

RESEARCH ARTICLE

# Genetic variants in *ALDH1B1* and alcohol dependence risk in a British and Irish population: A bioinformatic and genetic study

Michael J. Way<sup>1,2</sup>, M. Adam Ali<sup>1,2</sup>, Andrew McQuillin<sup>1</sup>, Marsha Y. Morgan<sup>2\*</sup>

**1** Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London, United Kingdom, **2** UCL Institute for Liver & Digestive Health, Department of Medicine, Royal Free Campus, University College London, London, United Kingdom

\* [marsha.morgan@ucl.ac.uk](mailto:marsha.morgan@ucl.ac.uk)



**OPEN ACCESS**

**Citation:** Way MJ, Ali MA, McQuillin A, Morgan MY (2017) Genetic variants in *ALDH1B1* and alcohol dependence risk in a British and Irish population: A bioinformatic and genetic study. PLoS ONE 12(6): e0177009. <https://doi.org/10.1371/journal.pone.0177009>

**Editor:** Yong-Gang Yao, Kunming Institute of Zoology, Chinese Academy of Sciences, CHINA

**Received:** October 12, 2016

**Accepted:** April 20, 2017

**Published:** June 8, 2017

**Copyright:** © 2017 Way et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. Further details may be obtained by contacting the corresponding author ([marsha.morgan@ucl.ac.uk](mailto:marsha.morgan@ucl.ac.uk)).

**Funding:** The research was funded by a PhD studentship award jointly funded by University College London and the Hobson Charity Limited.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Alcohol is metabolized in the liver *via* the enzymes alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH). Polymorphisms in the genes encoding these enzymes, which are common in East Asian populations, can alter enzyme kinetics and hence the risk of alcohol dependence and its sequelae. One of the most important genetic variants, in this regards, is the single nucleotide polymorphism (SNP) rs671 in *ALDH2*, the gene encoding the primary acetaldehyde metabolizing enzyme ALDH2. However, the protective allele of rs671 is absent in most Europeans although *ALDH1B1*, which shares significant sequence homology with *ALDH2*, contains several, potentially functional, missense SNPs that do occur in European populations. The aims of this study were: (i) to use bioinformatic techniques to characterize the possible effects of selected variants in *ALDH1B1* on protein structure and function; and, (ii) to genotype three missense and one stop-gain, protein-altering, non-synonymous SNPs in 1478 alcohol dependent cases and 1254 controls of matched British and Irish ancestry. No significant allelic associations were observed between the three missense SNPs and alcohol dependence risk. The minor allele frequency of rs142427338 (Gln378Ter) was higher in alcohol dependent cases than in controls (allelic  $P = 0.19$ , OR = 2.98, [0.62–14.37]) but as this SNP is very rare the study was likely underpowered to detect an association with alcohol dependence risk. This potential association will need to be further evaluated in other large, independent European populations.

## Introduction

Alcohol is metabolized in the liver to acetaldehyde and acetate *via* the enzymes alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH). Functional polymorphisms in the genes encoding ADH and ALDH are associated with alterations in enzyme kinetics [1, 2], which, in turn, may determine the rates of production and removal of the toxic intermediate metabolite acetaldehyde. High circulating acetaldehyde concentrations are associated with a number of unpleasant systemic effects, including: facial flushing, tachycardia, nausea,

headache and even collapse and death [3, 4]. In consequence, individuals carrying certain functional *ADH* and *ALDH* variants tend to limit their alcohol consumption and so are, in effect, protected from developing alcohol dependence and the medical consequences of harmful drinking [5–8].

There are significant differences in the frequencies of these functional *ADH* and *ALDH* variants between ethnic groups. The prevalence of these variants is higher in the Far East and very much lower in Europe. Thus, for example, the missense single nucleotide polymorphism (SNP) rs1229984 (Arg48His) in *ADH1B*, which increases the maximal velocity at which alcohol is oxidized to acetaldehyde by over 100-fold [3, 9], is found in 19 to 91% of East-Asians and 10 to 70% of West-Asians [7], but at rates ranging from zero to 10% in other populations [10,11]. Nevertheless, there is consistent evidence that the protection against the development of alcohol dependence conferred by carriage of this variant extends to these populations as well [7,12–15].

Similarly, a missense SNP, rs671 (Glu504Lys) in *ALDH2*, which inactivates the mitochondrial isoform of ALDH2 causing a loss of up to 90% of its enzymatic activity and 20-fold elevations in circulating acetaldehyde levels following alcohol consumption, is found in 30 to 50% of East-Asians [16] but is absent in Europeans [17]. However, other genes, for example *ALDH1B1*, have the capacity to oxidize acetaldehyde at physiologically relevant rates [18] and may be of importance in relation to alcohol dependence in European populations. *ALDH1B1* is located on chromosome 9 and its encoded protein, ALDH1B1, has 75% amino acid sequence homology with ALDH2. In addition, in common with ALDH2, it is expressed in liver mitochondria and is predicted to form homo-tetramers comprised of four ALDH1B1 monomer subunits [18]. Computational analysis predicts that ALDH1B1 monomer subunits preferentially hetero-tetramerize with ALDH2 rather than forming homo-tetramers amongst themselves; this provides a hypothetical explanation for why inactivation of ALDH2 in carriers of the *ALDH2* 504Lys allele is not compensated for by ALDH1B1 [19].

A number of missense SNPs have been identified in *ALDH1B1*, which are thought likely to alter the structure of its protein product the enzyme ALDH1B1 [20–22]. Jackson et al [19], undertook computational modelling of human recombinant ALDH1B1 and investigated the functional effects of three non-synonymous SNPs in *ALDH1B1* viz: rs2073478 (Arg107Leu), rs2228093 (Ala86Val) and rs4878199 (Val253Met) *in vitro*. They showed that presence of the Val86 allele of rs2228093 abolished enzyme activity whereas the other two SNPs appeared to have no effect. The association between two of these SNPs, rs2073478 (Arg107Leu) and rs2228093 (Ala86Val) and alcohol consumption levels and the risk for developing alcohol dependence has been explored but with no robust or replicated findings [20,23–25]. Thus, there is some functional evidence suggesting that non-synonymous SNPs in *ALDH1B1* might affect enzyme activity and hence alcohol metabolism. There is also limited and conflicting genetic evidence for their role in modifying the risk for developing alcohol dependence in man.

The primary aims of this study were: (i) to use bioinformatic techniques to identify and characterize non-synonymous SNPs in *ALDH1B1*; and, (ii) to investigate, using a case control design, whether these functional variants associate with alcohol dependence risk in a large, carefully characterized population of British and Irish ancestry.

