Open access • Posted Content • DOI:10.1101/2021.06.02.21258094

Genotype-stratified GWAS meta-analysis reveals novel loci associated with alcohol consumption — Source link

Koyanagi Yn, Masahiro Nakatochi, Hidemi Ito, Yumiko Kasugai ...+37 more authors

Institutions: Nagoya University, Tohoku University, Kyoto University, Kyushu University ...+10 more institutions

Published on: 05 Jun 2021 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: ADH1B, Genome-wide association study and Genetic architecture

Genotype-stratified GWAS meta-analysis reveals novel loci associated with alcohol consumption

Yuriko N. Koyanagi[1#], Masahiro Nakatochi[2#*], Hidemi Ito[1,3], Yumiko Kasugai[4,5], Akira Narita[6], Takahisa Kawaguchi[7], Hiroaki Ikezaki[8,9], Asahi Hishida[10], Megumi Hara[11], Toshiro Takezaki[12], Teruhide Koyama[13], Haruo Mikami[14], Sadao Suzuki[15], Sakurako Katsuura-Kamano[16], Kiyonori Kuriki[17], Yasuyuki Nakamura[18], Kenji Takeuchi[10], Atsushi Hozawa[19], Kengo Kinoshita[6], Yoichi Sutoh[20], Kozo Tanno[21,22], Atsushi Shimizu[20,23], Isao Oze[4], Yukino Kawakatsu[4], Yukari Taniyama[1], Issei Imoto[24], Yasuharu Tabara[7], Meiko Takahashi[7], Kazuya Setoh[7], Shiori Suzuki[25], Atsushi Goto[25,26], Ryoko Katagiri[25], Taiki Yamaji[25], Norie Sawada[25], Shoichiro Tsugane[25,27], Kenji Wakai[10], Masayuki Yamamoto[6], Makoto Sasaki[28,29], Fumihiko Matsuda[7], Motoki Iwasaki[25], Paul Brennan[30], Keitaro Matsuo[4,5*]

1. Division of Cancer Information and Control, Department of Preventive Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan.
2. Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan.
3. Department of Descriptive Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan.
4. Division of Cancer Epidemiology and Prevention, Department of Preventive Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan.
5. Department of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan.
6. Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan.
7. Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.
8. Department of General Internal Medicine, Kyushu University Hospital, Fukuoka, Japan.
9. Department of Comprehensive General Internal Medicine, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan.
10. Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan.
11. Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, Japan.
12. Department of International Island and Community Medicine, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan.
13. Department of Epidemiology for Community Health and Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan.
14. Cancer Prevention Center, Chiba Cancer Center Research Institute, Chiba, Japan.
15. Department of Public Health, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.
16. Department of Preventive Medicine, Tokushima University Graduate School of Biomedical Sciences, Tokushima, Japan.

17. Laboratory of Public Health, Division of Nutritional Sciences, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan.
18. Department of Public Health, Shiga University of Medical Science, Otsu, Japan.
19. Department of Preventive Medicine and Epidemiology, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan.
20. Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan.
21. Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University, Iwate, Japan.
22. Division of Clinical Research and Epidemiology, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan.
23. Division of Biomedical Information Analysis, Institute for Biomedical Sciences, Iwate Medical University, Iwate, Japan.
24. Aichi Cancer Center Research Institute, Nagoya, Japan.
25. Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan.
26. Department of Health Data Science, Graduate School of Data Science, Yokohama City University, Yokohama, Japan.
27. National Institute of Health and Nutrition, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan.
28. Division of Ultrahigh Field MRI, Institute for Biomedical Sciences, Iwate Medical University, Iwate, Japan.
29. Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan.
30. Genomic Epidemiology Branch, International Agency for Research on Cancer, Lyon, France.

[#]These authors contributed equally to this work.

***Correspondence to: Masahiro Nakatochi and Keitaro Matsuo**

Masahiro Nakatochi, PhD, Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, 1-1-20 Daiko-Minami, Higashi-ku, Nagoya, Aichi, 461-8673, Japan
Phone/Fax: +81-52-719-1923, E-mail: mnakatochi@met.nagoya-u.ac.jp

Keitaro Matsuo, MD, PhD, MSc, Division of Cancer Epidemiology and Prevention, Department of Preventive Medicine, Aichi Cancer Center Research Institute, 1-1 Kanokoden Chikusa-ku, Nagoya, Aichi, 464-8681, Japan
Phone: +81-52-764-2982, Fax: +81-52-763-5233, E-mail: kmatsuo@aichi-cc.jp

Keywords: Genotype-stratified GWAS, *ALDH2* rs671, alcohol consumption, SNP-SNP interaction, gene-gene interaction

77 **Conflicts of interest:** The authors declare no potential conflicts of interest.

