



Published in final edited form as:

Alcohol Clin Exp Res. 2011 February ; 35(2): 304–316. doi:10.1111/j.1530-0277.2010.01346.x.

Haplotype-Based Study of the Association of Alcohol Metabolizing Genes with Alcohol Dependence in Four Independent Populations

Jixia Liu, PhD^{1,*}, Zhifeng Zhou, PhD¹, Colin A. Hodgkinson, PhD¹, Qiaoping Yuan, PhD¹, Pei-Hong Shen, MS¹, Connie J. Mulligan, PhD², Alex Wang, BS^{1,2}, Rebecca R. Gray, PhD², Alec Roy, MD³, Matti Virkkunen, MD^{4,5}, David Goldman, MD¹, and Mary-Anne Enoch, MD¹

¹ Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, NIH, Bethesda, MD 20892

² Department of Anthropology, University of Florida, Gainesville, FL 32610

³ Mental Health and Behavioral Science, Department of Veterans Affairs New Jersey Health Care System, East Orange, N.J.

⁴ Institute of Clinical Medicine, Department of Psychiatry, University of Helsinki, Finland (MV)

⁵ Kellokoski Hosapital, Kellokoski, Finland (MV)

Abstract

Background—Ethanol is metabolized by two rate limiting reactions: alcohol dehydrogenases (ADH) convert ethanol to acetaldehyde, subsequently metabolized to acetate by aldehyde dehydrogenases (ALDH). Approximately 50% of East Asians have genetic variants that significantly impair this pathway and influence alcohol dependence (AD) vulnerability. We investigated whether variation in alcohol metabolism genes might alter the AD risk in four non-East Asian populations by performing systematic haplotype association analyses in order to maximize the chances of capturing functional variation.

Methods—Haplotype-tagging SNPs were genotyped using the Illumina GoldenGate platform. Genotypes were available for 40 SNPs across the ADH genes cluster and 24 SNPs across the two ALDH genes in four diverse samples that included cases (lifetime AD) and controls (no Axis I disorders). The case, control sample sizes were: Finnish Caucasians: 232, 194; African Americans: 267, 422; Plains American Indians: 226, 110; Southwestern American (SW) Indians: 317, 72.

Results—In all four populations, as well as HapMap populations, five haplotype blocks were identified across the ADH gene cluster: (1) *ADH5-ADH4*; (2) *ADH6-ADH1A-ADH1B*; (3) *ADH1C*; (4) intergenic; (5) *ADH7*. The *ALDH1A1* gene was defined by four blocks and *ALDH2* by one block. No haplotype or SNP association results were significant after correction for multiple comparisons; however several results, particularly for *ALDH1A1* and *ADH4*, replicated earlier findings. There was an *ALDH1A1* block 1 and 2 (extending from intron 5 to the 3' UTR) yin yang haplotype (haplotypes that have opposite allelic configuration) association with AD in the Finns driven by SNPs rs3764435 and rs2303317 respectively, and an *ALDH1A1* block 3

Corresponding author: Jixia Liu PhD NIH/NIAAA/LNG 5625 Fishers Lane, Room 3S32, MSC 9412 Bethesda, MD 20892-9412 Tel: 301-846-5309 Fax: 301-480-2839 liujixia@mail.nih.gov.

SUPPLEMENTARY FIGURE 1

Haplotype Block Structure across Four Populations and Three HapMap Populations for the *ALDH* gene on Chromosome 12 Haplotype blocks are outlined. Linkage disequilibrium between SNPs within haplotype blocks was set at $D' \geq 0.80$.

Left panel: HapMap populations: YRI = Yoruban Africans; CEU = U.S. Caucasians; ASN = Japanese / Chinese.

(including the promoter region) yin yang haplotype association in SW Indians driven by 5 SNPs, all in allelic identity. The *ADH4* SNP rs3762894 was associated with AD in Plains Indians.

Conclusions—The systematic evaluation of alcohol metabolizing genes in four non-East Asian populations has shown only modest associations with AD, largely for *ALDH1A1* and *ADH4*. A concentration of signals for AD with *ALDH1A1* yin yang haplotypes in several populations warrants further study.

Keywords

Alcohol dependence; alcohol dehydrogenases (ADH); aldehyde dehydrogenases (ALDH); haplotype association; *ALDH1A1*

INTRODUCTION

Since at least 10,000 B.C., almost all societies have found ways to ferment various substances to produce ethanol for consumption. Ethanol is metabolized largely in the liver by alcohol dehydrogenases (ADH) to acetaldehyde which is then converted to acetate by aldehyde dehydrogenases (ALDH). Since diverging from the ancestral African populations, East Asians (Chinese, Japanese and Koreans) have acquired ADH variants and an inactive ALDH variant that promote acetaldehyde accumulation after consumption of small quantities of alcohol. This results in the very unpleasant facial flushing syndrome that is protective against heavy drinking and therefore alcohol dependence (AD). A few ADH variants have also been shown to influence alcohol consumption in non-Asian populations.

The seven human ADH genes cluster together in a 365 kb region on chromosome 4q21-24. This chromosomal region has been linked to AD in Caucasians and American Indians (Long et al., 1998; Reich et al., 1998) and maximum drinks ever consumed over 24 hours in Caucasians (Saccone et al., 2000). Most ADH genes are highly expressed in liver with the exception of *ADH7* (Edenberg, 2007). ADH enzymes have been categorized into 5 classes based on structural similarity and kinetic properties. The class I enzymes encoded by the *ADH1A*, *ADH1B* and *ADH1C* genes contribute about 70% of the total ethanol oxidizing capacity, and the class II enzyme encoded by *ADH4* contributes about 30% (Hurley, 2002; Lee et al., 2004). The class III *ADH5* enzyme is the only enzyme detectable in brain. Class IV *ADH7* is mainly expressed in the upper digestive tract where it oxidizes ethanol at high concentrations. The class V *ADH6* enzyme catalyzes a widely variety of substrates such as ethanol, retinol etc., but it is less efficient in ethanol metabolism.

Functional polymorphisms in the class I genes: *ADH1B**2 (Arg48His, rs1229984); *ADH1B**3 (Arg370Cys, rs2066702), and the *ADH1C**2 linked SNPs (Arg272Gln, rs1693482; Ile350Val, rs698), have been shown to influence the risk for alcohol dependence (AD) in several ethnic groups especially East Asian populations (Osier et al., 1999; Shen et al., 1997). The enzymes encoded by *ADH1B**2 and *ADH1B**3 have a 40-fold and 30-fold respectively higher activity than the enzyme encoded by *ADH1B**1 (Burnell et al., 1987; Edenberg 2007). The enzyme encoded by the *ADH1C**2 haplotype (Gln272Val350) has lower activity than the enzyme encoded by the more common *ADH1C**1 haplotype (Arg272Ile350) (Hoog et al, 1986). *ADH1C**1 is in strong linkage disequilibrium with *ADH1B**2 (Chen et al., 1999; Choi et al., 2005; Osier et al., 1999). The higher enzyme activity encoded by *ADH1B**2, *ADH1B**3 and *ADH1C**1 enables more rapid conversion of ethanol to acetaldehyde, which causes facial flushing and aversive effects after consuming alcohol (Osier et al., 1999). The *ADH1B**2 allele has been shown to be protective against excessive alcohol consumption and AD, particularly in East Asians who have a high frequency of this allele (Chen et al., 1999; Crabb et al., 2004; Thomasson et al., 1991;

Whitfield, 2002) but also in European and African populations (Luczak et al., 2002; Sherva et al., 2009; Whitfield, 2002). The *ADH1B**3 allele has been shown to have a significant protective effect on risk for AD in African-Americans and American Indians (Edenberg et al., 2006; Wall et al., 2003).

Noncoding variants in *ADH4* and *ADH7* have also been shown to be risk factors for AD in European American families (Edenberg et al., 2006; Luo et al., 2006a). An *ADH7* variant, *ADH7* StyI site rs1154458 (not included in our study), has been reported to have an epistatic effect with *ADH1B**2 (Arg48His) for protection against AD in Taiwanese Han and European populations (Han et al., 2005; Luo et al., 2006b; Osier et al., 2004).

Acetaldehyde is converted to acetate primarily by the mitochondrial enzyme ALDH2, encoded by the *ALDH2* gene located on chromosome 12q24.2 (Maly et al., 1999). The variant *ALDH2**2 (Glu504Lys -- this substitution corresponds to position 487 in the mature ALDH2 enzyme, rs671) encodes an inactive ALDH2 enzyme (Impraim et al., 1982; Shibuya and Yoshida, 1988a; Shibuya and Yoshida, 1988b; Xiao et al., 1995). About 50% of Japanese and Chinese individuals carry an *ALDH2**2 allele and show an alcohol flush reaction after the consumption of relatively small amounts of alcohol (Harada et al., 1982; Peng et al., 2002). Thus the *ALDH2**2 allele is protective against heavy drinking and therefore AD (Chai et al., 2005; Crabb et al., 1989; Harada et al., 1982; Higuchi et al., 2004; Peng et al., 2002; Yokoyama et al., 2005; Zintzaras et al., 2006). In addition, individuals with the combination of the higher activity *ADH1B**2 and the inactive *ALDH2**2 have a particularly low risk for AD (Chen et al., 1999). The *ALDH2**2 allele has not been found in non-East Asian populations.

