principle component approach as implemented in simpleM (Gao et al., 2010). We then obtained the effective number of independent traits (Nyholt, 2004). Because we were interested in identifying variants in the *ADH* region that might potentially have functional effects on any of the AUD phenotypes, we adopted a hierarchical testing approach for error control (Peterson et al., 2016). Briefly, the association tests performed on each variant is treated as a family of hypotheses; for each variant, a P value is obtained for the intersection of individual hypotheses. False discovery rate (FDR) adjustment (referred to as vFDR) is then applied to the collection of the p -values with an FDR target error rate q_1 . Once a set of variants is selected, within the family of each selected variant, the FDR procedure with adjustment for selection bias (referred to as sFDR) is

applied to test individual hypotheses with a target level q_2 . At each stage of the testing procedure, we used the M_{eff} -based FDR procedure (Li and Ji, 2005). We report results for $q_1 = q_2 = 0.1$.

For single variant association analyses, we included all variants with MAF over 1% in either cohort and derive M_{eff} from these variants. Of 1657 and 1640 total variants for the NA and EA cohorts, the effective number of independent variants turned out to be 358 and 425 respectively. The effective number of independent traits, computed over the phenotypes residualized over covariates, is 1.32 and 1.35 for AI and EA.

Variant selections and multiple-variant association

As many variants in the *ADH* region are correlated, their effects are in general not independent from each other. Moreover, while single-variant association tests may be effective in identifying individual variants having relatively the largest effects, variants that individually have smaller effects yet collectively might contribute to the phenotypic variance can easily go undetected. One way to model collective associations from many variants is through multiple-regression. Given the much larger number of variants relative to the number of samples, we need to regularize the regression to reduce the number of predictors. Because we were primarily interested in finding a subset of *ADH* variants that were most associated with AUD phenotypes, we used generalized least squares with lasso on variants across the whole *ADH* region (Tibshirani, 1994). Ten rounds of 10-fold cross-validation were used to select the penalty parameter λ for each phenotype and cohort.

Liver eQTL analysis and functional predictions

Liver-specific *cis*-eQTL data were obtained from the Genotype-Tissue Expression (GTEx) Project version V6p (Aguet et al., 2016). Only *ADH* genes were considered. In addition to the significant variants from the association analysis, the SNPs in LD with the significant variants were also examined. The threshold for LD was $D' \geq 0.5$ and $R^2 \geq 0.1$ (a low R^2 threshold allows for large allele frequency differences between variants). If a variant resides within an exon of a gene, PolyPhen-2 was used to assess the potential functional impact of the variant (Adzhubei et al., 2010).

Rare variant association analysis

The sequence kernel association test (SKAT) (Wu et al., 2011) implemented within a linear mixed model was used for rare variant tests of association. We tested for each genes separately and also for the entire region. For each *ADH* gene, we included all its rare variants (MAF $\leq 1\%$). For the region, we included all rare variants in the *ADH* region with MAF $\leq 1\%$. Since not all variants in this region are potentially causal and effects may be in either direction, SKAT that uses a multiple regression model is presumably more powerful than other burden tests. The kernel function was the default beta distribution density function $Beta(1,25)$. For gene-based rare variant test, we assumed that the 7 genes were independent and used the same hierarchical testing approach described above to correct for multiple comparisons.

Full details are given in Supplemental Information: Materials and Methods.

Results

No variant in the *ADH* region was significantly associated with either of the two AUD related phenotypes for the Euro-American SFFS cohort (see Table 3). For the American Indians, variant rs187709743, 3.7kb downstream of *ADH4*, is the top variant in the *ADH* region that was associated with both the alcohol-related-life-events and the alcohol-dependence-drinking-symptoms phenotypes (variant discovery vFDR = 0.003, see Figure 1 to visualize the variant locations, and Figure S1 for LocusZoom plots). At slightly less significant levels, variant rs116822782 upstream of *ADH4* is also associated with both phenotypes. Two additional variants, downstream and upstream of *ADH1C* respectively, are only significantly associated with the alcohol-related-life-events (Table 3 and Figure 1). As described in the Methods section, the FDRs were obtained by a hierarchical approach adjusting the nominal *p* values using M_{eff} -based FDR procedures at each stage. All significant association effects for AIs are positive, indicating that the variants are potentially risk alleles. Note that with the exception of the rare variant rs116822782 whose allele frequency is around 1% in both cohorts, the other three SNPs all have much higher allele frequencies in AI than in EA. Of the two top variants, rs187709743 and rs55960155, while they both are rather common in the AIs (MAF=14.8%, 8.4% respectively), the former is rare (MAF=0.7%) and the latter absent in the EAs.

Because none of the significant findings were within the genes themselves, in order to investigate whether these variants might confer any functional effects, we examined the liver-specific *cis*-eQTLs effects of the *ADH* genes and the significant variants or those in LD with them. The top SNP-gene associations are listed in Table S2. Of the four variants that were associated with AUD in AI (Table 3), the most significant liver eQTL was found on SNPs that are in LD with rs187709743 ($D' = 1$, $R^2 = 0.15-0.29$) for genes *ADH5* and *ADH4* (Table S2). Additionally, SNPs in LD with rs55960155 ($D' = 0.98-1$, $R^2 = 0.12-0.13$) were most significantly associated with *ADH1C* expression in liver. The results of differential association analysis between the two cohorts are described in the Supplemental Results and listed in Tables S4 and S5.