78

79 **Word count:** 3,390 words (Introduction through Discussion)

80

81 **Note:** This manuscript includes 5 figures, 8 supplementary tables, and 6 supplementary figures.

82

Abstract

An East Asian-specific variant on *aldehyde dehydrogenase 2* (*ALDH2* rs671, G>A) is the major genetic determinant of alcohol consumption. We performed an rs671 genotype-stratified genome-wide association study meta-analysis in up to 40,679 individuals from Japanese populations to uncover additional loci associated with alcohol consumption in an rs671-dependent manner. No loci satisfied the genome-wide significance threshold in wild-type homozygotes (GG), but six loci (*ADH1B*, *ALDH1B1*, *ALDH1A1*, *ALDH2*, *GOT2*, and *MYOM1- MYL12A*) did so in heterozygotes (GA). Of these, three loci (*ALDH2*, *GOT2*, and *MYOM1- MYL12A*) were novel, and two (*ADH1B* and *ALDH1B1*) showed genome-wide significant interaction with rs671. Our results identify a new genetic architecture associated with alcohol consumption, and shed additional light on the genetic characteristics of alcohol consumption among East Asians.

Alcohol consumption is a major contributor to mortality and influences risk for various human diseases and disorders[1]. Even moderate consumption may have a substantial impact on mortality[2]. Indeed, the latest Global Burden of Disease study on alcohol use states that the level of consumption should be reduced to zero to minimize health risk[1]. Alcohol consumption has been considered a heritable trait[3,4]. The number of genetic studies on alcohol consumption is increasing[5-14], and the genetic variants that are consistently reproducible are those of genes encoding alcohol-metabolizing enzymes[5,7,8,10-14]. Ingested alcohol is predominantly metabolized to acetaldehyde through alcohol dehydrogenase (ADH) enzymes, and aldehyde dehydrogenase (ALDH) enzymes further catalyze the oxidation of acetaldehyde to acetate[15]. Notably, rs671 (c.1510G>A [p.Glu504Lys]), a functional single nucleotide polymorphism (SNP) in the *ALDH2* gene which is highly prevalent in East Asians[16], is a strong and well-known determinant of alcohol consumption. Every previous genome-wide association study (GWAS) in East Asians[5,7,8,11,14] identified the strongest signals in the rs671 variant (or variants in strong linkage disequilibrium [LD] with rs671), ranging from $P < 1.0 \times 10^{-58}$ ($n = 2,834$)[5] to $P < 1.0 \times 10^{-4,740}$ ($n = 165,084$)[14].

Among ALDH isoforms, ALDH2 has by far the highest affinity for acetaldehyde (Km <1 μM) and is primarily responsible for its oxidation[16,17]. Because the *ALDH2* rs671variant inactivates ALDH2 enzymatic activity, individuals who are heterozygous (GA) or homozygous (AA) for this variant experience a rapid accumulation of blood acetaldehyde after alcohol ingestion[16]. This variant thereby increases exposure to the unpleasant effects of acetaldehyde (e.g. flushing, headache,

palpitation, and nausea), which in consequence significantly reduces alcohol consumption and thereby confers a protective effect against alcohol-induced carcinogenesis[18]. Conversely, heterozygotes (GA) who drink alcohol experience increased susceptibility to carcinogenesis, in particular for head and neck and esophageal cancers, due to higher concentrations of acetaldehyde, one of the most likely carcinogens in alcohol[15]. With regard to variant homozygotes (AA), however, these have rarely evidenced an increased cancer risk associated with alcohol, because they are unable to oxidize acetaldehyde, a characteristic which is highly correlated with nondrinking[19,20]. In contrast, heterozygotes having 16-18% of normal enzyme activity[21,22] show a broader range of alcohol consumption. Overall, the highest risk group for alcohol-related cancers are heterozygotes[23,24], and alcohol consumption among genotypes of rs671 shows distinct genetic heterogeneity.

Here, to further decipher the genetic architecture of alcohol consumption in consideration of the status of this unique variant of rs671, we conducted a meta-analysis of rs671 genotype-stratified GWASs comprising up to 40,679 Japanese individuals. Using rs671 genotype-stratified analyses, we tested the hypothesis that variants associated with alcohol consumption exhibit rs671 genotype-dependent associations, and sought novel variants conferred by genetic interaction of the rs671 genotype. We considered that this approach might help uncover loci with differential influence on alcohol consumption among genotypes, and enable the detection of loci whose effects were indistinct in previous GWASs.