ALDH1A1, located on chromosome 9q21.13, encodes the liver cytosolic isozyme ALDH1 that is particularly important for acetaldehyde elimination in those individuals who do not express the active mitochondrial ALDH2 (Bosron et al., 1993). Two polymorphisms have been identified in the *ALDH1A1* promoter region: *ALDH1A1**2, a 17bp deletion (-416/-432) and *ALDH1A1**3, a 3bp insertion (-524bp) (Spence et al., 2003). Although both *ALDH1A1**2 and *ALDH1A1**3 occur at low frequencies (0.012~0.035 for *ALDH1A1**2 in Asian, Caucasian and African populations and 0.029 for *ALDH1A1**3 in African Americans), some studies have shown associations with AD in Native Americans and African-derived populations (Ehlers et al., 2004; Moore et al., 2007; Spence et al., 2003). Furthermore, a recent study in Finnish Caucasian men has shown evidence of *ALDH1A1* SNP association with AD and alcohol consumption (Lind et al., 2008).

In our study we investigated whether variation in alcohol metabolism genes might alter the risk for AD in non-East Asian populations. In order to increase the chances of capturing unknown functional variation, we performed systematic haplotype association analyses based on 64 haplotype tagging SNPs across *ALDH1A1*, *ALDH2* and all seven genes in the ADH cluster. Our study is the first to provide a comprehensive analysis of alcohol metabolism genes in four diverse, independent populations. We tested for haplotype associations with AD in Finnish Caucasians, African Americans, Plains American Indians and Southwestern American Indians.

Materials and Methods

Four population samples were the focus of this study, each with lifetime diagnoses of AD: Finnish Caucasians, African Americans, Plains American Indians and Southwestern American (SW) Indians. Informed consent was obtained according to human research protocols approved by the Institutional Review Board (IRB) of the University of Helsinki for the Finnish sample; the IRBs of the Veterans Affairs New Jersey Healthcare System

(VANJHCS) and the New Jersey Medical School for the African American sample and the IRB of the National Institute on Alcohol Abuse and Alcoholism for both the Plains Indian and the SW Indian samples. Both tribal councils approved the study.

Finnish Caucasians

A total of 426 Finnish Caucasian men were recruited from the same source population in Finland: 232 incarcerated alcoholics (mean (SD) age = 31.5 (9.8)) and 194 controls (mean (SD) age = 31.7 (9.1)). The Structured Clinical Interview for DSM-III-R was administered by psychiatrists to both alcoholics and controls (Spitzer et al, 1992). Blind-rated DSM-III-R lifetime psychiatric diagnoses were obtained. For a more detailed description see Lappalainen et al., 1998.

African Americans

This sample of 267 alcohol dependent African American men (mean (SD) age = 45.4 (8.5)) was recruited from the Substance Abuse Treatment Program at the VANJHCS. Lifetime psychiatric diagnoses were derived using the Structured Clinical Interview for DSM-IV (SCID) (Spitzer et al, 1995). Seventy six men had a single diagnosis of AD; 191 had AD comorbid with drug dependence (cocaine and / or heroin). A total of 422 controls (175 men, 247 women) (mean (SD) age = 34.6 (9.8)) were recruited from the same geographical region as the alcoholics through outpatient diabetes or ophthalmology clinics. All controls had a semi-structured psychiatric interview and were without a lifetime history of any substance abuse or dependence or major Axis 1 psychiatric disorder. For further details see Roy et al, 2007.

Plains American Indians

Community volunteers (N = 336) were recruited from a Plains Indian tribe. There were 226 individuals (121 men, 105 women) with AD (mean (SD) age = 41.5 (11.8)) and 110 controls (35 men, 75 women) (mean (SD) age = 42.9 (17.4)). Probands were initially ascertained at random from the tribal register, and the families of alcoholic probands were extended. Although most participants derived from one large, multigenerational pedigree, the average sharing of descent between any two individuals was only 0.003. Blind-rated DSM-III-R lifetime psychiatric diagnoses were derived from the Schedule for Affective Disorders and Schizophrenia-Lifetime Version (SADS-L) (Endicott and Spitzer, 1978). An additional criterion for AD, drinking heavily for a year or more, was incorporated to establish a clear pattern of long-term alcohol use. For further details see (Enoch et al., 2006).

SW American Indians

A total of 389 participants were recruited from a SW American Indian tribe: 317 individuals with AD (174 men, 143 women, mean (SD) age = 37.4 (14.2) yrs) and 72 controls who were free of psychiatric disorders (16 men, 56 women, mean (SD) age = 35.8 (13.1) yrs). Most participants derived from three large interrelated pedigrees however the average sharing of descent between any two individuals was 0.012 which is equivalent to the relationship between second cousins once removed and third cousins. Blind-rated DSM-III-R lifetime psychiatric diagnoses were derived from the SADS-L. For a more detailed description see Robin et al., 1998.

Genotyping

Genomic regions containing sequence 5 kb upstream and 1 kb downstream of the seven ADH (*ADH1-7*) and two ALDH (*ALDH1A1* and *ALDH2*) genes were retrieved from NCBI Human Build 35.1. Haplotype tagging SNPs were identified using a previously described design pipeline (Hodgkinson et al., 2008). Seventy six SNPs spanning the nine genes were

genotyped using the Illumina GoldenGate platform (Hodgkinson et al., 2008). The average genotyping completion rate was 94.8~99.7%, genotyping replication rate was 99.4~99.9%, the average genotyping error rate was 0.05~0.3% across all four datasets.

Twelve SNPs that were monomorphic in more than two populations were excluded and therefore the analyses were performed using 64 SNPs. Marker information for the 40 SNPs spanning the seven ADH genes (364.7 kb), the 17 SNPs spanning ALDH1A1 (136.6 kb), and the 7 SNPs spanning ALDH2 (35.2 kb) is presented in Table 1. Alleles 1 and 2 are defined using the Illumina top/bottom strand convention (http://www.illumina.com/documents/products/technotes/technote_topbot.pdf) which is based upon the SNP context sequence. The Illumina alleles 1 and allele 2 were converted to the corresponding dbSNP alleles (<http://www.ncbi.nlm.nih.gov/>).

Twelve SNPs, listed in Supplementary Table 1, were not in Hardy Weinberg (HW) equilibrium in the controls. Eleven of these SNPs were out of HW equilibrium in only one population; one SNP was not in HW equilibrium in two populations. The fact that no SNP was out of HW equilibrium across all 4 populations indicated that this was not due to genotyping error therefore all 12 SNPs were included in the analyses.

Assessment of population stratification using ancestry informative markers

The samples were genotyped for 186 ancestry markers (AIMS) also using the Illumina GoldenGate platform (Hodgkinson et al., 2008). The same AIMS were also genotyped in 1051 individuals from the 51 worldwide populations represented in the HGDP-CEPH Human Genome Diversity Cell Line Panel (<http://www.cephb.fr/HGDP-CEPH-Panel>). *Structure 2.2* (<http://pritch.bsd.uchicago.edu/software.html>) was run simultaneously using the AIMS genotypes from our samples and the 51 CEPH populations to identify population substructure and compute individual ethnic factor scores. This ancestry assessment identifies seven ethnic factors that have been previously described in detail (Hodgkinson et al., 2008). The mean / median ethnic factor scores > 0.02 were as follows: Finns: European 0.85 / 0.92, Asian 0.08 / 0.02; Plains Indians: Americas 0.72 / 0.76, Far East 0.13 / 0.08, European 0.06 / 0.02, Asian 0.04 / 0.02; SW Indians: Americas 0.81 / 0.86, Far East 0.08 / 0.04, European 0.04 / 0.01, Asian 0.03 / 0.01; African Americans: African 0.77 / 0.81, European 0.09 / 0.04, Mid Eastern 0.06 / 0.04, Asian 0.06 / 0.03. Ethnic factor scores were included in the logistic regression analyses if they had a significant effect.

Three SNPs (rs1154456, rs1154469 and rs1154470) in ADH7 block2 and two ALDH2 SNPs (rs10849970 and rs7296651) were associated with the far east ethnic factor score.

Statistics

Linkage disequilibrium (LD) D' values were evaluated for all marker pairs and haplotype blocks were identified using Haploview. A minimum D' value of 0.80 (Gabriel confidence intervals for 'strong LD' of 0.70 to 0.98) was used to define block boundaries. Reference haplotype blocks were derived from the HapMap African, Caucasian and the two Asian reference populations (Figures 1, 2, 3). Haplotype block structure was compared between the reference HapMap populations and our four study populations. Haplotypes were derived using the program Phase 2.0 (Li and Stephens, 2003). Comparisons of haplotype frequencies between cases and controls were performed using Fishers test (<http://www.quantitativeskills.com/sisa/statistics/fisher.htm>). In secondary analyses, allele frequencies of individual SNPs were compared between cases and controls using Fishers test.

To correct for multiple testing, the false discovery rate (FDR) based procedure was implemented using R package qvalue version 1.1 to obtain the q-value corresponding to

each p value in four populations (Storey and Tibshirani, 2003). R package takes a list of p-values resulting from the simultaneous testing of many hypotheses and estimates their q-values. The q-value of a test measures the proportion of false positives incurred when that particular test is called significant. We assumed that all haplotypes had an equal chance of association with AD. Therefore we conducted 35, 34, 39 and 45 tests in FDR analyses in Plains Indians, Southwestern American Indians, Finnish Caucasians and African Americans respectively.