## Materials and methods

### Bioinformatic analysis

**Selecting variants in *ALDH1B1*.** A search for nonsense variants with any notated frequency, and missense variants with minor allele frequencies (MAF) exceeding 5% in

**Table 1. PolyPhen-2\* functional prediction analysis of the three *ALDH1B1* missense variants.**

SNP identity	Amino acid change	PolyPhen-2 score (0–1)	PolyPhen-2 prediction
rs2073478	Arg107Leu	0.46	Possibly damaging
rs2228093	Ala86Val	0.99	Probably damaging
rs4878199	Val253Met	0	Benign

\*Adzhubei et al., 2010 [26]

Abbreviations: SNP = single nucleotide polymorphism

<https://doi.org/10.1371/journal.pone.0177009.t001>

populations of European ancestry was undertaken in the 1000 Genomes Project build 37 phase I integrated data in the relevant region (chr9: 38,395,746–38,397,299) [21].

Four, non-synonymous variants fulfilled the selection criteria: three were missense variants; rs2073478 (Arg107Leu), rs2228093 (Ala86Val) and rs4878199 (Val253Met); and one stop/gain variant, rs142427338 (Gln378Ter). These variants are common in the reference 1000 genomes European ancestry sub-populations [21]. The allele frequencies for rs2073478 (Arg107Leu) and rs2228093 (Ala86Val) are largely similar in the British, Toscani and Iberian groups, with the frequency distributions in the Finnish group varying most noticeably from the rest (S1 Fig).

**Missense variant functionality prediction.** The likelihood that the three missense variants were damaging was investigated using the functional prediction software PolyPhen-2 [26]; rs2073478 (Arg107Leu) and rs2228093 (Ala86Val) were predicted to be damaging while the mutation in the third, rs4878199 (Val253Met) was predicted to be benign (Table 1). This software cannot be used to predict the potential damaging effects of loss of function variants, such as rs142427338 (Gln378Ter), as the rare allele of this SNP encodes a stop codon.

**Structural homology modelling.** Structural homology models of the monomeric unit of wild type ALDH1B1 (GI: 25777730) were generated using the I-TASSER computational modelling server [27–29]. The top ranked homology model was derived from an alignment with the human ALDH2 protein structure (PDB 1a4zA) (S1 Table). The accuracy of this model was estimated by the I-TASSER server using the topology modelling score [30]. It was subsequently visualized using UCSF Chimera [31]. The model of the wild-type ALDH1B1 was subsequently modified by introducing relevant amino acid substitutions using the rotamer function of UCSF Chimera [31] to allow visualization of the positioning of the three missense variants within the overall structure.

Homology modelling was also used to visualize the structural effects of the stop gain variant rs142427338 (Gln378Ter) and hence the potential damaging effects of this SNP.

**Molecular dynamic simulations.** Molecular dynamics simulations were performed on both the wild-type and the rs2228093 (Ala86Val) variant *ALDH1B1* homology models using GROMACS 5.0 [32, 33]; the rs2228093 variant was selected for study as it is predicted to be the most damaging of the three missense SNPs. The force-field used in all simulations was OPLA-AA/L all-atom force field [34]. The structures were immersed in cubic boxes (size = 10.8 nm<sup>3</sup>) containing 39842 water molecules. Sodium ions were added for charge neutralization. The Particle Mesh Ewald method [35] was used to treat long range electrostatic interactions and the structures relaxed through steepest descent energy minimization runs until the maximum force was less than 1000 kJ.mol<sup>-1</sup>.nm<sup>-2</sup>. The solvent and ions were then equilibrated with the starting structure through two steps of molecular dynamics equilibration runs. The first equilibration was run with a constant number of particles, volume and temperature (NVT) over a period of 100 picoseconds; the presence of a stable temperature profile

throughout the run confirmed equilibrium. The second equilibration was run with a constant number of particles, pressure and temperature (NPT) over a period of 100 picoseconds; the presence of stable pressure and density over the run confirmed equilibrium. The final production runs of the molecular dynamics were performed over 100 nanoseconds under NPT conditions.

**Trajectory analysis.** The atomic movements within the output trajectory files from the molecular dynamics simulations were analyzed by root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyses as performed in GROMACS 5.0 [32,33]. RMSF values were calculated between time points of the simulation at equilibrium indicated by a plateau in the trace of the RMSD trajectories. The RMSD and RMSF values of the variant and wild-type structures were directly compared by visualization of the data points over the entire 100 ns of simulation or the amino-acid sequence of ALDH1B1, respectively. The RMSD and RMSF values were also compared by their average values (mean  $\pm$ 1 standard deviation). Trajectories were animated using visual molecular dynamics (VMD) software [36].

## Genetic association study

**Alcohol dependent cases.** Individuals attending a variety of UK community and hospital-based services providing support and treatment for alcohol use disorders, between 1997 and 2014, were screened for eligibility. They were included in the study if: (i) they met the criteria for alcohol dependence specified in the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV)* [37] or the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)* [38]; (ii) if they were of English, Scottish, Welsh or Irish but not Jewish descent with a maximum of one grandparent of European ancestry; and, (iii) they were not related to any of the other included subjects.

**Controls.** Ancestrally-matched controls were recruited from London branches of the National Health Service (NHS) blood transfusion service, from General Practitioners' surgeries, from amongst university students, and from the general public through the National Institute of Health Research (NIHR) funded Mental Health Research Network (MHRN). Individuals were excluded if they met the schedule for affective disorders and schizophrenia lifetime version (SADS-L) criteria for a history of depression, bipolar disorder, schizophrenia or alcohol/drug use disorders [39]; none of the control subjects had a family history of bipolar disorder, schizophrenia or alcohol dependence; none currently drank alcohol above a weekly maximum of 168 g for men or 112 g for women, nor had done so at any time in the past. In addition, DNA from a separate set of controls of British ancestry, not screened for psychiatric or alcohol use disorders, was purchased from the European Collection of Cell Cultures (ECACC; Health Protection Agency Culture Collections, Salisbury, UK).

None of the cases or controls utilized in this study had participated in any previous studies on ALDH genes polymorphisms or in GWAS studies of alcohol dependence.

**DNA extraction and genotyping.** Genomic DNA was extracted from whole blood from using a standard cell lysis, phenol chloroform technique [40]. The ECACC DNA was pre-extracted from transformed lymphoblastoid cell lines.

Primers for genotyping were designed using PrimerPicker software (<http://www.lgcgroup.com/genotyping/>) (Table 2). Genotyping for the *ALH1B1* SNPs was carried out, in-house, using fluorescent competitive allele specific PCR reagents (KASPar; LGC Genomics, Hoddeston, UK). Amplification and detection was undertaken using a LightCycler<sup>®</sup> 480 real time PCR system (Roche Applied Science, Burgess Hill, UK). Genotype calling was performed automatically by built-in Roche software of cluster plots with some manual editing of calls.