Results

Characteristics of study participants

A total of 40,679 individuals were included in this GWAS meta-analysis of five Japanese cohorts, namely the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC)[25] Study (n = 4,958), the Japan Multi-Institutional Collaborative Cohort (J-MICC)[26,27] Study (n = 13,236), the Japan Public Health Center-based Prospective (JPHC)[28] Study (n = 10,037), the Tohoku Medical Megabank Community-Based Cohort (TMM)[29] Study (n = 7,857), and the Nagahama Prospective Cohort for Comprehensive Human Bioscience (Nagahama) Study[30] (n = 4,591), after imputation and quality control of individual subject genotype data (Supplementary Information and Supplementary Tables 1 and 2). Median self-reported daily alcohol intake, mean age of participants, number of ever/never drinkers, and proportion of male participants were obtained from the studies and are shown in Supplementary Table 1. The number of participants included in each analysis for daily alcohol intake and drinking status was as follows: rs671 wild-type homozygotes (GG)-only analysis, n = 23,398 and 24,514; heterozygotes (GA)-only analysis, n = 13,385 and 13,848; unstratified analysis, n = 39,077 and 40,679; and interaction analysis, n = 36,783 and 38,362, respectively (Supplementary Table 3). As the number of variant homozygotes (AA) was small (n = 2,294 for daily alcohol intake; n = 2,317 for drinking status) and included only 101 subjects in the ever drinking group (Supplementary Table 3), association analysis in variant homozygotes only was not conducted, and these subjects were excluded from the interaction analyses. Quantile-quantile

plots revealed no evidence of genomic inflation (Supplementary Figures 1 and 2) and the genomic inflation factors ranged from 0.99 to 1.02 for meta-analyses.

ALDH2 rs671 genotype-stratified, unstratified, and interaction GWAS meta-analyses

Major results of the genotype-stratified GWAS meta-analysis are summarized in Figures 1-4. In wild-type homozygotes (GG), no genome-wide significant loci were detected for either daily alcohol intake or drinking status (Figures 1a and 2a). In heterozygotes (GA), on the other hand, six and four loci satisfied the genome-wide significance ($P < 5.0 \times 10^{-8}$) for daily alcohol intake and drinking status, respectively (Figures 1b, 2b, 3 and 4). These included three loci that were previously implicated in alcohol consumption, namely chromosome 4q23[10-14], *ALDH1B1*[14], and *ALDH1A1*[14]. Of the three remaining loci, two loci (*GOT2* at 16q21, and *MYOM1- MYL12A* at 18p11.31) have not been reported in previous GWASs of alcohol consumption, and one locus (chromosome 12q24.12) is the same locus as rs671. Regional association plots for these novel loci are shown in Figure 5. The lead SNPs are rs56884502 in chromosome 12q24.12 (*ALDH2*), rs73550818 in *GOT2*, and rs572435541 in *MYOM1-MYL12A* for daily alcohol intake, and rs7978737 in chromosome 12q24.12 (*ACAD10-ALDH2*) for drinking status. The unstratified GWAS showed three hits (rs1260326 in *GCKR* for daily alcohol intake; rs1229984 in *ADH1B* and rs671 in *ALDH2* for daily alcohol intake and drinking status) (Figures 1c, 2c, 3 and 4), all of which have been previously reported[5,7-14].

Regional association plots for all identified regions other than those in Figure 5 are shown in Supplementary Figures 3 and 4.

A further analysis evaluating variant-rs671 interaction detected two loci (chromosome 4q23 and *ALDH1B1*) showing genome-wide significant interaction with rs671 (Figures 1d, 2d, 3 and 4 and Supplementary Figures 3c and 4c). This is the first identification of the interactive effects of these two loci on rs671, although that of rs1229984 in *ADH1B* at 4q23 on rs671 was suggested from a meta-analysis of studies using candidate gene-based approaches[31]. Among the other loci reaching genome-wide significance in either rs671-stratified or unstratified analysis, three loci (*ALDH1A1*, chromosome 12q24.12, and *GOT2*) demonstrated interaction with rs671 with a suggestive significance level ($P < 5.0 \times 10^{-6}$) for daily alcohol intake and/or drinking status (Figures 3 and 4). Regarding the multiple hits on chromosome 4q23 observed through these analyses, rs1813977 (interaction), rs35333426 (for drinking status among heterozygotes), and rs10005290 (for daily alcohol intake among heterozygotes) were in strong LD with a functional SNP of *ADH1B* (rs1229984) (all r^2 and *D'* values are 0.71 and 1.00 in 1000 Genome Project phase 3-Japanese in Tokyo (1KGP- JPT), respectively). We further applied a random effects model[32] given that estimates in *ALDH1B1* rs2228093 showed between-study heterogeneity (*P* values from test of heterogeneity <0.05) (Figures 3 and 4), but the results did not change substantially ($P = 8.34 \times 10^{-11}$ in the interaction analysis for daily alcohol intake; and $P = 4.38 \times 10^{-12}$ and $P = 1.62 \times 10^{-11}$ in the analyses of heterozygotes only and interaction, respectively, for drinking status).