RESULTS

Haplotype block structure across the ADH gene cluster

The extent of LD between the 40 ADH SNPs was determined in the four study populations and in the three HapMap populations (Figure 1). For the 365 kb ADH gene cluster, five haplotype blocks were defined: block 1 (13 SNPs): ADH5-ADH4; block 2 (13 SNPs): ADH6-ADH1A-ADH1B; block 3 (5 SNPs): ADH1C; block 4 (2 SNPs): intergenic region between ADH1C and ADH7; block 5(6 SNPs): ADH7. LD across the ADH gene cluster is similar across the four populations (Figure 1, right panel). The haplotype block structure derived from our study is consistent with the haplotype block structure of the equivalent HapMap populations, based on visualization (Figure 1, left panel). As expected, the haplotype diversity was greatest in populations with African ancestry.

ADH Gene Cluster haplotypic analyses

Block 1: ADH5-ADH4—From Table 2 it can be seen that there was one predominant haplotype that ranged in frequency from 0.89 in Plains Indian to 0.50 in African American controls. There was no haplotype association with AD. One ADH4 SNP (SNP12, rs3762894) showed an association with AD in Plains Indians (controls=0.02, AD=0.004, $p=0.04$, $r^2 = 0.005$) and showed marginal association in African Americans in the opposite direction (controls=0.18, AD=0.22, $p=0.08$) (Table 1).

Block 2: ADH6-ADH1A-ADH1B—There was considerable haplotypic diversity across the 4 populations. In the SW Indians, 8 haplotypes accounted for 99% of the haplotype diversity whereas in the African Americans, 12 haplotypes accounted for 81% of the diversity. There were 2 predominant haplotypes that were common in the SW Indians, the Plains Indians and the Caucasians but there were no associations with AD. None of the haplotypes listed in Table 2 included the functional *ADH1B**2 rs1229984 Arg48His variant since it was either non-existent or occurred at a very low frequency (0.01) in these populations.

One minor haplotype indicated in Table 2, was significantly more common in SW Indian controls than in AD subjects ($\chi^2 = 8.7$, 1 df, $p = 0.007$) and showed a trend effect in the same direction in the African Americans ($\chi^2 = 2.6$, 1 df, $p = 0.11$). This haplotype was not present in the other two groups.

Block 3: ADH1C—There were two or three predominant haplotypes but no association with AD (Table 3). The *ADH1C**1 and *ADH1C**2 linked SNPs (Arg272Gln, rs1693482; Ile350Val, rs698), have been associated with AD in previous studies in several populations. Both variants were common across the four study populations with frequencies in controls ranging from 0.15~0.50. Nevertheless we did not detect any association between the functional *ADH1C**1 and *ADH1C**2 haplotypes and AD across the four populations (Table 3).

Haplotype block 4 and block 5: ADH7—The two-SNP haplotype block 4 encompasses the intergenic region between the *ADH1C* and *ADH7* haplotype blocks (Figure 1) and in all four populations includes two yin yang haplotypes with differing frequencies (Table 4). ‘Yin yang’ is the term used to describe haplotypes that have opposite allelic configurations, in this case 11 and 22. In SW Indians only, the yin yang haplotypes were associated with AD ($\chi^2=4.6$, 1df, $p=0.03$).

In block 5 that includes ADH7 no haplotypic association with AD was observed (Table 4).

ALDH1A1 haplotypic analyses

Fourteen haplotype tagging SNPs were selected from the 17 *ALDH1A1* SNPs listed in Table 1. The extent of LD between the 14 SNPs was determined in the four study populations (Figure 2, right panel) and in the three HapMap populations (Figure 2, left panel). Because the *ALDH1A1* haplotype blocks are not exactly the same across the four populations and haplotype diversity differs in the four populations, we split the 14 SNP configuration into four small blocks to avoid artificial haplotypes and to increase the power of haplotype analysis.

In all four *ALDH1A1* haplotype blocks, two major (yin yang) haplotypes were observed across the four populations, although as usual there was more haplotypic diversity in the African Americans. The yin-yang haplotypes accounted for 0.92~0.98 of the haplotype diversity in Plains Indians, 0.92~1 in SW Indians, 0.66~0.98 in Finnish Caucasians but only 0.32~0.73 in African Americans (Table 5).

As indicated in Table 5, one of the yin yang haplotypes in both block 1 and block 2 was associated with AD in the Finns ($\chi^2 = 4.03$, 1 df, $p = 0.04$; $\chi^2 = 5.86$, 1 df, $p = 0.02$, respectively). Within the Finnish block 1, SNP 41 rs3764435 appears to be driving the haplotype association with AD: $\chi^2 = 5.2$, 1 df, $p = 0.03$, $r^2 = 0.0124$. Within block 2, SNP 45 rs2303317 appears to be driving the haplotype association with AD: $\chi^2 = 5.5$, 1 df, $p = 0.03$, $r^2 = 0.005$. (Table 1)

In block 1, two minor yin yang haplotypes were associated with AD: haplotype 211 showed association in African Americans (AD=0.05, Controls=0.02, $\chi^2 = 7.62$, 1 df, $p = 0.01$) and 122 showed association in Finnish Caucasians (AD=0.01, Controls=0.03, $\chi^2 = 6.2$, 1 df, $p = 0.02$). However, these haplotypes may have lower attributable risk and thus these associations are less important because of the relatively low frequencies of these haplotypes.

In block 3 one of the major yin yang haplotypes showed an association with AD in SW Indians ($\chi^2 = 5.71$, 1 df, $p = 0.02$) (Table 5). Of the 7 SNPs within this haplotype block, 5 SNPs (rs1424482, rs8187876, rs2249978, rs1418187 and rs4745209) are in allelic identity (MAF in: AD = 0.42, controls = 0.49) and are associated with AD: $\chi^2 = 4.0 - 5.7$, 1 df, $p = 0.02 - 0.05$, $r^2 = 0.005 - 0.008$. These 5 SNPs are driving the haplotype association with AD in SW Indians.

ALDH2 haplotypic analyses

All *ALDH2* SNPs were in strong LD in all four populations with the exception of rs10849970 in Finnish Caucasians (Supplementary Figure 1). In the Plains Indians only, one haplotype was less common in alcoholics (0.04) than in controls (0.08) ($\chi^2=4.5$, df =1, $p=0.03$) (Table 6). This association appears to be driven by SNPs rs2238152, rs4648328 and rs4646778, all of which have a frequency of 0.04 in AD and 0.08 in controls ($\chi^2=4.5$, df =1, $p=0.05$). The relatively high q values by FDR analysis showed that this association is probably due to chance (data not show).

In African Americans, one haplotype (that includes rs7311852 which is monomorphic in the other samples) was less common in AD individuals (0.05) than in controls (0.08) ($\chi^2=4.6$, $df=1$, $p=0.03$) (Table 6). This association was driven by rs7311852 (controls=0.08, AD=0.05, $\chi^2=4.6$, 1 df , $p=0.03$). However, when the European ethnic factor score was included as a covariate in the logistic regression model, the p value changed to 0.05 ($\chi^2=3.9$, 1 df) indicating that the positive association was influenced by population stratification in this admixed population.

Corrections for Multiple Testing

The minimum q-value corresponding to the significant p-value was 0.58 (35 tests) in Plains Indians, 0.31 (39 tests) in SW Indians, 0.34 (34 tests) in Finnish Caucasians and 0.27 (45 tests) in African Americans. These high q-values indicate that none of the nominally significant haplotype associations with AD were statistically significant.

DISCUSSION

The present study provides a comprehensive test of association between AD and haplotypes in the ADH gene cluster and two ALDH genes in four populations: Plains American Indians, SW American Indians, Finnish Caucasians and African Americans. Our study did not find evidence of any significant associations once the results were corrected for multiple comparisons. However there are several results that are of interest, particularly for *ALDH1A1* and *ADH4*, since they replicate other findings in the literature.

Our study has shown an *ALDH1A1* haplotype association with AD in SW American Indians. In *ALDH1A1* haplotype block 3, one of the major yin yang haplotypes was more abundant in alcoholics whereas the other yin yang haplotype was more abundant in controls. Of the seven SNPs in this haplotype block, five were in allelic identity and associated with AD suggesting that this block might have been subjected to selective pressure. In the Plains Indians the same five SNPs were also in allelic identity unlike in the Finns and African Americans, suggesting that this selective pressure might be unique to American Indians. The *ALDH1A1* block 3 yin-yang haplotype patterns in our four study populations are consistent with the phased haplotype patterns derived from the Asian, Caucasian and African HapMap populations (Figure 3).

The *ALDH1A1* haplotype block 3 includes the promoter, the minimal promoter region (-91 to +53) that contains regulatory elements and the CCAAT box region intron1 (Yanagawa et al., 1995). It is of interest that two low frequency promoter polymorphisms (0.012 – 0.035), a 17bp deletion (-416/-432) and a 3bp insertion (-524) (*ALDH1A1**2 and *ALDH1A1**3 respectively) have previously been associated with AD in Southwest California Indians (Ehlers et al., 2004) and Caribbean populations (Moore et al., 2007). In secondary analyses we genotyped 91 yin-yin and 93 yang-yang SW American Indians for the *ALDH1A1**2 and *ALDH1A1**3 variants but neither variant was detected.