Using the regularized multiple-regression analysis, 37 variants in the *ADH* region were selected in AI for the alcohol-related-life-events phenotype ($p\text{-value} = 1.1 \times 10^{-11}$ for multiple-regression using selected variants) and 33 for the alcohol-dependence-drinking-symptoms ($p = 5.8 \times 10^{-10}$), of which 13 were shared. In the EA cohort, 20 variants ($p = 4.8 \times 10^{-8}$) and 21 variants ($p = 5.7 \times 10^{-10}$) were selected for the two phenotypes respectively, of which eight were shared (see Figure 1 and Table S3). Note that the *p*-values reported here were biased because variants were already selected for relevance. Only two variants were selected in both cohorts: a new variant chr4:100068799 at 3,350 bp upstream of *ADH4* and chr4:100092848 (rs60239449) at intergenic region between *ADH4* and *ADH6*. Of variants selected in the AI cohort, 11 were absent from the EA cohort, of which four were common (MAF $\geq 5\%$) in the AI; of those selected in the EA, one variant was absent from the AI. Three out of the 26 variants (30%) selected that were located within the genes were on exons.

We further conducted rare variant analysis. For gene-based tests (see Table S6), *ADH5* was the most significant (nominal $p = 0.027$) for alcohol-related-life-events in AI, and *ADH4*

was the most significant for alcohol-dependence-drinking symptoms for both cohorts (nominal $p = 0.081$ for AI, 0.047 for SFFS). However, no gene remained significant after multiple-testing corrections. Because all of the observed significant single variant associations and the majority of the selected variants (70%) in the two cohorts were either upstream or downstream from an *ADH* gene or in the intergenic regions, we did another test that included all rare variants with MAF $\geq 1\%$ across the entire *ADH* region. There were 1865 and 3043 rare variants for AI and SFFS respectively. Nearly all participants had some rare variants in the region. That SFFS EA cohort had many more rare variants in the region is likely due to its larger sample size. The only association with a nominal p -value less than 0.05 ($p = 0.043$) was observed for the alcohol-related-life-events in AI. Once corrected for multiple traits, its FDR is 0.057 (Table 4).

Discussion

The primary aims of this study were to assess common and rare genomic variations in the chromosomal region containing the seven alcohol dehydrogenase genes (*ADH7-ADH1C-ADH1B-ADH1A-ADH6-ADH4-ADH5*) that encode the primary alcohol metabolic enzyme family in an American Indian and a European American population, and to test their associations with alcohol use disorder symptomatology. We conducted single variant, multi-variant and rare-variant association analyses on the number of alcohol dependence drinking symptoms and weighted alcohol-related life events.

Four *ADH* variants were significantly associated with the alcohol-related-life-events phenotype in AIs, while no variant was significantly associated with either phenotype in EAs. They are upstream or downstream of *ADH4* and *ADH1C*. The two variants near *ADH4* were also significantly associated with the alcohol-dependence-drinking-symptoms. Of note, the most significant variant, rs187709743, is common in AI at near 15% MAF but rare in EA with 0.7% MAF. Although there was no direct eQTL data available for the significant variants, those in strong LD with rs187709743 exhibited *cis*-eQTL signals for *ADH4* and *ADH5* in the liver (Table S2). It appears that the haplotypes containing the minor allele of rs187709743 may increase the expression of *ADH5* and decrease the expression of *ADH4*. *ADH5* is expressed in most tissues and encodes χ -ADH that has very low affinity for ethanol. Its effects on the risk for AUD, however, is unclear. *ADH4* is expressed primarily in the liver and the cornea, and encodes π -ADH that contributes to ethanol oxidation at high or moderate concentrations. Variants in *ADH4* and in the intergenic region between *ADH4* and *ADH5* have been found to be associated with alcohol dependence in the COGA cohort, which is mainly comprised of European- and African-Americans, using the pedigree disequilibrium test (Edenberg et al., 2006) and in another Euro-American population using structured association tests (Luo et al., 2005b). Of the 12 significant SNPs in (Edenberg et al., 2006) and three significant SNPs in (Luo et al., 2005b), 14 were in our cohorts. All 14 variants are in LD with rs187709743 ($D' = 1$, R^2 varies and is small in general due to allele frequency differences) in AI, and the minor allele of rs187709743 is on the same haplotypes with the major alleles of these variants with the exception of rs4148886 (MAF=44%). In addition, the minor allele of variant rs3762894 has been reported to be protective against alcohol withdrawal in the same AI population (Gizer et al., 2011a). In the present AI cohort, rs187709743 and rs3762894 (MAF=11%) are in LD ($D' = 1$), and the minor allele of

rs187709743 and the major allele of rs3762894 are on the same haplotypes. This evidence consistently suggests that haplotypes containing the minor allele of rs187709743 might confer risk for AUD by reducing the expression level of *ADH4* in liver, which might potentially slow down the oxidation of ethanol (and the accumulation of acetaldehyde) at high concentrations, thus mitigating the adverse effect of acetaldehyde. These effects are likely unique to admixed AI populations, as rs187709743 is rare in the EA cohort and in general rare or absent in all but admixed American (AMR, MAF=11%) populations in the 1000 Genome database (Siva, 2008).