**Table 2. Primers used for genotyping the four selected SNPs in *ALDH1B1*.**

SNP identity	Sequence of primers
rs2073478	GAAGGTGACCAAGTTCATGCTCCACTAGGTCTGCCAGGC
	GAAGGTCGGAGTCAACGGATTGCTCCACTAGGTCTGCCAGGA
	GGTCCCCATGGCGCCGGAT
	AGCGGGGCCGGCTGCTGAA
rs2228093	GAAGGTGACCAAGTTCATGCTGTGAAAGCAGCCCGGGAAGC
	GAAGGTCGGAGTCAACGGATTCTGTGAAAGCAGCCCGGGAAGT
	GAGGCATCCATCCGGCGCCAT
	CATGGGGACCCAGGCGGAA
rs4878199	GAAGGTGACCAAGTTCATGCTGAAGGCAACTTTGTCAACATCCAC
	GAAGGTCGGAGTCAACGGATTGTGAAGGCAACTTTGTCAACATCCAT
	GTGCGGCCATCGCCAGCA
	CCCAACAGCAGGTGCGGCCAT
rs142427338	GAAGGTGACCAAGTTCATGCTAGGCTACATCCAGCTTGGCC
	GAAGGTCGGAGTCAACGGATTCTAGGCTACATCCAGCTTGGCT
	GCTCTCCGCCACAGAGGAGTTT
	CAGAGGAGTTTTGCGCCCTCCTT

Abbreviations: SNP = Single nucleotide polymorphism

<https://doi.org/10.1371/journal.pone.0177009.t002>

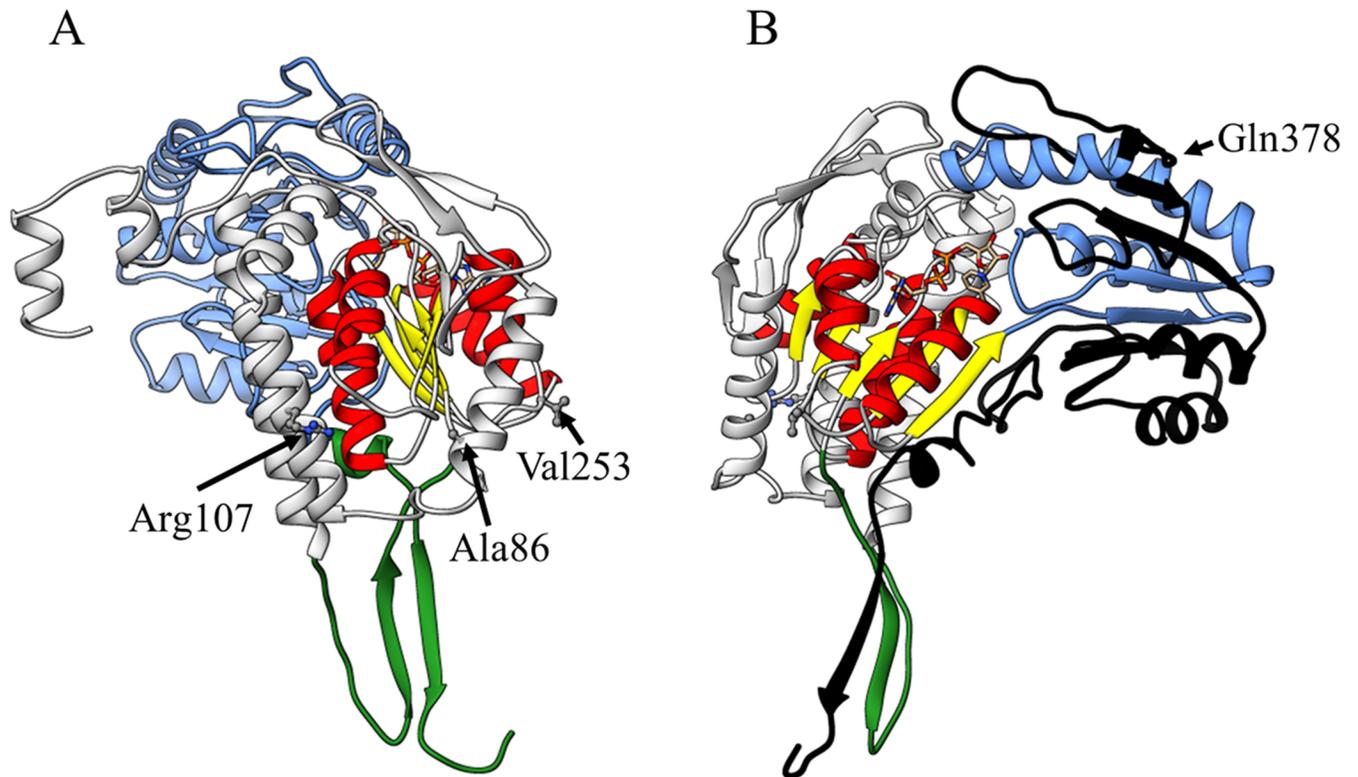
Approximately 12% of samples, randomly selected *a priori*, were genotyped in duplicate to ensure consistent allele calling.

**Sequencing of rs142427338.** Samples identified as carrying the minor allele of the stop gain variant rs142427338 (Gln378Ter), together with samples of uncertain genotype for this SNP, were directly sequenced in order to validate the KASPar genotype calling. Sequencing primers were designed using Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (Forward–CTGACATGGAGCATGCCGT; Reverse–CAGATCCCGGGTGAACACAG). A standard touchdown PCR protocol was performed [41] on 25ng of genomic DNA using a GeneAmp PCR System 9700 machine (Invitrogen, Paisley, UK) with an increased annealing temperature of 68°C to prevent non-specific amplification. PCR amplification success and specificity were determined by visualizing the fragments with agarose gel electrophoresis. Amplified PCR templates were cleaned using PEG precipitation (0.5M NaCl, 1 mM Tris-HCl pH 8.0, 0.1 mM EDTA, PEG 8000 20%, 1.75 mM MgCl<sub>2</sub>) prior to sequencing. Sequencing reactions were performed using a BigDye terminator kit (Applied Biosystems®; Haywards Heath; UK) with an ABI 3730xl DNA Analyzer (Applied Biosystems®, Life Technologies, California; USA). Sequencing data were read and analyzed using the Staden package [42].

**Data analysis.** Test for primary allelic associations, missingness, deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium were performed using PLINK version 1.9 [43,44]. Samples with conflicting calls (<0.05% of total) were excluded from further analyses.