We also found a yin yang haplotype association in Finnish Caucasians in haplotype blocks 1 and 2 that extends from intron 5 to the 3' end of the gene. A recent study also in Finnish Caucasians found significant associations between AD and SNPs in intron 8 and the 3' UTR (Lind et al., 2008), the same locations as in our study. A study in European Americans found an association between rs8187974 (intron10), located in an *ALDH1A1* splice site, and AD as well as max drinks / 24 hours (Sherva et al., 2009). The closest SNP to rs8187974 in our study was rs63319 (at a distance of 2082bp), but rs63319 did not show any association with AD. Moreover, another study in which 1105 Irish Caucasians were genotyped for the same SNPs as in our study showed no association between AD and *ALDH1A1* variation (Kuo et al., 2008).

There are two distinct, major isozymes in mammalian liver: ALDH2 is mitochondrial in origin and has a high affinity for acetaldehyde whereas ALDH1 is cytosolic in origin and has a low affinity for acetaldehyde. Therefore it is not clear how *ALDH1A1* may be implicated in AD in individuals with functioning ALDH2 enzymes. Nevertheless ALDH1 deficiency has been associated with sensitivity to alcohol (Agarwal, 1981; Harada et al., 1981) and with facial flushing (Harada et al., 1981; Yoshida et al., 1989) and fatty liver (Thomas et al., 1982).

The class I enzymes encoded by *ADH1A*, *ADH1B* and *ADH1C* contribute about 70% of the total ethanol oxidizing capacity. One minor haplotype in the *ADH6-ADH1A-ADH1B* haplotype block was significantly more common in SW Indian controls than in AD subjects ($p = 0.007$) and showed a trend effect in the same direction in the African Americans ($p = 0.11$). This haplotype was not present in the other two groups. However, these associated haplotypes did not include the *ADH1B*2* functional SNP rs1229984 (His48Arg). We found that rs1229984 was either not present or occurred at a low frequency (0.01) in our four samples and therefore we cannot comment on the potential influence of this functional polymorphism on AD with our current samples sizes. In contrast, Sherva et al were able to show an association between rs1229984 and AD in a sample of 1588 European Americans (Sherva et al., 2009).

Two *ADH1C* SNPs, rs698 and rs1693482, distinguish *ADH1C*1* from *ADH1C*2*. The two SNPs are in very high LD, as shown in previous studies and in our study. Both *ADH1C*1* and *ADH1C*2* are abundant in the four samples in this study. In individuals homozygous for the *ADH1B*1* allele, as is the case for nearly all non-Asian individuals, the ethanol-oxidizing differences between the enzymes encoded by *ADH1C*1* and *ADH1C*2* are low (Edenberg et al., 2006; Lee et al., 2004; Matsuo et al., 2007; Wall et al., 2005). This might explain the lack of association with AD in our datasets.

The class II enzyme encoded by *ADH4* contributes about 30% the total ethanol oxidizing capacity. *ADH4* has previously been associated with AD in European Americans (Luo et al., 2006a). Our positive result for *ADH4* is supported by the findings in the study from the Collaborative Study on the Genetics of Alcoholism (COGA) in which 110 SNPs across the *ADH* gene cluster were genotyped in 1860 European American individuals from 218 families (Edenberg et al, 2006). The pedigree disequilibrium test (PDT) showed that 12 SNPs spanning *ADH4* were significantly associated with AD and one of the three haplotype tagging SNPs was rs3762894. In our study this SNP was associated with AD in the Plains Indians and showed a trend association in the African Americans. Although the COGA sample size was large ($N = 1860$), the best SNP association p value was 0.004. In Edenberg's study, the MAF was 0.17 for all European Americans. Our study detected the same frequencies in the Finnish Caucasians (0.16 in AD, 0.12 in controls) and this is similar to the frequency in HapMap European Americans (0.13). An effect size cannot be derived from the PDT and therefore we were unable to directly compare our results in the Finnish Caucasians with the European Americans in Edenberg et al, 2006.

Kuo et al (2008) tested the association between AD and the same set of SNPs as in our study, genotyped by the identical method (Hodgkinson et al., 2008), in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) Sample ($n = 1105$) (Kuo et al., 2008). Likewise with the Caucasian sample in our study, there were few haplotype and SNP associations with AD. Out of all the SNPs tested in the Kuo et al study, two survived corrections for multiple testing: the *ADH5* SNP rs1154414 and the *ADH1B* SNP rs1353621. These findings were not replicated in our study. We performed power analyses based on the effect sizes for these two SNPs derived from the IASPSAD study. For *ADH5* rs1154414 we calculated that we had 95% power to detect an effect in the Finnish Caucasian sample. The

fact that we did not detect an association between rs1154414 and AD was due to the fact that the MAF differed between AD subjects in the IASPSAD study (0.09) and the Finnish Caucasians in our study (0.15) although the MAF in controls was the same in both studies (0.12). For ADH1B rs1353621, we calculated that we had 93% power to detect the same effect size in the Finnish Caucasians as had been detected in the IASPSAD study. However the MAF in the Finnish cases (0.30) and controls (0.33) were similar and differed from the MAF in cases (0.43) and controls (0.37) in the IASPSAD study.

ADH7 shows no LD with other ADH genes. This is consistent with the findings of earlier studies (Edenberg et al., 2006; Kuo et al., 2008) and HapMap data. The intergenic haplotype block includes the 3'UTR and the nearby 3' region of *ADH7*, indicating that the haplotype association with AD that we found in SW Indians might implicate a regulatory locus. However, relatively high *q* values by FDR analysis showed that this association is likely to be due to chance. Osier et al reported a possible epistatic role of *ADH7* and *ADH1B**2 in protection against alcohol abuse and/or AD by the haplotype rs1154458-rs1229984 (Osier et al., 2004). However, in our study we found no evidence of LD between SNPs in *AD7* and *ADH1B*.

It should be noted that there was considerable allele and haplotype frequency variation between the four populations for all genes, including variations between the two American Indian tribes. It could have been expected that of the four populations tested, the Plains Indians and SW Indians would be most likely to have alcohol metabolizing gene variants similar to Asians since their ancestors are thought to have migrated across the Bering Straits from Asia (Goebel et al., 2008). However these two American Indian tribes do not have the Asian ADH and ALDH2 gene variants which is consistent with proposals of a population bottleneck at the time of entry to the Americas that would have eliminated much genetic variation in American Indian populations (Kitchen et al., 2008). The considerable allelic variation could result in population stratification that can produce false-positive or false-negative results in case-control studies. However we included ethnic factor scores derived from 186 AIMs as covariates in our analyses when significant and therefore population stratification is unlikely to have influenced our results.

Yin yang haplotypes are defined as two high-frequency haplotypes with exactly opposite allelic configurations, for example the haplotypes 1222221 and 2111112 in *ALDH1A1* block 3. It has been shown that the proportion of the genome spanned by yin yang haplotypes is approximately 75%-85% and their conservation across all populations suggests that they represent ancient chromosomal regions that are likely to predate the African diaspora (Zhang et al., 2003). In accordance with HapMap data we found that yin yang haplotypes were present in all four populations although, as expected, at lower frequencies in the sample with African ancestry. At this point in time the significance of yin yang haplotypes is not clear. Using coalescence simulation, Zhang et al, 2003 have shown that yin yang haplotypes can be explained by strictly neutral evolution in a well-mixed population however another study indicates that yin yang haplotypes can be explained by the maintenance of ancient lineages by selection (Curtis and Vine, 2010).

There are a few potential limitations to our study. Different diagnostic criteria (DSM-III-R and DSM-IV) and different psychiatric instruments were used in this study. However, the interviews of the Finnish Caucasians and African Americans were conducted by two experienced clinical psychiatrists and the interviews of the Plains and Southwestern Indians were conducted by two clinical social workers who were fully versed in tribal customs and cultures. Therefore we are confident in the validity of the AD diagnosis. The controls were standardized across all four samples in that they had no lifetime Axis 1 disorders however nicotine dependence was not an exclusion criteria since we only had smoking history for the

Plains Indians. This resulted in effectively increasing the heterogeneity of the AD samples since the samples included AD alone and AD comorbid with other disorders. Two samples had significant comorbidity: cocaine and heroin dependence in the African Americans and ASPD in the Finnish Caucasians. Nevertheless the comorbid disorders are unlikely to have any associations with alcohol metabolizing genes. Because of the relatively small samples sizes across the four populations we were unable to determine potentially important rarer haplotype associations. Moreover, rare and uncommon haplotypes estimated using any method are subject to higher error because of increased sampling variance and genotyping errors. However we were able to analyze common haplotypic variation across these genes; for example the *ALDH1A1* block 3 yin yang haplotypes are fairly abundant.

In conclusion, the systematic evaluation of common variation in alcohol metabolizing genes in four non East Asian populations has shown only modest associations with AD, largely for *ALDH1A1* and *ADH4*. A concentration of signals for AD with *ALDH1A1* yin yang haplotypes in several populations warrants further research. Our study is the first to provide a comprehensive analysis of alcohol metabolism genes in four independent populations. The fact that no strong genetic associations were found may reflect the heterogeneity of the disease and the large role that environmental factors play in the development of alcohol dependence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institute on Alcohol Abuse and Alcoholism, NIH and in part by grant RO1 DA 10336-02 to AR from the National Institute of Drug Abuse, NIH and by NIH grant R03 AA12906 to C.J.M. We gratefully acknowledge the work performed by Robert Robin and Bernard Albaugh in recruiting and interviewing the SW Indians and Plains Indians, respectively. Longina Akhtar and Elisa Moore provided technical support. Danielle Muchnik, Angelica González-Oliver, David Mayorga, Ben Burkley, and Jena Chojnowski assisted in genotyping samples.