Variant rs55960155 is common (MAF=8.4%) in the AI cohort and absent from the EA samples altogether. Variants in strong LD with rs55960155 showed eQTL signals for *ADH1C* and *ADH6* in liver (Table S2). It appears that haplotypes containing the minor allele of rs55960155 may decrease the expression level of *ADH1C* and increase that of *ADH6*. While not much is known about the functional aspect of *ADH6*, there is substantial evidence suggesting that the *ADH1C*1* allele is protective against alcohol dependence in populations such as East Asian (Matsushita and Higuchi, 2017, Higuchi et al., 1995, Thomasson et al., 1991, Thomasson et al., 1994), European (Kuo et al., 2008) and American Indian (Mulligan et al., 2003), although the effect may be due to the LD between *ADH1C*1* and *ADH1B*2*. In addition, *ADH1C*2* was shown to increase the risk for alcoholism in a Mexican American sample (Konishi et al., 2003). In this AI cohort, the minor allele of rs55960155 appears on the same haplotypes with *ADH1C*2* (AF=41%) and *ADH1B*1* (AF=91%) alleles, although neither *ADH1C*2* nor *ADH1B*1* is a significant risk allele in AI (nominal *p*-values were between 0.025—0.030, and were not significant after correction for multiple comparisons). There was additional evidence to suggest that *ADH1C* gene expression differed by two-fold between various haplotypes in a region located from 3,496 to 3,008 bp upstream of the transcription start site of the gene (Chen et al., 2005), and rs55960155 is in this region, being 3,120 bp upstream of *ADH1C*. Taken together, it suggests that haplotypes containing the minor allele of rs55960155 might increase the risk for AUD in AIs by reducing the gene expression of *ADH1C* in liver. The most significant variants for the EA cohort were in or near *ADH7* but none remained significant once corrected for multiple comparisons (Table 3).

In addition to the single variant results, 57 and 33 genomic variants across the *ADH* gene cluster region were selected in the AI and the EA cohorts, respectively, to be associated with at least one of the AUD phenotypes through regularized multi-variant regressions (Figure 1 and Table S3). Two of the top variants around *ADH4* in the single variant association tests were also in the selected subset in AI. The two top variants in and near *ADH7*, although not significant in the single variant association tests for EA, were in the selected subset for EA. Although most of the selected variants in EA were co-located in the same regions with those in AI, all but two variants were distinct to the respective cohorts. That only three variants reside within exons suggests that the potential functional impacts of most of the associated variants are regulatory. Of the three exonic variants, rs6413444 on *ADH1C* is synonymous and associated with alcohol-related-life-events in EA. The other two are nonsynonymous: rs1229984 was associated with both phenotypes in AI, and rs1573496 with alcohol-related-life-events in AI. Notice that the A-allele of variant rs1229984 corresponds to the *ADH1B*2* allele that is protective against AUD in East Asian populations. The allele frequency of this

protective variant, however, was only 4% and 1% in the AI and the EA cohorts respectively. Variant rs1573496 is located in an exon of *ADH7* and predicted by PolyPhen-2 to be probably damaging with a probability score of 0.973. Of note, variant rs1800759 in the promoter region of *ADH4* (Edenberg et al., 2006, Luo et al., 2005a) was selected in association with alcohol-related-life-events in EA. Overall, largely distinct subsets of variants were selected for AI and EA respectively although some of them were in the same regions.

Ethanol is primarily oxidized in liver. Given the *ADH* enzyme concentrations and properties in liver, it has been theorized that the class I enzymes encoded by *ADH1A/B/C* and class II enzyme encoded by *ADH4* contribute the most to alcohol metabolism (Hurley et al., 2002). Apart from the well-known *ADH1B* alleles that have different frequencies in different ancestral populations resulting in protection effects against alcoholism in populations such as East Asian and African, the other *ADH* genes that are found more prominently associated with AUD are *ADH4* and *ADH1C*. These are consistent with our findings. In association studies across populations and cohorts, the desire is often to identify variants shared by different samples that contribute to a phenotype, either through replication or by combining the cohorts directly, or indirectly through meta-analysis. The increased statistical power gained by these approaches aid in the discovery process. It is however equally important to determine how genomic variations may differentially affect populations. The differential effects not only provide insights into risk or protection for the disorders across ethnic groups but may also have potential clinical relevance. Moreover, for admixed populations such as American Indians, alleles from different ancestral backgrounds might come together to affect risk. Indeed, we have identified variants that are either unique or more common in American Indians than in the European Americans that might confer further risk for AUD by potentially reducing the levels of the two most relevant classes of *ADH* enzymes. All of the variants we identified in the single variant associations are upstream or downstream from a gene or in the intergenic regions, implying their potential involvement in regulatory functions, which is also consistent with our findings in the multi-variant associations. Although they are in LD with variants in the coding regions, those variants themselves did not exhibit significant individual associations with the AUD phenotypes in either AI or EA.