### Ethics

United Kingdom National Health Service Multicentre Research Ethics Committee approval was granted for this study (MREC/03/11/090). This was ratified by the Research and Development Departments associated with the individual participating centres. Written informed consent was obtained from all subjects prior to inclusion.



**Fig 1. Homology model structure of the wild-type ALDH1B1 primary sequence (GI: 25777730).** This model structure is annotated and rotated to illustrate the positionings of: A) the three missense variants; and, B) the stop codon variant.

<https://doi.org/10.1371/journal.pone.0177009.g001>

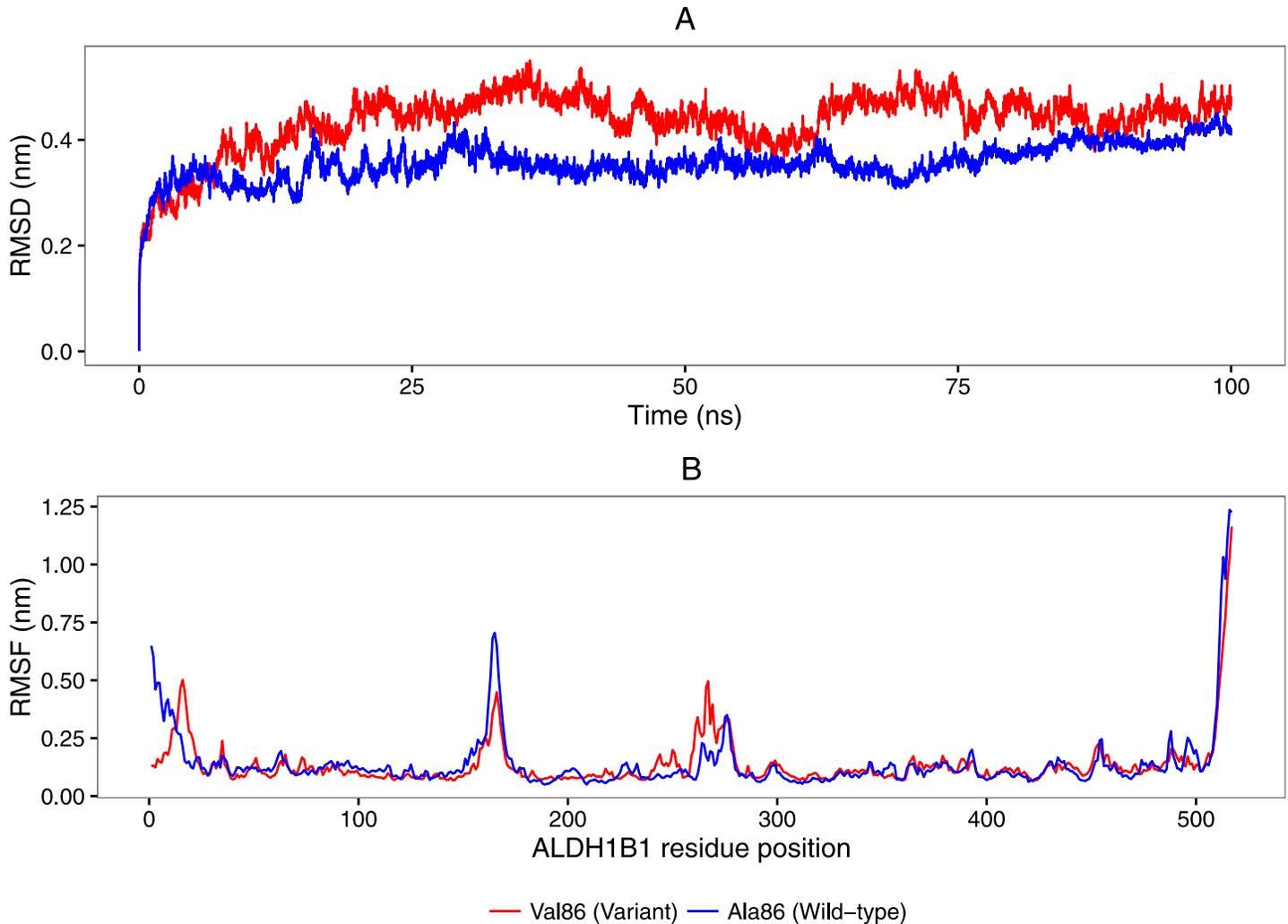
## Results

### Bioinformatic analysis

**Predicted Structure of ALDH1B1.** The top ranked predicted structure for the ALDH1B1 monomeric unit had a 95% structural overlap and 76% amino acid sequence identity with the ALDH2 structure 1A4Z. The estimated template modelling score between the homology model structure and the undetermined native structure was  $0.69 \pm 0.12$ .

The key structural features of the predicted ALDH1B1 monomer subunit were: (i) a dinucleotide binding domain; (ii) an oligomerization domain; and, (iii) a catalytic domain (Fig 1A; S1 Video). The three missense variants rs2073478 (Arg107Leu), rs2228093 (Ala86Val) and rs4878199 (Val253Met) are predicted to locate to the co-enzyme domain facing the outer surface of the enzyme (Fig 1A). The nonsense variant rs142427338 (Gln378Ter) is predicted to result in a premature transcription termination codon within ALDH1B1 and hence loss of 139 amino acids in the catalytic domain of the enzyme crucial for its function (Fig 1B).

**Molecular dynamics simulations.** The RMSD of the carbon  $\alpha$ backbone of the wild-type ALDH1B1 monomer indicates that the structures attained equilibrium deviation after 10 ns (Fig 2A). The RMSF indicates that this monomer contains four major regions of structural flexibility *viz*: (i) a region around residues 1–17 in the N-terminal, which lies within the mitochondrial targeting sequence of the enzyme; (ii & iii) two regions around residues 150–170 and 510–517 which, despite being distant in the primary sequence, are sterically close within the monomer structure and encompass the majority of the oligomerization domain; and, (iv) a



**Fig 2. Molecular dynamics simulations performed on the structural homology models of the wild type and the 86Val variant structure of ALDH1B1.** A—The RMSD\* of carbon  $\alpha$  backbone atoms over the course of the entire 100 ns molecular dynamics simulation for the wild-type (blue) and Val86 variant (red) structure. B—The RMSF\*\* of carbon  $\alpha$  backbone atoms over the last 90 ns of the molecular dynamics simulation for the wild-type (blue) and Val86 variant (red) structure. \* The root mean squared deviation (RMSD) is a measure the average distance between a number of atomistic positions between a reference structural model and the structural model at a point in time during the simulation. Its variability over the time-course of the simulation indicates the scale of atomistic movement in a structural model and whether it has achieved dynamic equilibrium. \*\* The root mean squared fluctuation (RMSF) is a calculation of the average RMSD in a well-defined atomistic position, typically atoms present in key regions of the protein structure, over a specified time-period of the simulation. The RMSF highlights differences in movement between in a protein structure over a time period of a simulation and is useful for comparing two or more simulation trajectories.