REFERENCES

- Agarwal DPM-T, D.; Harada, S.; Goedde, HW.; Du, R. Sixth Int. Cong. Hum. Genet. Vol. 102. Jerusalem: 1981. Mechanism of biological sensitivity to alcohol: inherited deficiency of aldehyde dehydrogenase isoenzyme I in Mongoloids. only
- Bosron WF, Ehrig T, Li TK. Genetic factors in alcohol metabolism and alcoholism. *Semin Liver Dis.* 1993; 13:126–135. [PubMed: 8337601]
- Burnell JC, Carr LG, Dwulet FE, Edenberg HJ, Li TK, Bosron WF. The human beta 3 alcohol dehydrogenase subunit differs from beta 1 by a Cys for Arg-369 substitution which decreases NAD(H) binding. *Biochem Biophys Res Commun.* 1987; 146:1127–1133. [PubMed: 3619918]
- Chai YG, Oh DY, Chung EK, Kim GS, Kim L, Lee YS, Choi IG. Alcohol and aldehyde dehydrogenase polymorphisms in men with type I and Type II alcoholism. *Am J Psychiatry.* 2005; 162:1003–1005. [PubMed: 15863807]
- Chen CC, Lu RB, Chen YC, Wang MF, Chang YC, Li TK, Yin SJ. Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *Am J Hum Genet.* 1999; 65:795–807. [PubMed: 10441588]
- Choi IG, Son HG, Yang BH, Kim SH, Lee JS, Chai YG, Son BK, Kee BS, Park BL, Kim LH, Choi YH, Shin HD. Scanning of genetic effects of alcohol metabolism gene (*ADH1B* and *ADH1C*) polymorphisms on the risk of alcoholism. *Hum Mutat.* 2005; 26:224–234. [PubMed: 16086315]
- Crabb DW, Edenberg HJ, Bosron WF, Li TK. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive *ALDH2(2)* allele is dominant. *J Clin Invest.* 1989; 83:314–316. [PubMed: 2562960]

- Crabb DW, Matsumoto M, Chang D, You M. Overview of the role of alcohol dehydrogenase and aldehyde dehydrogenase and their variants in the genesis of alcohol-related pathology. *Proc Nutr Soc.* 2004; 63:49–63. [PubMed: 15099407]
- Curtis D, Vine AE. Yin yang haplotypes revisited - long, disparate haplotypes observed in European populations in regions of increased homozygosity. *Hum Hered.* 2010; 69:184–192. [PubMed: 20203523]
- 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, 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. *Hum Mol Genet.* 2006; 15:1539–1549. [PubMed: 16571603]
- 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.* 2004; 28:1481–1486. [PubMed: 15597079]
- Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry.* 1978; 35:837–844. [PubMed: 678037]
- Enoch MA, Waheed JF, Harris CR, Albaugh B, Goldman D. Sex differences in the influence of COMT Val158Met on alcoholism and smoking in plains American Indians. *Alcohol Clin Exp Res.* 2006b; 30:399–406. [PubMed: 16499480]
- Goebel T, Waters MR, O'Rourke DH. The late Pleistocene dispersal of modern humans in the Americas. *Science.* 2008; 319:1497–1502. [PubMed: 18339930]
- Han Y, Oota H, Osier MV, Pakstis AJ, Speed WC, Odunsi A, Okonofua F, Kajuna SL, Karoma NJ, Kungulilo S, Grigorenko E, Zhukova OV, Bonne-Tamir B, Lu RB, Parnas J, Schulz LO, Kidd JR, Kidd KK. Considerable haplotype diversity within the 23kb encompassing the ADH7 gene. *Alcohol Clin Exp Res.* 2005; 29:2091–2100. [PubMed: 16385178]
- Harada S, Agarwal DP, Goedde HW. Aldehyde dehydrogenase deficiency as cause of facial flushing reaction to alcohol in Japanese. *Lancet.* 1981; 2:982. [PubMed: 6117742]
- Harada S, Agarwal DP, Goedde HW, Tagaki S, Ishikawa B. Possible protective role against alcoholism for aldehyde dehydrogenase isozyme deficiency in Japan. *Lancet.* 1982; 2:827. [PubMed: 6126701]
- Higuchi S, Matsushita S, Masaki T, Yokoyama A, Kimura M, Suzuki G, Mochizuki H. Influence of genetic variations of ethanol-metabolizing enzymes on phenotypes of alcohol-related disorders. *Ann N Y Acad Sci.* 2004; 1025:472–480. [PubMed: 15542751]
- Hodgkinson CA, Yuan Q, Xu K, Shen PH, Heinz E, Lobos EA, Binder EB, Cubells J, Ehlers CL, Gelernter J, Mann J, Riley B, Roy A, Tabakoff B, Todd RD, Zhou Z, Goldman D. Addictions biology: haplotype-based analysis for 130 candidate genes on a single array. *Alcohol Alcohol.* 2008; 43:505–515. [PubMed: 18477577]
- Hurley, TD.; E, HJ.; Li, T-K. *Pharmacogenomics: The Search for Individualized Therapies.* Wiley-VCH; Weinheim, Germany: 2002. The pharmacogenomics of alcoholism; p. 417-441.
- Impraim C, Wang G, Yoshida A. Structural mutation in a major human aldehyde dehydrogenase gene results in loss of enzyme activity. *Am J Hum Genet.* 1982; 34:837–841. [PubMed: 7180842]
- Kitchen A, Miyamoto MM, Mulligan CJ. A three-stage colonization model for the peopling of the Americas. *PLoS One.* 2008; 3:e1596. [PubMed: 18270583]
- 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. *Alcohol Clin Exp Res.* 2008; 32:785–795. [PubMed: 18331377]
- Lee SL, Hoog JO, Yin SJ. Functionality of allelic variations in human alcohol dehydrogenase gene family: assessment of a functional window for protection against alcoholism. *Pharmacogenetics.* 2004; 14:725–732. [PubMed: 15564879]
- Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics.* 2003; 165:2213–2233. [PubMed: 14704198]

- Lind PA, Eriksson CJ, Wilhelmsen KC. The role of aldehyde dehydrogenase-1 (ALDH1A1) polymorphisms in harmful alcohol consumption in a Finnish population. *Hum Genomics*. 2008; 3:24–35. [PubMed: 19129088]
- Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, Bennett PH, Goldman D. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosomal-wide scan in an American Indian population. *Am J Med Genet*. 1998; 81:216–221. [PubMed: 9603607]
- Luczak SE, Shea SH, Carr LG, Li TK, Wall TL. Binge drinking in Jewish and non-Jewish white college students. *Alcohol Clin Exp Res*. 2002; 26:1773–1778. [PubMed: 12500100]
- 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*. 2006a; 31:1085–1095. [PubMed: 16237392]
- Luo X, Kranzler HR, Zuo L, Wang S, Schork NJ, Gelernter J. Diplotype trend regression analysis of the ADH gene cluster and the ALDH2 gene: multiple significant associations with alcohol dependence. *Am J Hum Genet*. 2006b; 78:973–987. [PubMed: 16685648]
- Maly IP, Crotet V, Sasse D. Spatial distribution of human liver aldehyde dehydrogenase isoenzymes. *Histochem Cell Biol*. 1999; 111:461–466. [PubMed: 10429968]
- Matsuo K, Hiraki A, Hirose K, Ito H, Suzuki T, Wakai K, Tajima K. Impact of the alcohol-dehydrogenase (ADH) 1C and ADH1B polymorphisms on drinking behavior in nonalcoholic Japanese. *Hum Mutat*. 2007; 28:506–510. [PubMed: 17285601]
- Moore S, Montane-Jaime K, Shafe S, Joseph R, Crooks H, Carr LG, Ehlers CL. Association of ALDH1 promoter polymorphisms with alcohol-related phenotypes in Trinidad and Tobago. *J Stud Alcohol Drugs*. 2007; 68:192–196. [PubMed: 17286337]
- Osier M, Pakstis AJ, Kidd JR, Lee JF, Yin SJ, Ko HC, Edenberg HJ, Lu RB, Kidd KK. Linkage disequilibrium at the ADH2 and ADH3 loci and risk of alcoholism. *Am J Hum Genet*. 1999; 64:1147–1157. [PubMed: 10090900]
- Osier MV, Lu RB, Pakstis AJ, Kidd JR, Huang SY, Kidd KK. Possible epistatic role of ADH7 in the protection against alcoholism. *Am J Med Genet B Neuropsychiatr Genet*. 2004; 126B:19–22. [PubMed: 15048643]
- Peng GS, Yin JH, Wang MF, Lee JT, Hsu YD, Yin SJ. Alcohol sensitivity in Taiwanese men with different alcohol and aldehyde dehydrogenase genotypes. *J Formos Med Assoc*. 2002; 101:769–774. [PubMed: 12517056]
- 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 Jr, 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; 81:207–215. [PubMed: 9603606]
- Robin RW, Long JC, Rasmussen JK, Albaugh B, Goldman D. Relationship of binge drinking to alcohol dependence, other psychiatric disorders, and behavioral problems in an American Indian tribe. *Alcohol Clin Exp Res*. 1998; 22:518–523. [PubMed: 9581662]
- Saccone NL, Kwon JM, Corbett J, Goate A, Rochberg N, Edenberg HJ, Foroud T, Li TK, Begleiter H, Reich T, Rice JP. A genome screen of maximum number of drinks as an alcoholism phenotype. *Am J Med Genet*. 2000; 96:632–637. [PubMed: 11054770]
- Shen YC, Fan JH, Edenberg HJ, Li TK, Cui YH, Wang YF, Tian CH, Zhou CF, Zhou RL, Wang J, Zhao ZL, Xia GY. Polymorphism of ADH and ALDH genes among four ethnic groups in China and effects upon the risk for alcoholism. *Alcohol Clin Exp Res*. 1997; 21:1272–1277. [PubMed: 9347089]
- Sherva R, Rice JP, Neuman RJ, Rochberg N, Saccone NL, Bierut LJ. Associations and interactions between SNPs in the alcohol metabolizing genes and alcoholism phenotypes in European Americans. *Alcohol Clin Exp Res*. 2009; 33:848–857. [PubMed: 19298322]
- Shibuya A, Yoshida A. Frequency of the atypical aldehyde dehydrogenase-2 gene (ALDH2(2)) in Japanese and Caucasians. *Am J Hum Genet*. 1988a; 43:741–743. [PubMed: 3189337]