There are several limitations to our study. These cohorts do not represent all American Indians or all European Americans as both samples were not ascertained using random sampling techniques. The AI cohort is a community sample. The EA cohort is comprised of semi-targeted recruits who met certain screening criteria for alcohol dependence, and a smaller set of unselected general population sample with similar demographics (Gizer et al., 2011b). As a result, the two cohorts have a large mean age difference (Table 1) and the EA cohort has on average higher numbers of alcohol-related life events and alcohol-dependence drinking symptoms (Table S1). These differences could have potentially influenced the study results, even though age and age-squared were included as covariates, and with quantitative traits there is no need to adjust for prevalence with linear regressions. There might also be a potential violation of the homoscedasticity assumption. The sample sizes were moderate in both cohorts relative to the number of variants we tested, even after taking variant correlations into consideration. Although we used the effective number of independent variants in multiple testing corrections, the computation was based on correlations between

variants that are usually low when variants are of rather different frequencies. Thus, their actual independence is somewhat obscured. This might explain why no association remained significant in the single variant analysis for the EA cohort even though previous works have reported significant relations between individual *ADH* variants and alcoholism in Euro-American samples; nonetheless, those studies were either based on linkage analysis or using a relatively small number of genotyped variants. Similarly, the number of significant variants in *ADH* were also much fewer than our previous work in AI, although many of the variants are in LD with each other. Additionally, because many variants in the *ADH* region are in high LD, it is difficult to infer the actual causal variants with respect to AUD.

Limitations notwithstanding, the present study has identified individual *ADH* variants that might confer risk for alcohol-related symptomatology in American Indians but not for European Americans. These variants are either unique to the AIs or more common in the AIs than in the EAs. When *ADH* variants were considered collectively, largely distinct subsets of common and less frequent variants were selected for AI and EA, respectively, yet many of them co-located with each other in the same regions. The rare variants across *ADH* on the other hand seem to contribute to alcohol symptomatology in AI but not in EA. Our results suggest that *ADH* variants, both common and rare are more likely to impact risk for alcohol-related symptomatology in this AI population than in this EA sample; and variants that might affect AUD are likely different although often convergent onto the same genomic regions between the two populations. We remark however that *ADH* variants only explain a small fraction of the potential genetic contribution to alcohol symptomatology. Further, overall genetic variance can only explain part of the AUD variations. Unmeasured risk factors, some specific to American Indians/Alaska Natives or to European Americans, could also potentially explain some of the relationship between AUD and *ADH* variants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: Funding for this study was provided by grants from National Institutes of Health (NIH): National Institute on Alcohol Abuse and Alcoholism (NIAAA) K25 AA025095 to QP, NIAAA and National Center on Minority Health and Health Disparities (NCMHD) 5R37 AA010201 to CLE, and National Institute on Drug Abuse (NIDA) R01 DA030976 to C.L.E., IRG, K.C.W.

We would like to acknowledge and thank the following people for their roles in 1) the genotyping effort: Chris Bizon, Scott Chasse, Piotr Mieczkowski, Ewa Patrycja Malc, Joshua Sailsbery, and Phil Owens; 2) ancestry analysis: Ondrej Libiger; and 3) recruiting participants and collection and preparation of the clinical data: David Gilder, Corinne Kim, Evie Phillips, Gina Stouffer, Susan Lopez, Linda Corey and Derek Wills.

Funding for this study was provided by grants from National Institutes of Health (NIH): National Institute on Alcohol Abuse and Alcoholism (NIAAA) K25 AA025095 to QP, NIAAA and National Center on Minority Health and Health Disparities (NCMHD) 5R37 AA010201 to CLE, and National Institute on Drug Abuse (NIDA) R01 DA030976 to C.L.E., IRG, K.C.W. NIAAA, NCMHD and NIDA had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