Supplementary Table 4 shows functional annotation results and allele frequencies across different ancestries for the lead SNPs. Five SNPs, namely rs1260326 on *GCKR*, rs1229984 on *ADH1B*, rs2228093 on *ALDH1B1*, rs818787929 on *ALDH1A1*, and rs671 on *ALDH2*, are non-synonymous. According to the 1KGP database, three SNPs (rs8187929, rs671, and rs572435541) are polymorphic, with a minor allele frequency (MAF) of 0.046, 0.170, and 0.009 in the East Asian (EAS) population, respectively. In contrast, they are monomorphic in the European (EUR) population. The *ADH1B* rs1229884 C allele is major in the EUR population (AF = 0.970) but minor in the EAS population (AF = 0.300).

Effect of a novel SNP within the same locus as rs671

Figures 3 and 4 show the direction of effects of the identified variants other than rs671 under each analysis. Notably, with regard to the novel SNP in chromosome 12q24.12 (*ALDH2* rs56884502, T>A), the A allele of rs56884502, which was associated with decreasing daily alcohol intake in the rs671 heterozygotes (β = -0.217), showed the opposite direction of effect in the unstratified analysis (β = 0.274) (Figure 3). This apparently conflicting result was due to strong LD between rs56884502 and rs671. The 1KGP-JPT (n = 104) and our own direct genotyped data from subjects in the HERPACC Study (n = 96) indicated that there were only three rs56884502-rs671 haplotypes, namely T-G, A-G and T-A (Supplementary Table 5). The respective LD coefficients of r^2 and D' were <0.1 and 1.0 (Supplementary Table 5). Results from LD analysis based on the 1KGP JPT data

208 (n = 104) of rs671 and the 73 SNPs at 12q24.12 which showed genome-wide significance for daily
209 alcohol intake among the heterozygotes (Supplementary Table 6) are shown in Supplementary
210 Figure 5. The pairwise *D'* figure showed that all 73 SNPs were in complete LD in terms of *D'*
211 (Supplementary Figure 5). Further, the three indicated haplotypes of rs671 and these SNPs could
212 explain >99% (Supplementary Figure 6). Therefore, when evaluated without stratification,
213 rs56884502 A allele, which formed a haplotype with rs671 G allele only, was associated with
214 increasing drinking intensity, by reflecting the effect of rs671 G allele. However, when stratified,
215 rs56884502 A allele turned out to have the opposite direction of effect–decreasing drinking intensity.
216 The other lead SNP in chromosome 12q24.12 for drinking status (rs7978737) was in LD with
217 rs56884502 (r^2 = 0.97 in 1KGP JPT), and accordingly showed the same phenomenon (Figure 4).
218
219 **Associations of previously reported loci**
220 Among the previously reported loci in the EUR population[9,11,12,33,34] other than those replicated with
221 a genome-wide significance level in this study, we observed nominal evidence of association (P <
222 0.05) for nine loci in the unstratified analysis, six loci in wild-type homozygotes, and seven loci in
223 heterozygotes (Supplementary Table 7).
224
225 **eQTL analysis of novel SNPs**

Of the detected novel variants, rs56884502 and rs7978737 in chromosome 12q24.12 and rs73550818 in *GOT2* were found to be eQTL using the GTEx database (Supplementary Table 8). rs56884502 A allele and rs7978737 T allele are associated with decreased expression of *ALDH2* in multiple tissues. rs73550818 A allele is associated with increased expression of *GOT2* in liver ($P = 1.0\times10^{-8}$).