- Shibuya A, Yoshida A. Genotypes of alcohol-metabolizing enzymes in Japanese with alcohol liver diseases: a strong association of the usual Caucasian-type aldehyde dehydrogenase gene (ALDH1(2)) with the disease. *Am J Hum Genet.* 1988b; 43:744–748. [PubMed: 3189338]
- Spence JP, Liang T, Eriksson CJ, Taylor RE, Wall TL, Ehlers CL, Carr LG. Evaluation of aldehyde dehydrogenase 1 promoter polymorphisms identified in human populations. *Alcohol Clin Exp Res.* 2003; 27:1389–1394. [PubMed: 14506398]
- Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry.* 1992; 49:624–629. [PubMed: 1637252]
- Spitzer, RL.; Williams, JBW.; Gibbon, M. Structured Clinical Interview for DSM-IV (SCID). New York State Psychiatric Institute, Biometrics Research; New York: 1995.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003; 100:9440–9445. [PubMed: 12883005]
- Thomas M, Halsall S, Peters TJ. Role of hepatic acetaldehyde dehydrogenase in alcoholism: demonstration of persistent reduction of cytosolic activity in abstaining patients. *Lancet.* 1982; 2:1057–1058. [PubMed: 6127541]
- 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. *Am J Hum Genet.* 1991; 48:677–681. [PubMed: 2014795]
- Wall TL, Carr LG, Ehlers CL. Protective association of genetic variation in alcohol dehydrogenase with alcohol dependence in Native American Mission Indians. *Am J Psychiatry.* 2003; 160:41–46. [PubMed: 12505800]
- Wall TL, Shea SH, Luczak SE, Cook TA, Carr LG. Genetic associations of alcohol dehydrogenase with alcohol use disorders and endophenotypes in white college students. *J Abnorm Psychol.* 2005; 114:456–465. [PubMed: 16117582]
- Whitfield JB. Alcohol dehydrogenase and alcohol dependence: variation in genotype-associated risk between populations. *Am J Hum Genet.* 2002; 71:1247–1250. author reply 1250–1251. [PubMed: 12452180]
- Xiao Q, Weiner H, Johnston T, Crabb DW. The aldehyde dehydrogenase ALDH2*2 allele exhibits dominance over ALDH2*1 in transduced HeLa cells. *J Clin Invest.* 1995; 96:2180–2186. [PubMed: 7593603]
- Yanagawa Y, Chen JC, Hsu LC, Yoshida A. The transcriptional regulation of human aldehyde dehydrogenase I gene. The structural and functional analysis of the promoter. *J Biol Chem.* 1995; 270:17521–17527. [PubMed: 7615557]
- Yokoyama M, Yokoyama A, Yokoyama T, Funazu K, Hamana G, Kondo S, Yamashita T, Nakamura H. Hangover susceptibility in relation to aldehyde dehydrogenase-2 genotype, alcohol flushing, and mean corpuscular volume in Japanese workers. *Alcohol Clin Exp Res.* 2005; 29:1165–1171. [PubMed: 16046871]
- Yoshida A, Dave V, Ward RJ, Peters TJ. Cytosolic aldehyde dehydrogenase (ALDH1) variants found in alcohol flushers. *Ann Hum Genet.* 1989; 53:1–7. [PubMed: 2729894]
- Zhang J, Rowe WL, Clark AG, Buetow KH. Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. *Am J Hum Genet.* 2003; 73:1073–1081. [PubMed: 14560401]
- Zintzaras E, Stefanidis I, Santos M, Vidal F. Do alcohol-metabolizing enzyme gene polymorphisms increase the risk of alcoholism and alcoholic liver disease? *Hepatology.* 2006; 43:352–361. [PubMed: 16440362]

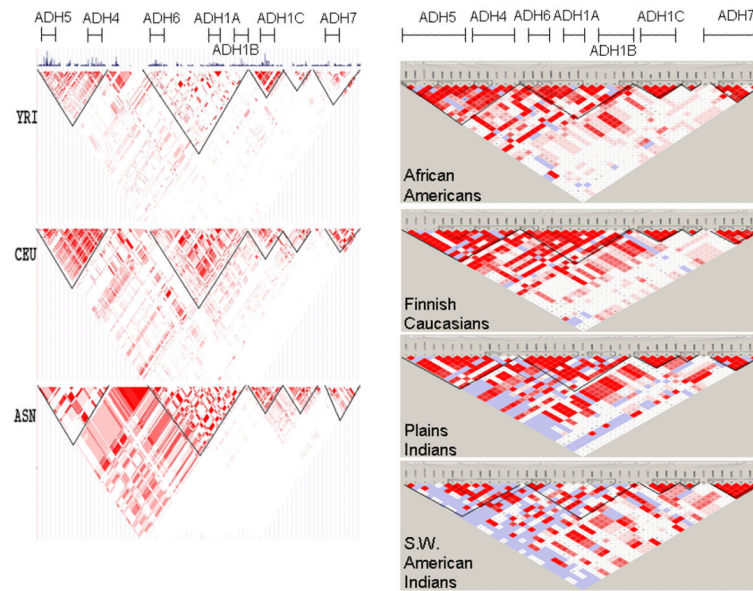


FIGURE 1. Haplotype Block Structure across Four Populations and Three HapMap Populations for the ADH Gene Cluster on Chromosome 4

Haplotype blocks are outlined. Linkage disequilibrium between SNPs within haplotype blocks was set at $D' \geq 0.80$.

Left panel: HapMap populations: YRI = Yoruban Africans; CEU = U.S. Caucasians; ASN = Japanese / Chinese

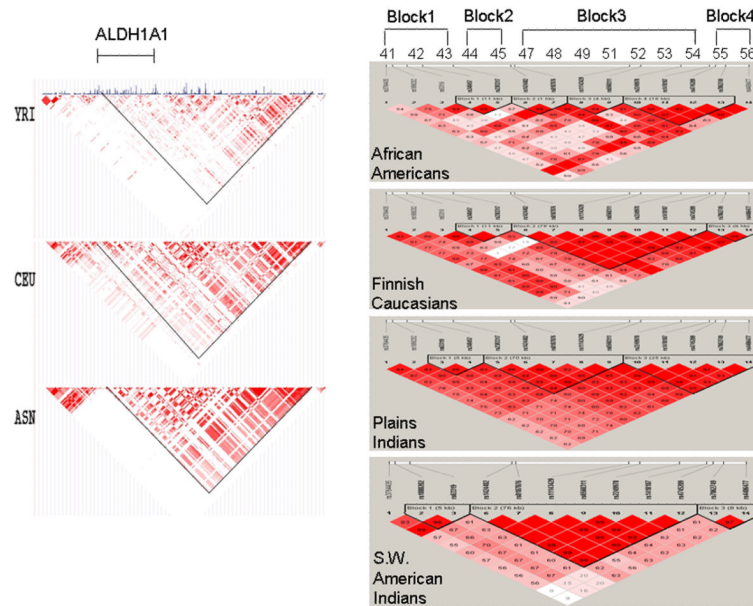


FIGURE 2. Haplotype Block Structure across Four Populations and Three HapMap Populations for the *ALDH1A1* Gene on Chromosome 9

Haplotype blocks are outlined. Linkage disequilibrium between SNPs within haplotype blocks was set at $D' \geq 0.80$.

Right panel: SNPs 41 – 56 (as listed in Table 1) are labeled.