References

- ADZHUBEI IA, SCHMIDT S, PESHKIN L, RAMENSKY VE, GERASIMOVA A, BORK P, KONDRASHOV AS, SUNYAEV SR. A method and server for predicting damaging missense mutations. *Nature methods*. 2010; 7:248–249. [PubMed: 20354512]
- AGUET F, BROWN AA, CASTEL S, DAVIS JR, MOHAMMADI P, SEGRE AV, ZAPPALA Z, ABELL NS, FRESARD L, GAMAZON ER, GELFAND E, GLOUDEMANS MJ, HE Y, HORMOZDIARI F, LI X, LI X, LIU B, GARRIDO-MARTIN D, ONGEN H, PALOWITCH JJ, PARK Y, PETERSON CB, QUON G, RIPKE S, SHABALIN AA, SHIMKO TC, STROBER BJ, SULLIVAN TJ, TERAN NA, TSANG EK, ZHANG H, ZHOU Y-H, BATTLE A, BUSTAMONTE CD, COX NJ, ENGELHARDT BE, ESKIN E, GETZ G, KELLIS M, LI G, MACARTHUR DG, NOBEL AB, SABBATI C, WEN X, WRIGHT FA, LAPPALAINEN T, ARDLIE KG, DERMITZAKIS ET, BROWN CD, MONTGOMERY SB. Local genetic effects on gene expression across 44 human tissues. 2016 bioRxiv.
- AMERICAN PSYCHIATRIC ASSOCIATION. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC: American Psychiatric Association; 2013.
- AMERICAN PSYCHIATRIC ASSOCIATION TASK FORCE ON DSM-IV. Diagnostic and Statistical Manual of Mental Disorder (DSM-IV). Washington, DC: American Psychiatric Association; 1994.
- BIZON C, SPIEGEL M, CHASSE SA, GIZER IR, LI Y, MALC EP, MIECZKOWSKI PA, SAILSBERY JK, WANG X, EHLERS CL, WILHELMSSEN KC. Variant calling in low-coverage whole genome sequencing of a Native American population sample. *BMC Genomics*. 2014; 15:85. [PubMed: 24479562]
- BOSRON WF, EHRIG T, LI TK. Genetic factors in alcohol metabolism and alcoholism. *Semin Liver Dis*. 1993; 13:126–35. [PubMed: 8337601]
- BUCHOLZ KK, CADORET R, CLONINGER CR, DINWIDDIE SH, HESSELBROCK VM, NURNBERGER JI JR, REICH T, SCHMIDT I, SCHUCKIT MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol*. 1994; 55:149–58. [PubMed: 8189735]
- CARR LG, XU Y, HO WH, EDENBERG HJ. Nucleotide Sequence of the ADH23 Gene Encoding the Human Alcohol Dehydrogenase β 3 Subunit. *Alcoholism: Clinical and Experimental Research*. 1989; 13:594–596.
- CHEN HJ, TIAN H, EDENBERG HJ. Natural haplotypes in the regulatory sequences affect human alcohol dehydrogenase 1C (ADH1C) gene expression. *Human Mutation*. 2005; 25:150–155. [PubMed: 15643610]
- CHEN YC, PENG GS, WANG MF, TSAO TP, YIN SJ. Polymorphism of ethanol-metabolism genes and alcoholism: correlation of allelic variations with the pharmacokinetic and pharmacodynamic consequences. *Chem Biol Interact*. 2009; 178:2–7. [PubMed: 19014920]
- COHEN, J., COHEN, P., WEST, SG., AIKEN, LS. Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers; 2003.
- COMPTON WM, THOMAS YF, STINSON FS, GRANT BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry*. 2007; 64:566–76. [PubMed: 17485608]
- DEPRISTO MA, BANKS E, POPLIN R, GARIMELLA KV, MAGUIRE JR, HARTL C, PHILIPPAKIS AA, DEL ANGEL G, RIVAS MA, HANNA M, MCKENNA A, FENNELL TJ, KERNYTSKY AM, SIVACHENKO AY, CIBULSKIS K, GABRIEL SB, ALTSHULER D, DALY MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*. 2011; 43:491–8. [PubMed: 21478889]
- EDENBERG HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health*. 2007; 30:5–13. [PubMed: 17718394]
- EDENBERG HJ, FOROUD T. REVIEW: The genetics of alcoholism: identifying specific genes through family studies. *Addiction Biology*. 2006; 11:386–396. [PubMed: 16961766]
- EDENBERG HJ, XUEI X, CHEN HJ, TIAN H, WETHERILL LF, DICK DM, ALMASY L, BIERUT L, BUCHOLZ KK, GOATE A, HESSELBROCK V, KUPERMAN S, NURNBERGER J,