Discussion

We report here the results of an rs671 (G>A) genotype-stratified GWAS meta-analysis of alcohol consumption with a total of 40,679 participants from five Japanese cohorts. While three loci (*GCKR*, *ADH1B*, and *ALDH2*) were identified in the unstratified analysis, the rs671 genotype-stratified GWAS identified no loci in wild-type homozygotes (GG) and six loci (*ADH1B*, *ALDH1B1*, *ALDH1A1*, *ALDH2*, *GOT2*, and *MYOM1- MYL12A*) in heterozygotes (GA). Of these, three loci (*ALDH2* at 12q24.12, *GOT2* at 16q21, and *MYOM1- MYL12A* at 18p11.31) are novel in the context of alcohol consumption. Further, the interaction GWAS identified for the first time two loci (*ADH1B* and *ALDH1B1*) showing genome-wide significant interaction with rs671.

The failure of other loci to reach a genome-wide significant level in rs671 wild-type homozygotes indicates that the rs671 GG genotype itself is strong enough to make a significant contribution to determining the alcohol consumption phenotype in this population. This is consistent with the observation that this phenotype has been resistant to gene discovery efforts in non-Asian populations, where rs671 is often monomorphic[35]. Further, the strongest signal for daily alcohol intake in rs671 heterozygotes was observed in *ADH1B* ($P < 5.0 \times 10^{-26}$), followed by *ALDH1B1* ($P = 7.2 \times 10^{-14}$) and then *ALDH1A1* ($P = 1.2 \times 10^{-10}$), all of which are associated with the concentration of acetaldehyde. ADH1B is the predominant isoform involved in alcohol oxidation, whereas ALDH1B1 and ALDH1A1 are the ALDH isoforms involved in acetaldehyde oxidation, with the second (Km 30 μM) and third (Km 50–180 μM) highest affinities for acetaldehyde, respectively[17].

249 The nonsynonymous lead SNP of rs1229984 (T>C [p.His48Arg]) found in the *ADH1B* coding region
250 is associated with slow alcohol metabolism, leading to the slow accumulation of acetaldehyde and
251 consequently greater alcohol consumption[19,20]. Although rs2228093 (C>T [p.Ala86Val]) in
252 *ALDH1B1* and rs8187929 (T>A [p.Ile177Phe]) in *ALDH1A1* were first shown to be associated with
253 drinking status in a previous Japanese GWAS[14], their effects on enzyme activity are not fully
254 understood. However, rs2228093 in *ALDH1B1* was also shown to possibly influence alcohol
255 consumption in European populations using a candidate gene approach[36,37]. In addition, the
256 protective effect of rs2228093 T allele against alcohol consumption observed in this study is
257 consistent with the results of previous studies using bioinformatic analyses, which predicted
258 disruption of the structural flexibility of the protein product[38] and catalytic inactivity[39] of ALDH1B1
259 in the presence of the rs2228093 T allele. Our present study genetically confirmed that, at least in
260 this population, alcohol consumption level is largely determined by the concentration of
261 acetaldehyde, because no significant signal was detected in rs671wild-type homozygotes whereas
262 signals in the genes encoding the second and third enzymes involved in the concentration of
263 acetaldehyde were identified in heterozygotes. Elucidating the functional contributions of rs2228093
264 in *ALDH1B1* and rs8187929 in *ALDH1A1* to alcohol/aldehyde metabolism requires further
265 investigation.

266 Using a stratified method based on rs671 genotype, we were able to uncover the effect of a
267 novel SNP at the same locus as rs671. ALDH2 is a tetramer which is regarded as a dimer of dimers.

The rs671 A allele is predicted to disrupt the structure of not only its own subunit but also its dimer partner, reducing the stability of the tetramic structure of ALDH2 and resulting in a dramatic reduction in enzyme activity[40]. This East Asian-specific SNP is considered to be a relatively young polymorphism[41] and to have been under strong recent natural selection pressure in the Japanese population[42]. On the other hand, the novel SNP at 12q24.12 of rs56884502 is a globally common SNP located <40Kb distant to rs671 (Supplementary Table 4). Given this evidence and the two LD measures of r^2 <0.1 and D' = 1.0 for rs56884502 and rs671, we speculate that rs56884502 arose prior to rs671, and that rs671 then arose on a different branch from rs56884502 in the rs56884502-rs671 T-G haplotype without subsequent historic recombination, finally resulting in the three haplotypes of T-G, A-G, and T-A. The protective effect of the rs56884502 A allele against alcohol consumption observed in rs671 heterozygotes can therefore be regarded as a protective effect of the rs56884502-rs671 T-A/A-G diplotype. Accordingly, we hypothesize that rs56884502 (or variants in LD) is associated with reduced enzyme activity via an effect on the rs671 A allele located on the opposite haplotype. However, rs56884502 is located in the intron of *ALDH2* (Supplementary Table 4) and the rs56884502 A allele was found to be associated with decreased expression of *ALDH2* (Supplementary Table 8). Further, none of the other 72 SNPs that were in LD with rs56884502 and showed genome-wide significance for daily alcohol intake are located within the coding region (Supplementary Table 6), suggesting that while these SNPs may be potentially associated with expression, they may have no direct effect on the protein structure or tetramer formation of ALDH2