Left panel: HapMap populations: YRI = Yoruban Africans; CEU = U.S. Caucasians; ASN = Japanese / Chinese

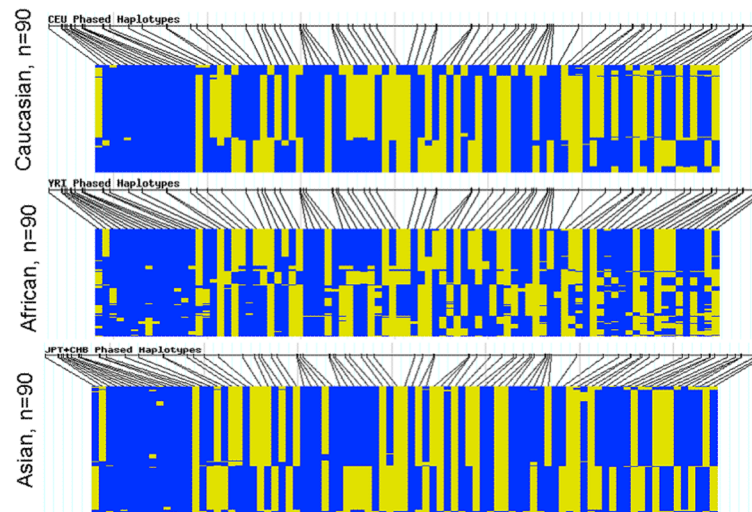


FIGURE 3. Yin Yang Haplotypes in HapMap *ALDH1A1* Haplotype Block 3
Black lines indicate the SNPs genotyped in the HapMap samples. The two colors indicate the two alleles for each SNP. The Caucasian and Asian populations show two major yin yang haplotypes. The Africans show more haplotype diversity.

Table 1

Characteristics of the 64 ADH and ALDH SNPs

ID	SNP	Allele1/Allele2	Gene	Minor Allele Frequency											
				Plains Indians		Southwest Indians		Finnish Caucasians		African Americans					
				AD	Controls	AD	Controls	AD	Controls	AD	Controls	AD	Controls		
1	rs7684986	C/T	ADH5	0.01	0.02	0	0	0.04	0.05	0.07	0.06				
2	rs17595424	G/T	ADH5	0.01	0.02	0	0	0.04	0.05	0.09	0.08				
3	rs1154414	C/T	ADH5	0	0.02	0.01	0.01	0.15	0.12	0.03	0.02				
4	rs7683704	C/T	ADH5	0.01	0.02	0.003	0	0.04	0.05	0.25	0.26				
5	rs1154405	A/G	ADH5	0.02	0.04	0.01	0	0.19	0.21	0.13	0.11				
6	rs1154401	C/G	ADH5	0.06	0.09	0.17	0.18	0.3	0.3	0.31	0.32				
7	rs1154400	C/T	ADH5	0.02	0.04	0.02	0.01	0.24	0.25	0.31	0.31				
8	rs1042364	A/G	ADH4	0.02	0.03	0.01	0	0.2	0.22	0.12	0.09				
9	rs1126671	A/G (Ile309Val)	ADH4	0.06	0.09	0.16	0.17	0.25	0.27	0.19	0.16				
10	rs6836440	A/G	ADH4	0.05	0.06	0.15	0.17	0.06	0.05	0.08	0.07				
11	rs2213035	C/T	ADH4	0.02	0.03	0.01	0	0.2	0.22	0.11	0.09				
12	rs3762894	C/T	ADH4	0.004	0.02	0.01	0.03	0.16	0.12	0.22	0.18				
13	rs1984364	G/T	ADH4	0.02	0.03	0.01	0	0.19	0.22	0.12	0.09				
14	rs3857224	C/T	ADH6	0.5	0.48	0.48	0.47	0.32	0.32	0.47	0.43				
15	rs6833176	C/G	ADH6	0.5	0.5	0.46	0.46	0.5	0.5	0.19	0.18				
16	rs4699733	C/G	ADH6	0.48	0.5	0.45	0.46	0.41	0.37	0.4	0.35				
17	rs10008281	A/C	ADH6	0.02	0.04	0.01	0.01	0.18	0.16	0.26	0.28				
18	rs3819197	C/T	ADH1A	0.02	0.04	0.01	0.01	0.19	0.18	0.26	0.27				
19	rs1229967	C/G	ADH1A	0.41	0.39	0.27	0.24	0.28	0.24	0.09	0.08				
20	rs13134764	A/T	ADH1A	0.02	0.02	0.02	0.01	0.3	0.32	0.1	0.08				
21	rs904092	A/G	ADH1A	0.4	0.39	0.3	0.28	0.22	0.19	0.22	0.24				
22	rs1042026	A/G	ADH1B	0.02	0.04	0.01	0.01	0.25	0.24	0.11	0.08				
23	rs17033	A/G	ADH1B	0.48	0.45	0.48	0.44	0.13	0.14	0.08	0.08				
24	rs1229984	A/G (His48Arg)	ADH1B	0	0	0.002	0.01	0.002	0.01	0.01	0.01				
25	rs1353621	A/G	ADH1B	0.02	0.02	0.02	0.01	0.3	0.33	0.1	0.09				
26	rs1159918	G/T	ADH1B	0.04	0.06	0.03	0.03	0.47	0.42	0.31	0.29				

ID	SNP	Allele1/Allele2	Gene	Minor Allele Frequency											
				Plains Indians		Southwest Indians		Finnish Caucasians		African Americans					
				AD	Controls	AD	Controls	AD	Controls	AD	Controls	AD	Controls		
27	rs1693426	A/G	ADH1C	0.48	0.45	0.4	0.46	0.49	0.5	0.16	0.14				
28	rs1693482	T/C(Gln272Arg)	ADH1C	0.48	0.45	0.4	0.46	0.49	0.5	0.17	0.15				
29	rs904096	G/T	ADH1C	0.48	0.45	0.4	0.46	0.49	0.5	0.17	0.15				
30	rs698	A/G(Ile350Val)	ADH1C	0.44	0.39	0.29	0.27	0.5	0.5	0.17	0.15				
31	rs1614972	C/T	ADH1C	0.08	0.07	0.04	0.04	0.3	0.28	0.48	0.49				
32	rs729147	A/G	Intergenic	0.49	0.46	0.3	0.38	0.25	0.28	0.17	0.14				
33	rs894369	C/G	ADH7	0.49	0.46	0.3	0.38	0.25	0.28	0.17	0.14				
34	rs1154454	C/T	ADH7	0.01	0.01	0.01	0.01	0.12	0.12	0.42	0.42				
35	rs1154456	C/T	ADH7	0.31	0.32	0.37	0.35	0.28	0.26	0.18	0.18				
36	rs1154460	A/G	ADH7	0.46	0.45	0.29	0.29	0.41	0.4	0.47	0.49				
37	rs971074	A/G(Arg230Arg)	ADH7	0.24	0.23	0.07	0.07	0.13	0.14	0.19	0.18				
38	rs1573496	C/G(Gly92Ala)	ADH7	0.03	0.04	0.04	0.04	0.12	0.13	0.03	0.02				
39	rs1154469	A/G	ADH7	0.31	0.32	0.36	0.35	0.34	0.34	0.18	0.18				
40	rs1154470	A/G	ADH7	0.31	0.32	0.37	0.35	0.34	0.34	0.15	0.14				
41	rs3764435	A/C	ALDH1A1	0.37	0.32	0.39	0.44	0.49	0.43	0.18	0.17				
42	rs1888202	C/G	ALDH1A1	0.39	0.34	0.38	0.42	0.49	0.48	0.18	0.18				
43	rs63319	A/C	ALDH1A1	0.35	0.33	0.43	0.47	0.49	0.45	0.3	0.31				
44	rs348457	C/G	ALDH1A1	0.3	0.3	0.5	0.42	0.42	0.48	0.44	0.42				
45	rs2303317	G/T	ALDH1A1	0.31	0.28	0.5	0.44	0.55	0.47	0.23	0.24				
46	rs2773806	A/G	ALDH1A1	0.01	0.01	0	0	0.02	0.04	0.26	0.29				
47	rs1424482	C/T	ALDH1A1	0.28	0.27	0.42	0.49	0.29	0.34	0.37	0.33				
48	rs8187876	A/G	ALDH1A1	0.26	0.26	0.41	0.49	0.1	0.07	0.21	0.21				
49	rs11143429	C/T	ALDH1A1	0.26	0.27	0.44	0.48	0.39	0.42	0.23	0.21				
50	rs1364451	A/G	ALDH1A1	0.01	0.02	0.01	0	0.13	0.16	0.27	0.3				
51	rs6560311	G/T	ALDH1A1	0.26	0.25	0.43	0.49	0.26	0.25	0.34	0.32				
52	rs2249978	C/T	ALDH1A1	0.29	0.27	0.42	0.49	0.25	0.25	0.37	0.35				
53	rs1418187	C/T	ALDH1A1	0.29	0.27	0.42	0.49	0.22	0.21	0.49	0.47				
54	rs4745209	C/T	ALDH1A1	0.29	0.27	0.42	0.49	0.22	0.21	0.4	0.4				
55	rs7862749	C/T	ALDH1A1	0.28	0.22	0.37	0.4	0.27	0.28	0.1	0.09				

ID	SNP	Allele1/Allele2	Gene	Minor Allele Frequency											
				Plains Indians		Southwest Indians		Finnish Caucasians		African Americans					
				AD	Controls	AD	Controls	AD	Controls	AD	Controls	AD	Controls		
56	rs4406477	C/T	ALDH1A1	0.29	0.24	0.38	0.41	0.32	0.36	0.48	0.49				
57	rs11143443	C/T	ALDH1A1	0	0.01	0	0	0	0.01	0.21	0.27				
58	rs2238151	C/T	ALDH2	0.49	0.46	0.35	0.35	0.34	0.38	0.14	0.11				
59	rs2238152	G/T	ALDH2	0.04	0.08	0.25	0.28	0.16	0.16	0.19	0.19				
60	rs4648328	C/T	ALDH2	0.04	0.08	0.25	0.28	0.16	0.16	0.18	0.19				
61	rs7311852	C/G	ALDH2	0	0	0	0	0	0	0.05	0.08				
62	rs10849970	A/G	ALDH2	0.29	0.31	0.22	0.23	0	0.01	0.05	0.06				
63	rs4646778	A/C	ALDH2	0.04	0.08	0.25	0.27	0.16	0.16	0.18	0.19				
64	rs7296651	C/G	ALDH2	0.33	0.4	0.46	0.47	0.16	0.17	0.34	0.31				

Note: SNPs showing (uncorrected) significant case-control associations are bolded.