- PORJESZ B, RICE J, SCHUCKIT M, TISCHFIELD J, BEGLEITER H, FOROUD T. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Human Molecular Genetics*. 2006; 15:1539–1549. [PubMed: 16571603]
- EHLERS CL, GILDER DA, WALL TL, PHILLIPS E, FEILER H, WILHELMSSEN KC. Genomic screen for loci associated with alcohol dependence in Mission Indians. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2004a; 129B:110–115.
- EHLERS CL, GIZER IR. Evidence for a genetic component for substance dependence in Native Americans. *American Journal of Psychiatry*. 2013; 170:154–164. [PubMed: 23377636]
- EHLERS CL, GIZER IR, GILDER DA, WILHELMSSEN KC. Linkage analyses of stimulant dependence, craving, and heavy use in American Indians. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2011; 156B:772–80.
- EHLERS CL, SPENCE JP, WALL TL, GILDER DA, CARR LG. Association of ALDH1 promoter polymorphisms with alcohol-related phenotypes in southwest California Indians. *Alcohol Clin Exp Res*. 2004b; 28:1481–6. [PubMed: 15597079]
- EHLERS CL, STOUFFER GM, COREY L, GILDER DA. The clinical course of DSM-5 alcohol use disorders in young adult native and Mexican Americans. *The American Journal on Addictions*. 2015; 24:713–721. [PubMed: 26346282]
- EHLERS CL, WALL TL, BETANCOURT M, GILDER DA. The clinical course of alcoholism in 243 Mission Indians. *American Journal of Psychiatry*. 2004c; 161:1204–10. [PubMed: 15229052]
- GAO X, BECKER LC, BECKER DM, STARMER JD, PROVINCE MA. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic epidemiology*. 2010; 34:100–105. [PubMed: 19434714]
- GELERNTER J, KRANZLER HR, SHERVA R, ALMASY L, KOESTERER R, SMITH AH, ANTON R, PREUSS UW, RIDINGER M, RUJESCU D, WODARZ N, ZILL P, ZHAO H, FARRER LA. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry*. 2014; 19:41–9. [PubMed: 24166409]
- GILDER DA, LAU P, COREY L, EHLERS CL. Factors associated with remission from alcohol dependence in an American Indian community group. *Am J Psychiatry*. 2008; 165:1172–8. [PubMed: 18519526]
- GILDER DA, WALL TL, EHLERS CL. Comorbidity of select anxiety and affective disorders with alcohol dependence in southwest California Indians. *Alcohol Clin Exp Res*. 2004; 28:1805–13. [PubMed: 15608596]
- GIZER IR, EDENBERG HJ, GILDER DA, WILHELMSSEN KC, EHLERS CL. Association of Alcohol Dehydrogenase Genes with Alcohol-Related Phenotypes in a Native American Community Sample. *Alcoholism: Clinical and Experimental Research*. 2011a; 35:2008.
- GIZER IR, EHLERS CL, VIETEN C, SEATON-SMITH KL, FEILER HS, LEE JV, SEGALL SK, GILDER DA, WILHELMSSEN KC. Linkage scan of alcohol dependence in the UCSF Family Alcoholism Study. *Drug and Alcohol Dependence*. 2011b; 113:125–132. [PubMed: 20817416]
- GRANT BF, GOLDSTEIN RB, SAHA TD, CHOU SP, JUNG J, ZHANG H, PICKERING RP, RUAN WJ, SMITH SM, HUANG B, HASIN DS. Epidemiology of DSM-5 alcohol use disorder: Results from the national epidemiologic survey on alcohol and related conditions III. *JAMA Psychiatry*. 2015; 72:757–766. [PubMed: 26039070]
- HESELBROCK M, EASTON C, BUCHOLZ KK, SCHUCKIT M, HESSELBROCK V. A validity study of the SSAGA--a comparison with the SCAN. *Addiction*. 1999; 94:1361–70. [PubMed: 10615721]
- HIGUCHI S, MATSUSHITA S, MURAYAMA M, TAKAGI S, HAYASHIDA M. Alcohol and aldehyde dehydrogenase polymorphisms and the risk for alcoholism. *American Journal of Psychiatry*. 1995; 152:1219–1221. [PubMed: 7625477]
- HURLEY TD, EDENBERG HJ, BOSRON WF. Expression and kinetic characterization of variants of human beta 1 beta 1 alcohol dehydrogenase containing substitutions at amino acid 47. *Journal of Biological Chemistry*. 1990; 265:16366–16372. [PubMed: 2398055]
- HURLEY, TD., EDENBERG, HJ., LI, T-K. *Pharmacogenomics*. Wiley-VCH Verlag GmbH & Co; KGaA: 2002. *Pharmacogenomics of Alcoholism*.

