Alcohol dehydrogenases, aldehyde dehydrogenases and alcohol use disorders: a critical review

Howard J. Edenberg*[12] and Jeanette N. McClintick[1]

Supplementary Information

Supplementary Figures:

Supplementary Figure 1. Location of eQTLs for *ADH1B* in adipose tissues. Positions of significant eQTLs for *ADH1B* in adipose tissue, from GTEx version 7 (GTEx_Consortium, 2013). Top line: Subcutaneous adipose, with position on human genome build GRCh37/hg19; next, Visceral adipose; bottom, Gene map.

Supplementary Figure 2. Linkage Disequilibrium (D') among ADH SNPs with $p<10^{-6}$. D' among SNPs in the ADH region from Supplementary Table 1 (3 SNPs at a distance were omitted to allow good visualization) was calculated and plotted using LDlink for 3 major population groups: A. European populations. B. East Asian populations. C. African populations. The locations of ADH genes are shown below, along with position of the SNPs along chromosome 4. Rs1229984 and rs2066702 are boxed.

Supplementary Figure 3. Linkage Disequilibrium (D') among ALDH SNPs with $p<5x10^{-8}$. D' among SNPs in the ALDH region from Supplementary Table 2 was calculated and plotted using LDlink for East Asian populations. Rs671 is boxed.

Supplementary Tables:

Supplementary Table 1. ADH SNPs associated with alcohol dependence or related traits. SNPs with $p<10^{-6}$ are shown in order of their position on chromosome 4 (GRCh38/hg38). Traits: AD = alcohol dependence, Consumption = Ln(drinks per week), Maxdrinks = largest number of drinks in 24 h, AUDIT total score, AUDIT-C = questions 1-3 (consumption), AUDIT-P =

questions 4-10 (problems), Drinker/non drinker status. Population: EUR = European, 23 = 23andMe EUR, UKB = UK Biobank EUR, AA = African American.

Supplementary Table 2. ALDH SNPs associated with alcohol dependence and related traits. SNPs with $p<5x10^{-8}$ are shown in order of their position on chromosome 12 (GRCh38/hg38). Traits: AD = alcohol dependence, Flushing, Consumption = Ln(drinks per week), Maxdrinks = largest number of drinks in 24 h, Drinker vs. Non-drinker.

Supplementary note. Analyses reported here have made use of many bioinformatic resources:

- The 1000 Genomes Project (The_1000_Genomes_Project_Consortium et al., 2015) data are at http://www.internationalgenome.org/home.
- The Genotype-Tissue Expression (GTEx) Project (GTEx_Consortium, 2013) was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.
- The Genome Aggregation Database (gnomAD) (Lek et al., 2016) data are at http://gnomad.broadinstitute.org/.
- LDlink is provided by the National Cancer Institute (Machiela and Chanock, 2018).
- ALFRED (Rajeevan et al., 2012) is at http://alfred.med.yale.edu.
- rAggr is at http://raggr.usc.edu.

Other coding alleles present at ≥1% in at least one population. Data from the Genome Aggregation Database; gnomAD version 2, *http://gnomad.broadinstitute.org* (Lek et al., 2016). Most have not been studied at the protein level.

ADH1A: Coding variations are essentially non-existent, with none having an allele frequency of 1% or above in any population studied.

ADH1B: There are only 2 other missense variations with allele frequencies over 1% in at least 1 population: rs113075608 (Trp16Arg in Africans) and rs41275699 (Ile65Thr in Finns), and these have not been studied at the protein level.

ADH1C: There are less common variants in *ADH1C* that are rarely studied, including Pro352Thr (rs35719513), found at low frequency in Native American populations (MAF = 0.039) and nearly never elsewhere (≤0.005) (Osier et al., 2002), and a stop codon at position 79 (rs283413) not found in many populations (overall frequency in the 1000 genomes populations = 0.007). There is only 1 other missense variation with allele frequencies over 1% in at least 1 population: rs35385902 (Arg48His in Africans) would be expected to affect kinetics as much as the same change in *ADH1B* does, but it has not been studied at the protein level.

ADH4: There are few coding variants in *ADH4*, one of which (Ile309Val, rs1126671) affects the stability of the enzyme and its binding of ethanol (Stromberg et al., 2002); it is relatively common in Europeans (MAF = 0.30) and Africans (MAF=0.14) but rare in East Asians (MAF ~ 0.001). There are 2 other missense variations with allele frequencies over 1% in several populations: rs1126673 (Val374Ile) and rs13125262 (Phe8Leu), that have not been studied at the protein level, plus an early terminator, rs3919370 (Tyr22Ter), that should abolish activity.

ADH5*:* There is only one known coding variation in *ADH5* with an allele frequency over 1%, rs28730623 (Leu163Ser), which has not been studied at the protein level.

ADH7: There are 3 missense variants with MAF above 1%: rs1573496 (Gly100Ala, unlikely to affect kinetics), rs59534319 (Lys246Glu, in Africans), and rs113993320 (Arg58His, in Africans), but they have not been studied at the enzymatic level.

ALDH1A1: Only 1 missense variant has MAF above 1%: rs8187929 (Ile177Phe) in East Asia, and it is unlikely to affect kinetics.

ALDH1B1: rs2073478 (Arg107Leu) is at high frequency in several populations, ranging from 0.297 to 0.495. rs2228093 (Ala86Val) ranges in MAF from 0.093 to 0.380. rs4878199 (Val253Met) ranges up to 0.146. rs113083991 (Val176Ile) ranges up to 0.90. There are also 2 frameshifts: rs538496304 (Gly193ValfsTer16) and rs201408956 (Gly388GlufsTer23) that are present at over 1% in at least one population.

ALDH2: rs201582342 (Pro92Thr) reaches 0.257 and rs201108880 (Val304Met) reaches 0.022.

Supplementary References

GTEx_Consortium (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580-585.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285-291.

Machiela MJ, Chanock SJ (2018) LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. Bioinformatics 34:887-889.

Osier MV, Pakstis AJ, Goldman D, Edenberg HJ, Kidd JR, Kidd KK (2002) A proline-threonine substitution in codon 351 of ADH1C is common in Native Americans. Alcohol Clin Exp Res 26:1759-1763.

Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK (2012) ALFRED: an allele frequency resource for research and teaching. Nucleic Acids Research 40:D1010-D1015.

Stromberg P, Svensson S, Hedberg JJ, Nordling E, Hoog JO (2002) Identification and characterisation of two allelic forms of human alcohol dehydrogenase 2. Cell Mol Life Sci 59:552-559.

The_1000_Genomes_Project_Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. Nature 526:68-74.