by interacting with the rs671 A allele on the opposite haplotype. Further, this potential effect on *ALDH2* expression is inconsistent with the lack of protective effect of the rs56884502 A allele in a population with rs671 wild-type homozygosity. One possible explanation is that this protective effect may be too small for detection in wild-type homozygotes, but if the causative exonic SNP may be hidden and the variant is rare, this would be difficult or impossible to impute using the 1KGP reference panel. Further elucidation of this rs671-dependent protective effect of rs56884502 will require deep whole-genome sequencing-based analysis and/or experimental study.

Our genome-wide analysis indicates the interactive effect of rs2228093 in *ALDH1B1* with rs671. ALDH1B1 is another mitochondrial ALDH which shares 75% similarity with ALDH2 at the amino acid sequence level, and is predicted to form a homotetramer, similarly to ALDH2[43]. Although ALDH1B1 is also able to oxidize acetaldehyde, individuals with the rs671 A allele are reported not to exhibit a compensatory increase in ALDH1B1 activity[44]. Further, a bioinformatics analysis predicted protein-protein interactions between ALDH2 and ALDH1B1, indicating that ALDH2 and ALDH1B1 subunits are likely to form heterotetramers[44]. These findings suggest the hypothesis that the rs671 A allele reduces the catalytic activity of ALDH1B1. They also explain the present finding of gene-gene interaction between the rs671 A allele and rs2228093 T allele, both of which are validated and predicted[39] to be associated with catalytic inactivity. Moreover, these findings may further explain the limited and conflicting genetic evidence for rs2228093 on drinking

in European populations[36,37], in which rs2228093 is polymorphic (MAF = 0.15 in 1KGP EUR) but rs671 is monomorphic.

The remaining newly identified loci associated with alcohol consumption in this study are *GOT2* and *MYOM1-MYL12A*. The lead SNPs in these loci are located within the non-coding region and the functional effects of these variants are unknown. However, the lead SNP in *GOT2* (rs73550818, C>A) is located in the intron of *GOT2* and showed a protective effect against alcohol consumption in heterozygotes; this may be a suitable target for future study, given that the rs73550818 A allele was shown to be significantly associated with increased levels of aspartate aminotransferase (AST), a biochemical marker for liver injury, in a previous GWAS of 134,154 Japanese individuals[45]. *GOT2* encodes the mitochondrial isoform of glutamic-oxaloacetic transaminase; this plays an important role in many processes, including amino acid metabolism, long-chain fatty acid uptake, and the urea and tricarboxylic acid cycles[46]. An *in vivo* study suggested that increased mitochondrial AST among alcoholics is a consequence of the pharmacologic upregulation of *GOT2* gene expression by ethanol, which further mediates fatty acid uptake, resulting in alcoholic fatty liver[46]. Considering that the rs73550818 A allele is associated with increased expression of *GOT2* in the liver (Supplementary Table 8), this allele might be associated with ethanol-induced liver injury. On the other hand, previous studies of rs671 showed significantly lower AST in heterozygotes than in wild-type homozygotes among drinkers[47], even after adjustment for alcohol intake[7]. An observational study of patients with alcoholic liver injury[48] and a study of

Aldh2 knockout mice[49] suggested a protective effect of the rs671 A allele on ethanol-induced liver injury. These findings suggest the opposite effects of the rs73550818 A and rs671 A alleles on ethanol-induced liver injury. The mechanism of the suggested interaction between rs73550818 and rs671 observed in our present study therefore warrants further investigation.

This study has several strengths. First, most of the included cohorts were population-based and included a large number of general Japanese individuals, and the possibility of selection bias is likely small. Second, the study involved a single ethnic group with a similar religious and cultural background, making it unlikely that these factors would bias the phenotype of alcohol consumption[50]. An important limitation is that data on alcohol consumption were self-reported. Nevertheless, these data were collected at baseline survey using validated questionnaires or their variants in all studies. Any misclassification bias is therefore likely to be non-differential, in which case the validity of our observed associations is likely to hold.