- KANG HM, SUL JH, SERVICE SK, ZAITLEN NA, KONG SY, FREIMER NB, SABATTI C, ESKIN E. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*. 2010; 42:348–354. [PubMed: 20208533]
- KONISHI T, CALVILLO M, LENG AIS, FENG J, LEE T, LEE H, SMITH JL, SIAL SH, BERMAN N, FRENCH S, EYSSELEIN V, LIN KM, YVONNE WAN YUJ. The ADH3*2 and CYP2E1 c2 alleles increase the risk of alcoholism in Mexican American men. *Experimental and Molecular Pathology*. 2003; 74:183–189. [PubMed: 12710951]
- KUO PH, KALSI G, PRESCOTT CA, HODGKINSON CA, GOLDMAN D, VAN DEN OORD EJ, ALEXANDER J, JIANG C, SULLIVAN PF, PATTERSON DG, WALSH D, KENDLER KS, RILEY BP. Association of ADH and ALDH Genes With Alcohol Dependence in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) Sample. *Alcoholism: Clinical and Experimental Research*. 2008; 32:785–795.
- LI J, JI L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*. 2005; 95:221–227. [PubMed: 16077740]
- LI Y, SIDORE C, KANG HM, BOEHNKE M, ABECASIS GR. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Research*. 2011; 21:940–51. [PubMed: 21460063]
- LUO X, KRANZLER HR, ZUO L, LAPPALAINEN J, YANG BZ, GELERNTER J. ADH4 Gene Variation is Associated with Alcohol Dependence and Drug Dependence in European Americans: Results from HWD Tests and Case-Control Association Studies. *Neuropsychopharmacology*. 2005a; 31:1085–1095.
- LUO X, KRANZLER HR, ZUO L, YANG BZ, LAPPALAINEN J, GELERNTER J. ADH4 gene variation is associated with alcohol and drug dependence: results from family controlled and population-structured association studies. *Pharmacogenet Genomics*. 2005b; 15:755–768. [PubMed: 16220108]
- MACGREGOR S, LIND PA, BUCHOLZ KK, HANSELL NK, MADDEN PAF, RICHTER MM, MONTGOMERY GW, MARTIN NG, HEATH AC, WHITFIELD JB. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum Mol Genet*. 2009; 18:580–93. [PubMed: 18996923]
- MATSUSHITA S, HIGUCHI S. Use of Asian samples in genetic research of alcohol use disorders: Genetic variation of alcohol metabolizing enzymes and the effects of acetaldehyde. *The American Journal on Addictions*. 2017 n/a-n/a.
- MULLIGAN CJ, ROBIN RW, OSIER MV, SAMBUUGHIN N, GOLDFARB LG, KITTLES RA, HESSELBROCK D, GOLDMAN D, LONG JC. Allelic variation at alcohol metabolism genes (ADH1B, ADH1C, ALDH2) and alcohol dependence in an American Indian population. *Human Genetics*. 2003; 113:325–336. [PubMed: 12884000]
- NORDEN-KRICHMAR TM, GIZER IR, LIBIGER O, WILHELMSSEN KC, EHLERS CL, SCHORK NJ. Correlation analysis of genetic admixture and social identification with body mass index in a Native American community. *Am J Hum Biol*. 2014; 26:347–60. [PubMed: 24757035]
- NYHOLT DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics*. 2004; 74:765–9. [PubMed: 14997420]
- PARK BL, KIM JW, CHEONG HS, KIM LH, LEE BC, SEO CH, KANG TC, NAM YW, KIM GB, SHIN HD, CHOI IG. Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Human Genetics*. 2013; 132:657–668. [PubMed: 23456092]
- PENG Q, GIZER IR, LIBIGER O, BIZON C, WILHELMSSEN KC, SCHORK NJ, EHLERS CL. Association and ancestry analysis of sequence variants in ADH and ALDH using alcohol-related phenotypes in a Native American community sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2014; 165:673–683.
- PETERSON CB, BOGOMOLOV M, BENJAMINI Y, SABATTI C. Many Phenotypes Without Many False Discoveries: Error Controlling Strategies for Multitrait Association Studies. *Genetic Epidemiology*. 2016; 40:45–56. [PubMed: 26626037]
- PRESCOTT CA, SULLIVAN PF, KUO PH, WEBB BT, VITTUM J, PATTERSON DG, THISELTON DL, MYERS JM, DEVITT M, HALBERSTADT LJ, ROBINSON VP, NEALE MC, VAN DEN

- OORD EJ, WALSH D, RILEY BP, KENDLER KS. Genomewide linkage study in the Irish affected sib pair study of alcohol dependence: evidence for a susceptibility region for symptoms of alcohol dependence on chromosome 4. *Mol Psychiatry*. 2006; 11:603–611. [PubMed: 16534506]
- REICH T, EDENBERG HJ, GOATE A, WILLIAMS JT, RICE JP, VAN EERDEWEGH P, FOROUD T, HESSELBROCK V, SCHUCKIT MA, BUCHOLZ K, PORJESZ B, LI TK, CONNEALLY PM, NURNBERGER JI, TISCHFIELD JA, CROWE RR, CLONINGER CR, WU W, SHEARS S, CARR K, CROSE C, WILLIG C, BEGLEITER H. Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet*. 1998:81.
- SEATON KL, CORNELL JL, WILHELMSSEN KC, VIETEN C. Effective strategies for recruiting families ascertained through alcoholic probands. *Alcohol Clin Exp Res*. 2004; 28:78–84. [PubMed: 14745304]
- SIVA N. 1000 Genomes project. *Nat Biotechnol*. 2008; 26:256.
- THOMASSON HR, CRABB DW, EDENBERG HJ, LI TK, HWU HG, CHEN CC, YEH EK, YIN SJ. Low Frequency of the ADH2*2 Allele among Atayal Natives of Taiwan with Alcohol Use Disorders. *Alcoholism: Clinical and Experimental Research*. 1994; 18:640–643.
- THOMASSON HR, EDENBERG HJ, CRABB DW, MAI XL, JEROME RE, LI TK, WANG SP, LIN YT, LU RB, YIN SJ. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *American Journal of Human Genetics*. 1991; 48:677–681. [PubMed: 2014795]
- TIBSHIRANI R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B*. 1994; 58:267–288.
- VAN DER AUWERA, GA., CARNEIRO, MO., HARTL, C., POPLIN, R., DEL ANGEL, G., LEVY-MOONSHINE, A., JORDAN, T., SHAKIR, K., ROAZEN, D., THIBAUT, J., BANKS, E., GARIMELLA, KV., ALTSHULER, D., GABRIEL, S., DEPRISTO, MA. Current Protocols in Bioinformatics. John Wiley & Sons, Inc; 2002. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline.
- VIETEN C, SEATON KL, FEILER HS, WILHELMSSEN KC. The University of California, San Francisco Family Alcoholism Study. I. Design, Methods, and Demographics. *Alcoholism: Clinical and Experimental Research*. 2004; 28:1509–1516.
- WALL TL, CARR LG, EHLERS CL. Protective association of genetic variation in alcohol dehydrogenase with alcohol dependence in Native American Mission Indians. *American Journal of Psychiatry*. 2003; 160:41–6. [PubMed: 12505800]
- WU MC, LEE S, CAI T, LI Y, BOEHNKE M, LIN X. Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*. 2011; 89:82–93. [PubMed: 21737059]