<https://doi.org/10.1371/journal.pone.0177009.g002>

region around residues 270–280 which contains an alpha-helix  $\alpha D$ , which is a key component of the dinucleotide binding domain within the Rossmann fold (Fig 2B; S1 Video).

The simulation trajectory of the Val86 variant monomer differed from that of the wild type (average RMSD  $\pm$  SD =  $0.44 \pm 0.055$  nm vs. average RMSD  $\pm$  SD =  $0.35 \pm 0.034$  nm respectively), suggesting that the Val86 substitution reduces the structural stability *in silico*. The RMSF data generally mirrored those of the wild-type, except for: (i) a reduction in the flexibility between residues in positions 150–170, which correspond to a region of the oligomerization domain; and, (ii) an increase of the flexibility between residues in positions 270–280, which correspond to a key component of the dinucleotide binding domain.

**Table 3. Single marker allelic associations for genotyped ALDH1B1 polymorphisms in alcohol dependent cases and controls.**

SNP	Cohort	Number	Minor allele	Genotype counts			MAF (%)	Significance (P value)*	OR (95% CI)
				TT	TG	GG			
rs2073478	Case	1446	G	508	698	240	40.7	0.75 (1)	1.03 (0.88–1.20)
	Control	1244		433	604	207	40.9		
rs2228093	Case	1450	T	1080	346	24	13.6	0.91 (1)	0.99 (0.89–1.11)
	Control	1245		939	282	24	13.3		
rs4878199	Case	1415	G	1299	114	2	4.2	0.42 (1)	0.90 (0.69–1.17)
	Control	1232		1122	106	4	4.6		
rs142427338	Case	1472	T	1466	6	0	0.2	0.19 (0.77)	2.98 (0.62–14.37)
	Control	1252		1250	2	0	0.1		

Abbreviations: SNP = Single Nucleotide Polymorphism; MAF = Minor Allele Frequency;  $\chi^2$  = Chi-squared statistic; OR = Odds Ratio; CI = Confidence Intervals.

\*The unadjusted significance value of an allelic Fishers exact test; the value in brackets has been adjusted for multiple testing using Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0177009.t003>

### Genetic association analysis

**Genotyping accuracy.** Genotyping was completed for all four SNPs with call rates all greater than 95%. All markers followed Hardy Weinberg Equilibrium (HWE; cut-off > 0.05) in cases and controls.

**Primary allelic associations of ALDH1B1 SNPs.** A total of 1478 alcohol dependent cases and 1254 controls were genotyped.

No significant differences were observed in the frequencies of the three missense ALDH1B1 SNPs between the cases and controls (Table 3). However, the prevalence of the allele of the rare rs142427338 (Gln378Ter) was higher in the cases (allelic = 0.19, odds ratio [OR] = 2.98, 95% confidence interval [CI] 0.62–14.37) (Table 3). The variants were in strong linkage disequilibrium.

### Discussion

The missense variant rs671 in ALDH2 significantly affects the risk of developing alcohol dependence and its sequelae in East Asian populations but this variant is largely absent in European populations. There is, however, significant sequence homology between ALDH2 and ALDH1B1 and similarities in their enzymes cellular positioning within the mitochondrion and in their tendency to forms homo-tetramers *in vitro*. Importantly, however, ALDH1B1, unlike ALDH2, contains several missense SNPs which are common in European populations. Nevertheless, it is unclear whether these variants affect the processing of acetaldehyde, and hence influence the risk of developing the adverse consequences associated with alcohol consumption. This was addressed in the present study, in which four non-synonymous variants in ALDH1B1, three missense and one stop gain, which result in changes to the primary sequence of ALDH1B1, were further characterized using functional prediction software, structural homology modelling and molecular dynamics simulation.

The structure of the ALDH1B1 monomer subunit was predicted using the top-ranked I-TASSER software [28]. The output metrics for the final homology model provided a high degree of confidence in the prediction [30]. The technique of molecular dynamics simulation

was used to infer the flexibility of this predicted ALDH1B1 structure and characterize the effects of the 86Val allelic variant on protein structure and function [45,46]. The simulations revealed four regions of flexibility. The first, located at the N-terminal of the protein (residues 1–17), encompassed the mitochondrial targeting sequence. The simulated flexibility in this mitochondrial targeting sequence agrees with Nuclear Magnetic Resonance (NMR) and crystallography studies, which have experimentally determined freedom of movement in this region in homologous ALDH proteins [47,48]. The second and third regions of flexibility (residues 150–170 and 510–517) lie in the oligomerization domain of ALDH1B1. Flexibility in this domain might reflect the fact that the simulation left this region exposed whereas in the putative ALDH1B1 tetramer the oligomerization domain would likely be buried and hence stabilized within the overall macromolecular structure. The fourth region of flexibility (residues 270–280) contains an alpha-helix  $\alpha D$  which is a key component of the dinucleotide binding domain. Structural studies of the homologous alpha-helix in ALDH2 comparing the *apo* and *holo* forms of the enzyme, have shown that movement of this alpha-helix is essential for isomerization of bound NAD, a significant and necessary feature of catalysis [49].

A comparison of the molecular dynamics trajectory of the predicted structures of the wild-type ALDH1B1 and the missense variant predicted to be most damaging *viz.* rs2228093 (Ala86Val) demonstrates a considerable increase in the flexibility of the 270 to 280 region. As this region of the protein is crucial for the binding of the coenzyme NAD, this observation could explain the catalytic inactivity of this variant previously observed *in vitro* [19]. Ultimately, however, this technique does not provide conclusive evidence of a functional effect of this genetic variant on protein structure but does provide a basis for later confirmation by X-ray crystallography or solution NMR. This, notwithstanding, other studies have reported [50] excellent correlations between molecular dynamics simulation and experimental data.