Alleles 1 & 2 are defined using the Illumina top/bottom strand convention (http://www.illumina.com/documents/products/technote_topbot.pdf) and have been converted to the corresponding dbSNP alleles (<http://www.ncbi.nlm.nih.gov/>).

Table 2

The ADH5-ADH4 and ADH6-ADH1A-ADH1B haplotype blocks; haplotype frequencies in alcoholics and controls

Haplotype	Plains Indians		Southwest Indians		Finnish Caucasians		African Americans	
	AD	Controls	AD	Controls	AD	Controls	AD	Controls
ADH5-ADH4 block								
111121211112	0.01	0.02			0.04	0.05	0.06	0.05
2211112221211							0.11	0.13
2212111212211	0.05	0.05	0.15	0.17	0.06	0.05	0.01	0.01
2212121221211	0.93	0.89	0.82	0.81	0.59	0.60	0.46	0.50
2212121221221							0.17	0.14
2212212111112	0.01	0.01	0.01		0.13	0.13	0.03	0.02
2222121221221		0.02	0.01	0.01	0.11	0.09	0.03	0.02
ADH6-ADH1A-ADH1B block								
1211121111211							0.16	0.18
1211121221212	0.01*	0.04	0.01	0.01	0.16	0.15	0.08	0.07
1212221211211			0.03*	0.08			0.11	0.14
1212221212211	0.47	0.44	0.44	0.43	0.13	0.15	0.05	0.05
2112222211222	0.02	0.01	0.01	0.01	0.13	0.14	0.03	0.03
2122211111211	0.38	0.37	0.22	0.22	0.20	0.17	0.04	0.03
2122211221212					0.05	0.05	0.01	0.01
2122221111211	0.02	0.02	0.07	0.06	0.01	0.02	0.01	0.01
2122221211211	0.05	0.07	0.15	0.17	0.07	0.06	0.03	0.04
2212222211222			0.01		0.13	0.15	0.06	0.04
2222221211211				0.01	0.03	0.02	0.17	0.14
1212221211212							0.06	0.07
§ N	452	220	634	144	464	386	534	844

§N: the number of haplotypes. AD = alcohol dependence.

* P value ≤ 0.05 .

The ADH5-ADH4 block includes SNP1 to SNP13; the ADH6-ADH1A-ADH1B block includes SNP14 to SNP26 as listed in Table 1.

The table lists haplotypes that have a frequency $\geq 5\%$ in at least one population.

Table 3

The ADH1C haplotype block: haplotype frequencies in alcoholics and controls

Haplotype	Plains Indians		Southwest Indians		Finnish Caucasians		African Americans	
	AD	Controls	AD	Controls	AD	Controls	AD	Controls
12112	0.45	0.49	0.56	0.50	0.20	0.22	0.36	0.34
21212	0.03	0.05	0.09	0.14	0.01			
21222	0.44	0.39	0.31	0.32	0.50	0.50	0.16	0.14
12111	0.08	0.07	0.04	0.04	0.30	0.28	0.48	0.51
ADH1C*1 21	0.53	0.56	0.60	0.54	0.50	0.50	0.84	0.85
ADH1C*2 12	0.44	0.39	0.31	0.32	0.50	0.50	0.16	0.14
§N	452	220	632	144	464	388	534	844

The table lists haplotypes that have a frequency $\geq 5\%$ in at least one population.

§N: the number of haplotypes. AD = alcohol dependence Markers in the ADH1C haplotypes are from left to right: SNP27 to SNP31 (as listed in Table 1).

Markers in ADH1C*1 and ADH1C*2 are: SNP28 (rs1693482 Gln272Arg) and SNP30 (rs698 Ile350Val).

The ADH1C*1 21 haplotype: Arg272, Ile350.

The ADH1C*2 12 haplotype: Gln272, Val350.

Table 4

The ADH7 haplotype block: haplotype frequencies in alcoholics and controls

Haplotype Blocks	Plains Indians		Southwest Indians		Finnish Caucasians		African Americans	
	AD	Controls	AD	Controls	AD	Controls	AD	Controls
Intergenic Block								
11	0.38	0.34	0.29*	0.38	0.76	0.72	0.83	0.86
22	0.63	0.66	0.71*	0.63	0.24	0.28	0.16	0.14
ADH7 Block								
212111	0.31	0.32	0.63	0.65	0.27	0.26	0.14	0.12
111122	0.21	0.20	0.03	0.03	0.01	0.01	0.16	0.15
111222	0.03	0.04	0.04	0.04	0.12	0.13	0.03	0.02
122122	0.46	0.45	0.29	0.28	0.53	0.52	0.47	0.48
122111					0.06	0.08		
112122							0.15	0.15
$\sum N$	448	218	634	144	462	388	534	844

The table lists haplotypes that have a frequency $\geq 5\%$ in at least one population.

$\sum N$: the number of haplotypes. AD = alcohol dependence

* P value < 0.05.

Intergenic block includes SNPs 32-33, ADH7 block includes SNPs 35-40 as listed in Table 1. SNP 34 (rs1154454) is not in either haplotype block.

Table 5

Haplotype blocks across ALDH1A1: haplotype frequencies in alcoholics and controls

Haplotype	Plains Indians		Southwest Indians		Finnish Caucasians		African Americans	
	AD	Controls	AD	Controls	AD	Controls	AD	Controls
Block1								
222	0.59	0.65	0.56	0.52	0.45*	0.38	0.12	0.10
211	0.02	0.01	0.03	0.01	0.01	0.02	0.05**	0.02
122	0.02				0.01*	0.03	0.04	0.06
112	0.02	0.04			0.03	0.03	0.12	0.13
111	0.33	0.27	0.34	0.40	0.44	0.47	0.64	0.63
Block2								
21	0.68	0.69	0.49	0.42	0.55*	0.47	0.23	0.24
22	0.02				0.03	0.05	0.33	0.34
12	0.29	0.28	0.50	0.56	0.42	0.48	0.44	0.42
Block3								
1222221	0.68	0.71	0.56	0.48	0.61	0.56	0.17	0.13
1212221	0.01		0.01	0.01	0.10	0.11	0.04	0.05
2212221		0.01			0.03*	0.05	0.14	0.16
2211112	0.01		0.01		0.13	0.14	0.08	0.09
2111112	0.24	0.25	0.40*	0.51	0.10	0.06	0.20	0.19
2212112							0.06	0.07
2212121							0.09	0.09
1211112							0.05	0.03
1222111							0.03	0.05
Block4								
21	0.72	0.76	0.63	0.59	0.67	0.63	0.52	0.51
22		0.01	0.01	0.01	0.06	0.09	0.38	0.40
12	0.28	0.22	0.37	0.40	0.26	0.27	0.10	0.09
§N	452	218	630	144	462	388	534	844

The table lists haplotypes that have a frequency $\geq 5\%$ in at least one population.

^SN: the number of haplotypes. AD = alcohol dependence

* p value <0.05

** p<0.01.

The numbering of SNPs corresponds to Table 1: Block 1: SNP1-43; Block 2: SNP44-45; Block 3: SNP47-49, 51-54; Block 4: SNP55-56. The bolded numbers indicate the frequencies of the yin yang haplotypes.

Table 6
The ALDH2 haplotype block: haplotype frequencies in alcoholics and controls

Haplotype	Plains Indians		SW Indians		Finnish Caucasians		African Americans		Haplotype		African Americans		
	AD	Controls	AD	Controls	AD	Controls	AD	Controls	SNP58-64	AD	Controls	AD	
SNP58-64 noSNP61													
222121	0.29	0.31	0.22	0.23	0.01	0.01	0.05	0.05	2222121	0.05	0.05	0.05	0.05
222222	0.17	0.15	0.18	0.15	0.21	0.21	0.20	0.20	2222222	0.20	0.20	0.20	0.20
211211	0.04*	0.08	0.24	0.24	0.16	0.16	0.18	0.19	2112211	0.18	0.19	0.18	0.19
122222	0.49	0.46	0.35	0.35	0.62	0.62	0.14	0.11	1222222	0.14	0.11	0.14	0.11
222221					0.01	0.01	0.43	0.44	2222221	0.38	0.44	0.38	0.37
§N	452	220	630	144	464	388	534	844	2221221	0.05*	0.08	534	844

The table lists haplotypes that have a frequency $\geq 5\%$ in at least one population.

§N: the number of haplotypes. AD = alcohol dependence

* P value < 0.05

The markers are in the order of SNP 58 to 64, as listed in Table 1. Analyses were undertaken without SNP61 which was monomorphic in all populations except African Americans. The analysis was then conducted with all 7 SNPs in African Americans.