Figure 1.
ADH variants selected through regularized multi-variant association with AUD phenotypes for the American Indian (red lines) and the Euro-American (blue lines) cohorts. The top two rows of solid lines are associated with alcohol-related-life-events and the bottom two rows of dotted lines are with alcohol-dependence-drinking-symptoms. The short black lines are the top variants from the single-variant association (see Table 3).

Table 1

Demographics of the two independent population samples.

Cohort	American Indian (AI)	Euro-American SFFS (EA)
Samples	742	1215
Families	170	763
Age ^a (yrs)	31.2±13.2 [18, 82]	47.7±12.1 [17, 84]
Gender	M: 317, F: 425	M: 625, F: 690
Ethnicity	Native American (Admixed)	European American 95% African American 3% Native American 1% Others 1%

^aIn the format of mean ± standard deviation [min, max]

Table 2
Life events of alcohol use disorders, from which the weighted alcohol-related-life-events phenotype is derived.

Severity weight 1	Severity weight 2	Severity weight 3
1. Arguments	13. Binges	25. Arrested for alcohol related behavior
2. Physical fights	14. Tolerance	26. Problems in love relationship
3. Problems at work/school	15. Interfered with work	27. Considered excessive drinker
4. Problems with family, friends	16. Self injury while drunk	28. Guilt
5. Hitting others without fighting	17. Decreased important activities	29. Wanted to quit 3+ times
6. Objections from family, friends	18. Inability to change drinking behavior	30. Withdrawal
7. Drank while in hazardous situations	19. Morning drinking	31. Unable to quit/cut down
8. Drank when not intended	20. Drank more than intended	32. Arrested for DUI
9. Lost friends	21. Used rules for drinking	33. Shakes
10. Blackouts	22. Little time for non-drinking activities	34. Continues despite health problems
11. Hit/threw things	23. Strong desire for alcohol	35. Health problems occurred
12. Hit family member	24. Psychological impairment	36. Sought professional help

Top association results for the *ADH* candidate gene regions for both American Indian (AI) and the Euro-American (EA) cohorts. EA does not have any significant associations.

Position	ref	alt	dbSNP	Gene	Location	PI ^a	vFDR ^b	sFDR ^c	β	MAF _{NA} ^d	MAF _{EA} ^d
American Indian											
100041125	G	C	rs187709743	ADH4	Downstr	C	0.003	7E-04	6.47	0.148	0.007
						S	-	0.002	0.45		
100066370	G	A	rs116822782	ADH4	Upstr	C	0.063	0.055	19.25	0.008	0.013
						S	-	0.023	1.58		
100253409	C	T	rs1497372	ADH1C	Downstr	C	0.064	0.026	3.40	0.365	0.219
100277037	C	T	rs55960155	ADH1C	Upstr	C	0.057	0.017	6.47	0.084	NA
Euro-American (SFFS)											
100347133	A	T	rs80329335	ADH7	Intron	S	0.661	0.832	0.19	0.140	0.240
100359912	A	G	rs1154473	ADH7	Upstr	C	0.439	0.439	2.40	0.323	0.379

^aPhenotype: C: Alcohol-related-life-events, S: Alcohol-dependence-drinking-symptoms;
^bvFDR: FDR for the discovery variants, i.e., the association with any AUD phenotype;
^csFDR: expected value of the average within family false discovery proportions across all selected variants;
^dMAF_{NA}: alternative allele frequency in AI; MAF_{EA}: alternative allele frequency in EA.

Table 4

Rare variants association between all variants of MAF 1% in *ADH* region and alcohol use disorder (AUD) phenotypes using SKAT test, with SKAT prior of $\beta(1,25)$.

Start Position	End Position	Cohort	N	Fraction with Rare	Total SNPs	Pass SNPs	Single SNPs	Alcohol life events		Drinking symptoms	
								FDR ^a	Q stat	FDR	Q stat
99982855	100366641	NA	742	97.7%	3542	1865	709	0.057	4.01E+05	0.143	2.02E+05
99982360	100366641	EA	1215	98.5%	5072	3043	973	0.699	2.20E+05	0.374	4.59E+05

^aFDR: *Meff*-based FDR *p*-value adjustment.