The variant, rs142427338 (Gln378Ter) in *ALDH1B1* encodes a stop codon rather than another amino-acid residue. Such 'loss of function' or 'nonsense' variants may exert functional effects at either the transcript level, *via* haploinsufficiency mechanisms, or at the protein level, *via* truncation mechanisms. Haploinsufficiency occurs when the mRNA transcript encoding a nonsense allele is eliminated *via* nonsense mediated decay (NMD). The process of NMD, however, requires splice sites for the detection of premature stop codons and thus does not occur on transcripts of intronless genes [51]; as *ALDH1B1* is intronless NMD seems unlikely. Thus, if this transcript is translated and the protein product does not invoke the unfolded protein response, the loss of function allele of *ALDH1B1* could result in loss of a sizeable portion of the final enzyme product encompassing the C-terminal 139 amino acids. This region of the protein contains the enzyme's catalytic residues and may also be important for tetramer formation. Hence, an ALDH1B1 protein truncated at position 378 would likely be non-functional. Clearly experimental studies would greatly facilitate functional understanding of this variant.

Despite the strength of evidence, from the bioinformatic analyses, for the presence of potentially damaging functional mutations in *ALDH1B1*, these coding variants were not significantly associated with alcohol dependence *per se* in a large well-characterized population of alcohol dependent cases and controls of British and Irish ancestry. However, the prevalence of the 378STOP allele in rs142427338 (Gln378Ter) was higher in cases than in controls although the difference was not significant. The MAFs of the three missense variants ranged from 4.6% to 40.9% and hence this study was sufficiently powered to detect significance between cases and controls. However, the stop gain variant is very rare with a MAF of 0.01; thus this study was likely underpowered to detect genetic association at the current sample size [52].

There are very few genetic association data available for comparison. Sherman et al. [20], found no association between carriage of either rs2073478 or rs2228093 and alcohol dependence risk in a cohort of 40 people of British ancestry genotyped using PCR-RFLP. Unfortunately, the

size of this population severely limits the statistical power at the expected effect size [53]. Additionally, the MAFs reported by Sherman *et al* [20] were notably different from those in the present study *viz* 29% and 41% respectively for rs2073478 and 24% and 13.3% respectively for rs2228093; this may reflect inaccuracies in the calculation of the population MAF due to the small size of the cohort and/or to inaccuracies arising from use of PCR-RFLP genotyping, which is prone to partial digestion and genotype reading errors [54]. However, the present study is sufficiently powered to allow confidence in the concordant, negative association [52].

The relationships between polymorphisms in *ALDH1B1* and alcohol drinking levels in the general population and alcohol-related hypersensitivity reactions have been examined in two large populations of Danish residents of Northern European origin [23–25]. The reported MAFs for the two SNPs examined were similar to those in the present study *viz* 39% and 41% respectively for rs2073478 and 12% and 13% respectively for rs2228093; thus the studies can be usefully compared.

In the first Danish study Husemoen *et al.* [23] obtained self-reported information on alcohol consumption from 1216 participants in the population-based Copenhagen Allergy Study and reported that carriage of the minor allele of rs2228093 was associated with non-drinking and *higher* weekly alcohol intake; these findings are paradoxical as the same directionality of allele effect between these related phenotypes would be expected.

In the second study, Linneberg *et al.* [25] combined the information on alcohol consumption obtained from their original cohort of 1216 with similarly collected information from an additional 6784 participants in a replication study. In addition they obtained self-reported information on alcohol hypersensitivity and allergic rhinitis from 957 (79%) of the original 1216 participants and 2419 (36%) of the 6784 replication cohort. In the combined population, carriers of the minor allele of rs2228093 reported significantly *higher* total alcohol consumption and an *increased* prevalence of alcohol hypersensitivity; this finding is again paradoxical as one would expect alcohol hypersensitivity to be associated with lower total alcohol consumption and thus expect a common risk allele between phenotypes. It is unclear why this apparent paradoxical effect occurs, although it is possible that the reported allergic reactions may have been too mild to warrant avoidance of alcohol. However, quantitative information on the severity of the response was not made available.

Emerging evidence suggests that rs2228093 has been under positive selection in East-Asian populations over the past 10,000 years [55]. Wang *et al.* [55] showed that several functional SNPs in alcohol metabolizing genes *viz* rs1229984 (*ADH1B*), rs671 (*ALDH2*), rs8187929 (*ALDH1A1*), rs3813867 (*CYP2E1*) and rs2031920 (*CYP2E1*), have been under moderate to strong selection, which appears to have been driven by the expansion of agriculture. Although these findings are not directly related to alcohol use phenotypes *per se*, they highlight the inter-related function of alcohol metabolizing genes and in particular the shared factors influencing allele frequencies in these populations. If selection pressures are similarly influencing these genes in Europeans, then it is possible that the resultant effects could confound genetic association studies of *ALDH1B1*, potentially explaining the negative findings observed in the present study and the paradoxical findings observed by the Danish groups [23,25]

## Conclusion

In conclusion: bioinformatic techniques have been used to investigate missense SNPs in *ALDH1B1*, in particular the 86Val allele of rs2228093, which is predicted to disrupt the structural flexibility of the protein product, the enzyme ALDH1B1. If, as indicated from the bioinformatic data, these changes influence NAD binding to the catalytic domain of the enzyme, or oligomerization between the enzyme subunits, then they could alter enzyme activity.

Functional and molecular cloning studies would be needed to further elucidate the effects of these polymorphisms and, in particular, their physiological role in alcohol pharmacokinetics. Although no genetic associations were identified between the four selected SNPs in *ALDH1B1* and alcohol dependence, the study was likely underpowered in relation to the stop gain variant rs142427338 due to its rarity in this population. Exploration of these *ALDH1B1* variants within other European populations is clearly warranted.

## Supporting information

**S1 Fig. *ALDH1B1* variant allele frequencies in four European ancestry 1000 genomes sub-populations.** Major and minor allele frequencies for the *ALDH1B1* variants rs2073478, rs2228093 and rs4878199 are shown. The four European ancestry sub-populations include: (A) Finnish in Finland (FIN); (B) British in England and Scotland (GBR); (C) Iberian Population in Spain (IBS); and, (D) Toscani in Italia (TSI).  
(TIFF)

**S1 Table. The top ten templates used by to generate the *ALDH1B1* structural homology model from the wild-type *ALDH1B1* primary sequence (GI: 25777730).** *PDB hit*: The protein databank ID of the template structure used to generate the *ALDH1B1* homology model. *Identity1*: The percentage sequence identity of the template structure and the *ALDH1B1* primary sequence. *Identity2*: The percentage sequence identity of the *ALDH1B1* primary sequence and the template structure; *Coverage*: The percentage of residues which align between the template structure and the *ALDH1B1* primary sequence. *Normalized Z-score*: A metric of template alignment accuracy: scores > 1 indicate good alignment.  
(DOC)

**S1 Video. An overview of the *ALDH1B1* structural model and molecular dynamics simulation.** A short video summarizing the key-features of the *ALDH1B1* structural model and molecular dynamics simulation.  
(ZIP)

**S1 File. Sequencing and genotype call data.** The double-anonymised genotype calling data and sequencing data used in the present study. This supplemental file directory contains two sub-directories: 1. "Sequencing", which contains the Sanger sequencing data (trace and sequence call data) for those samples that underwent sequencing verification for the variant rs142427338. 2. "Genotypes", which contains PLINK compatible data (.ped and .map) containing the final genotype calls for the four *ALDH1B1* variants analysed.  
(ZIP)

## Acknowledgments

The research was funded by a PhD studentship award jointly funded by University College London and the Hobson Charity Limited. We thank colleagues associated with the NIHR MHRN for assistance in the identification of cases, obtaining consent and collection of samples at the following NHS trusts: Sandwell Mental Health and Social Care; Northamptonshire Healthcare; Avon and Wiltshire Mental Health Partnership, Sheffield Health and Social Care; Tees Esk and Wear Valleys; Lincolnshire Partnership; Nottinghamshire Healthcare; Central and North West London; South Staffordshire and Shropshire Healthcare; Coventry and Warwickshire; and Dudley and Walsall Mental Health Partnership. Jit Saini, Kush Ruparelia, Radhika Kandaswamy, Alexandra Jarram, Giorgia Quadri, and Niamh O'Brien assisted with the processing of samples and DNA extraction.

## Author Contributions

**Conceptualization:** MJW MYM.

**Data curation:** MJW MYM.

**Formal analysis:** MJW MYM AMcQ.

**Funding acquisition:** MYM.

**Investigation:** MAA MJW.

**Methodology:** MJW MYM.

**Project administration:** MYM AMcQ.

**Resources:** MYM AMcQ.

**Supervision:** MYM AMcQ.

**Visualization:** MJW.

**Writing – original draft:** MYM MJW MAA.

**Writing – review & editing:** MYM MJW MAA AMcQ.

## References

1. Kopun M, Propping P. The kinetics of ethanol absorption and elimination in twins and supplementary repetitive experiments in singleton subjects. *Eur J Clin Pharmacol*. 1977; 11:337–344. PMID: [560303](#)
2. Hurley TD, Edenberg HJ. Genes encoding enzymes involved in ethanol metabolism. *Alcohol Res*. 2012; 34:339–344. PMID: [23134050](#)
3. Yin SJ, Bosron WF, Mages LJ, Li TK. Human liver alcohol dehydrogenase: purification and kinetic characterization of the  $\beta_2\beta_2$ ,  $\beta_2\beta_1$ ,  $\alpha\beta_2$ , and  $\beta_2\gamma_1$  "Oriental" isoenzymes. *Biochemistry*. 1984; 23:5847–5853. PMID: [6395883](#)
4. Goedde HW, Agarwal DP. Polymorphisms of aldehyde dehydrogenase and alcohol sensitivity. *Enzyme*. 1987; 37:29–44. PMID: [3106030](#)
5. Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li TK, et al. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *Am J Hum Genet*. 1991; 48:677–681. PMID: [2014795](#)
6. Chen CC, Lu RB, Chen YC, Wang MF, Chang YC, Li TK, et al. Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *Am J Hum Genet*. 1999; 65:795–807. <https://doi.org/10.1086/302540> PMID: [10441588](#)
7. Li D, Zhao H, Gelernter J. Strong association of the alcohol dehydrogenase 1B gene (*ADH1B*) with alcohol dependence and alcohol-induced medical diseases. *Biol Psychiatry*. 2011; 70:504–512. <https://doi.org/10.1016/j.biopsych.2011.02.024> PMID: [21497796](#)
8. Edenberg HJ. Genes contributing to the development of alcoholism: an overview. *Alcohol Res*. 2012; 34:336–338. PMID: [23134049](#)
9. Hurley TD, Edenberg HJ, Bosron WF. Expression and kinetic characterization of variants of human b1b1 alcohol dehydrogenase containing substitutions at amino acid 47\*. *J Biol Chem*. 1990; 265:16366–16372. PMID: [2398055](#)
10. Goedde HW, Agarwal DP, Fritze G, Meier-Tackmann D, Singh S, Beckmann G, et al. Distribution of *ADH2* and *ALDH2* genotypes in different populations. *Hum Genet*. 1992; 88:344–346. PMID: [1733836](#)
11. Li H, Mukherjee N, Soundararajan U, Tarnok Z, Barta C, Khaliq S, et al. Geographically separate increases in the frequency of the derived *ADH1B*\*47His allele in eastern and western Asia. *Am J Hum Genet*. 2007; 81:842–846. <https://doi.org/10.1086/521201> PMID: [17847010](#)
12. Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, et al. *ADH1B* is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol Psychiatry*. 2012; 17:445–450. <https://doi.org/10.1038/mp.2011.124> PMID: [21968928](#)

13. Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, et al. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry*. 2014; 19:41–49. <https://doi.org/10.1038/mp.2013.145> PMID: 24166409
14. Treutlein J, Frank J, Kiefer F, Rietschel M. *ADH1B* Arg48His allele frequency map: filling in the gap for Central Europe. *Biol Psychiatry*. 2014; 75:e15. <https://doi.org/10.1016/j.biopsych.2013.05.042> PMID: 24011820
15. Way M, McQuillin A, Saini J, Ruparelia K, Lydall GJ, Guerrini I, et al. Genetic variants in or near *ADH1B* and *ADH1C* affect susceptibility to alcohol dependence in a British and Irish population. *Addict Biol*. 2015; 20:594–604. <https://doi.org/10.1111/adb.12141> PMID: 24735490
16. Eng MY, Luczak SE, Wail TL. *ALDH2*, *ADH1B*, and *ADH1C* genotypes in Asians: a literature review. *Alcohol Res Health*. 2007; 30:22–27. PMID: 17718397
17. Chen YC, Peng GS, Tsao TP, Wang MF, Lu RB, Yin SJ. Pharmacokinetic and pharmacodynamic basis for overcoming acetaldehyde-induced adverse reaction in Asian alcoholics, heterozygous for the variant *ALDH2*\*2 gene allele. *Pharmacogenet Genomics*. 2009; 19:588–599. PMID: 19584771
18. Stagos D, Chen Y, Brocker C, Donald E, Jackson BC, Orlicky DJ, et al. Aldehyde dehydrogenase 1B1: molecular cloning and characterization of a novel mitochondrial acetaldehyde-metabolizing enzyme. *Drug Metab Dispos*. 2010; 38:1679–1687. <https://doi.org/10.1124/dmd.110.034678> PMID: 20616185
19. Jackson BC, Holmes RS, Backos DS, Reigan P, Thompson DC, Vasilio V. Comparative genomics, molecular evolution and computational modeling of *ALDH1B1* and *ALDH2*. *Chem Biol Interact*. 2013; 202:11–21. <https://doi.org/10.1016/j.cbi.2012.11.022> PMID: 23247008
20. Sherman D, Davé V, Hsu LC, Peters TJ, Yoshida A. Diverse polymorphism within a short coding region of the human aldehyde dehydrogenase-5 (*ALDH5*) gene. *Hum Genet*. 1993; 92:477–480. PMID: 8244338
21. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. <https://doi.org/10.1038/nature11632> PMID: 23128226
22. Way MJ. Computational modelling of *ALDH1B1* tetramer formation and the effect of coding variants. *Chem Biol Interact*. 2014; 207:23. <https://doi.org/10.1016/j.cbi.2013.10.019> PMID: 24211684
23. Husemoen LL, Fenger M, Friedrich N, Tolstrup JS, Beenfeldt Fredriksen S, Linneberg A. The association of *ADH* and *ALDH* gene variants with alcohol drinking habits and cardiovascular disease risk factors. *Alcohol Clin Exp Res*. 2008; 32:1984–1991. <https://doi.org/10.1111/j.1530-0277.2008.00780.x> PMID: 18782342
24. Husemoen LL, Jørgensen T, Borch-Johnsen K, Hansen T, Pedersen O, Linneberg A. The association of alcohol and alcohol metabolizing gene variants with diabetes and coronary heart disease risk factors in a white population. *PloS ONE*. 2010; 5:e11735. <https://doi.org/10.1371/journal.pone.0011735> PMID: 20700531
25. Linneberg A, Gonzalez-Quintela A, Vidal C, Jørgensen T, Fenger M, Hansen T, et al. Genetic determinants of both ethanol and acetaldehyde metabolism influence alcohol hypersensitivity and drinking behaviour among Scandinavians. *Clin Exp Allergy*. 2010; 40:123–130. <https://doi.org/10.1111/j.1365-2222.2009.03398.x> PMID: 20205700
26. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7:248–249. <https://doi.org/10.1038/nmeth0410-248> PMID: 20354512
27. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008; 9:40. <https://doi.org/10.1186/1471-2105-9-40> PMID: 18215316
28. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010; 5:725–738. <https://doi.org/10.1038/nprot.2010.5> PMID: 20360767
29. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015; 12:7–8. <https://doi.org/10.1038/nmeth.3213> PMID: 25549265
30. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins*. 2004; 57:702–710. <https://doi.org/10.1002/prot.20264> PMID: 15476259
31. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25:1605–1612. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
32. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013; 29:845–854. <https://doi.org/10.1093/bioinformatics/btt055> PMID: 23407358

33. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1:19–25.
34. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B*. 2001; 105:6474–6487.
35. Darden T, York D, Pedersen L. Particle mesh Ewald: an  $N \log(N)$  method for Ewald sums in large systems. *J Chem Phys*. 1993; 12:10089–10092.
36. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996; 14:33–8, 27–28. PMID: [8744570](#)
37. American Psychiatric Association. Diagnostic and statistical manual of mental disorders 4th ed. Washington DC: American Psychiatric Press; 1994.
38. World Health Organisation. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organisation; 1992.
39. Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry*. 1978; 35:837–844. PMID: [678037](#)
40. Sambrook J, Russell DW. *Molecular Cloning: a Laboratory Manual*. Fourth edition ed. Cold Spring Harbor, New York: Cold Spring Harbour Laboratory Press; 2001.
41. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc*. 2008; 3:1452–1456. <https://doi.org/10.1038/nprot.2008.133> PMID: [18772872](#)
42. Bonfield JK, Smith KF, Staden R. A new DNA sequence assembly program. *Nucleic Acids Res*. 1995; 23:4992–4999. PMID: [8559656](#)
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. <https://doi.org/10.1086/519795> PMID: [17701901](#)
44. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015; 4:7. <https://doi.org/10.1186/s13742-015-0047-8> PMID: [25722852](#)
45. Xin YN, Zhao Y, Lin ZH, Jiang X, Xuan SY, Huang J. Molecular dynamics simulation of *PNPLA3* I148M polymorphism reveals reduced substrate access to the catalytic cavity. *Proteins*. 2013; 81:406–414. <https://doi.org/10.1002/prot.24199> PMID: [23042597](#)
46. Liu M, Wang L, Sun X, Zhao X. Investigating the impact of Asp181 point mutations on interactions between PTP1B and phosphotyrosine substrate. *Sci Rep*. 2014; 28:5095.
47. Karlsake C, Piotto ME, Pak YK, Weiner H, Gorenstein DG. 2D NMR and structural model for a mitochondrial signal peptide bound to a micelle. *Biochemistry*. 1990; 29:9872–9878. PMID: [2271626](#)
48. Ni L, Zhou J, Hurley TD, Weiner H. Human liver mitochondrial aldehyde dehydrogenase: three-dimensional structure and the restoration of solubility and activity of chimeric forms. *Protein Sci*. 1999; 8:2784–2790. <https://doi.org/10.1110/ps.8.12.2784> PMID: [10631996](#)
49. Perez-Miller SJ, Hurley TD. Coenzyme isomerization is integral to catalysis in aldehyde dehydrogenase. *Biochemistry*. 2003; 42:7100–7109. <https://doi.org/10.1021/bi034182w> PMID: [12795606](#)
50. Smith LJ, Mark AE, Dobson CM, van Gunsteren WF. Comparison of MD simulations and NMR experiments for hen lysozyme. Analysis of local fluctuations, cooperative motions, and global changes. *Biochemistry*. 1995; 34:10918–10931. PMID: [7662673](#)
51. Brocke KS, Neu-Yilik G, Gehring NH, Hentze MW, Kulozik AE. The human intronless melanocortin 4-receptor gene is NMD insensitive. *Hum Mol Genet*. 2002; 11:331–335. PMID: [11823452](#)
52. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006; 38:209–213. <https://doi.org/10.1038/ng1706> PMID: [16415888](#)
53. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet*. 2014; 15:335–346. <https://doi.org/10.1038/nrg3706> PMID: [24739678](#)
54. Wu YY, Delgado R, Costello R, Sunderland T, Dukoff R, Csako G. Quantitative assessment of apolipoprotein E genotypes by image analysis of PCR-RFLP fragments. *Clin Chim Acta*. 2000; 293:213–221. PMID: [10699435](#)
55. Wang LX, Wen S, Wang CC, Zhou B, Li H. Molecular adaption of alcohol metabolism to agriculture in East Asia. *Quat Int*. 2016; 426:187–194.