Finally, we would like to note a merit of this particular type of genotype-stratified GWAS. If there is a phenotype of interest and a polymorphism that has a strong influence on that phenotype – in this case, alcohol consumption is the phenotype on which *ALDH2* rs671 has a decisive effect - this method is highly effective. The fact that many of the polymorphisms revealed in this study are related to alcohol metabolism may strongly support this notion. Although GWAS was originally conceived as hypothesis-free, the hypothesis-driven approach we used here worked effectively, indicating its potential in the search for new targetable loci. A phenomenon observed in this study is

generally termed gene-gene interaction or SNP-SNP interaction, and its existence has been identified using the candidate approach[51]. GWASs examining interactions with environmental factors using a statistical interaction term are not necessarily successful: in this study, the interaction term approach was not effective despite use of a strong partner, rs671. Accordingly, we propose that hypothesis-based genotype-stratified GWAS represents a promising new approach to discovery.

In conclusion, we performed an *ALDH2* rs671 genotype-stratified GWAS and successfully identified several loci that were associated with alcohol consumption in an rs671-dependent manner. This study further reveals the genetic structure of alcohol consumption, and should deepen our knowledge of the pathogenesis of alcohol-related diseases and disorders.

Methods

Study subjects and genotyping

We performed a genome-wide meta-analysis based on the Japanese Consortium of Genetic Epidemiology studies (J-CGE)[52] and the Nagahama Study[30], both of which comprised general Japanese populations. The J-CGE consisted of the following Japanese population-based and hospital-based studies: the HERPACC Study[25], the J-MICC Study[26,27], the JPHC Study[28], and the TMM Study[29]. Individual study descriptions and an overview of the characteristics of the study populations are provided in the Supplementary Information and Supplementary Table 1. Data and sample collection for the participating cohorts were approved by the respective research ethics committees. All participating studies obtained informed consent from all participants by following the protocols approved by their institutional ethical committees.

Phenotype

Information on alcohol consumption was collected by questionnaire in each study. Because the questionnaires were not homogeneous across the studies, we harmonized the two alcohol consumption phenotypes of drinking status (never versus ever drinker) and daily alcohol intake (g/day) in accordance with each study's criterion. Details are provided in the Supplementary Information.

372 **Quality control and genotype imputation**

373 Quality control for samples and SNPs was performed based on study-specific criteria (Supplemental
374 Table 2). Genotype data in each study were imputed separately based on the 1000 Genomes Project
375 reference panel (Phase 3, all ethnicities)[53]. Phasing was performed with the use of SHAPEIT (v2)[54],
376 and imputation was performed using minimac3[55], minimac4, or IMPUTE (v2)[56]. Information on the
377 study-specific genotyping, imputation, quality control, and analysis tools is provided in
378 Supplementary Table 2. After genotype imputation, further quality control was applied to each study.
379 SNPs with an imputation quality of $r^2 < 0.3$ for minimac3 or minimac4, info < 0.4 for IMPUTE2 or
380 an MAF of <0.01 were excluded.

381

382 **Association analysis of SNPs with daily alcohol intake and drinking status**

383 Association analysis of SNPs with daily alcohol intake and drinking status was performed on three
384 different subject groups: the entire population, subjects with the rs671 GG genotype only, and
385 subjects with the rs671 GA genotype only. Because the number of carriers with the rs671 AA
386 genotype was too small (Supplementary Table 3), association analysis in subjects with the rs671 AA
387 genotype only was not conducted. Daily alcohol intake was base-2 log-transformed ($\log_2$
388 (grammes/day + 1)). The association of daily alcohol intake with SNP allele dose for each study was
389 assessed by linear regression analysis with adjustment for age, age^2, sex, and the first 10 principal
390 components. The association of drinking status with SNP allele dose for each study was assessed by

391 logistic regression analysis with adjustment for age, age^2, sex, and the first 10 principal components.
392 The effect sizes and standard errors estimated in the association analysis were used in the subsequent
393 meta-analysis. The association analysis was conducted using EPACTS
394 (http://genome.sph.umich.edu/wiki/EPACTS), SNPTEST[57], or PLINK2[58].
395 Association analysis, including interaction terms, was performed to evaluate the differential effects
396 of each SNP on daily alcohol intake and drinking status between the GG and GA genotypes of rs671.
397 In the interaction analysis for daily alcohol intake, the linear regression models were fit as:

398 $$\log_2(y+1) = \beta_0 + \beta_{\text{rs671}} x_{\text{rs671}} + \beta_{\text{SNP}} x_{\text{SNP}} + \beta_{\text{interaction}} x_{\text{rs671}} x_{\text{SNP}} + \sum_k \beta_k c_k$$

399 where y is daily alcohol intake (grammes/day). x_{rs671} is the genotype of rs671. The GG genotype
400 is coded as 0, and the GA genotype is coded as 1. Carriers of the AA genotype were excluded from
401 the analysis. x_{SNP} is the imputed genotype coded as [0,2] for each SNP. c_k is a covariate composed
402 of age, age^2, sex, and the first 10 principal components. The effect sizes of the interaction term,
403 $\beta_{\text{interaction}}$, and its standard errors estimated in the association analysis were used in the subsequent
404 meta-analysis. In the interaction analysis for drinking status, the logistic regression model was fit as:

405 $$\ln\left(\frac{p_{\text{ever}}}{1-p_{\text{ever}}}\right) = \beta_0 + \beta_{\text{rs671}} x_{\text{rs671}} + \beta_{\text{SNP}} x_{\text{SNP}} + \beta_{\text{interaction}} x_{\text{rs671}} x_{\text{SNP}} + \sum_k \beta_k c_k$$

406 where p_{ever} is the probability that the subject is an ever drinker. Other variables and procedures are
407 as above. The association analysis, including the interaction term, was conducted using PLINK2[58].
408 To identify studies with inflated GWAS significance, which can result from population stratification,
409 we computed the genomic control λ[59]. Before the meta-analysis, all study-specific results in the

410 association analysis were corrected by multiplying the standard error of the effect size by λ if the λ of
411 that study was greater than 1.
412
413 **Meta-analysis**
414 The meta-analysis was performed with all Japanese subjects in the five cohorts (Supplementary
415 Table 1). The results of association analyses for each SNP across the studies were combined with
416 METAL software[60] by the fixed-effects inverse-variance-weighted method. Heterogeneity of effect
417 sizes was assessed by I^2 and Cochran's Q statistic. The meta-analysis included SNPs for which
418 genotype data were available from at least three studies with a total sample size of at least 20,000
419 individuals for unstratified GWAS or interaction GWAS or 10,000 individuals for rs671-stratified
420 GWAS. The genome-wide significance level α was set to a P value $<5 \times 10^{-8}$. P-values with $<1.0\times10^{-300}$ was calculated with Rmpfr of the R package.
421
422
423 **Functional annotations**
424 To investigate the function of the lead SNP identified in this study, we adopted a series of
425 bioinformatic approaches to collate functional annotations. We first used ANNOVAR[61] to obtain an
426 aggregate set of functional annotations — including gene locations and impacts of amino acid
427 substitutions based on prediction tools, such as SIFT, PolyPhen-2, and CADD — for SNPs with P
428 values $<5 \times10^{-8}$. We also explored eQTLs in tissues considered relevant to daily alcohol intake and

429 drinking status using the GTEx v8 database[62] with regard to the loci identified in this study. The
430 significant criteria for eQTL were based on the GTEx project: variants with a nominal *P* value below
431 the gene-level threshold were regarded as significant. This threshold was determined by permutation
432 tests in the GTEx project to keep the false discovery rate below 5%.
433
434 **Genotyping of rs56884502 and comparison with imputed genotype**
435 *ALDH2* rs56884502 was further genotyped using TaqMan Assays on a 7500 Real-Time PCR System
436 (Applied Biosystems, Foster City, CA, USA) in the selected 96 HERPACC samples which were also
437 genotyped by Illumina HumanCoreExome. We confirmed a 100% match between the imputed and
438 direct genotype data within these samples.
439
440 **Haplotype estimation of SNPs at 12q24**
441 We estimated haplotypes from genotypes of rs56884502 and rs671 at 12q24 for the HERPACC (*n* =
442 96) and 1KGP- JPT[53] (*n* = 104) samples. The genotype of rs56884502 for the HERPACC samples
443 was determined by the method described above, while that of rs671 for these samples was
444 determined by the Illumina HumanCoreExome SNP array. Furthermore, we estimated haplotypes
445 from genotypes of rs671 and 73 SNPs at 12q24.12 which showed genome-wide significance for
446 daily alcohol intake in the rs671 heterozygotes (GA) for the 1KGP- JPT samples (*n* = 104).
447 Haplotype estimation was performed by the Haploview software[63].

References

1. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **392**, 1015-1035 (2018).
2. Wood, A.M. *et al.* Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *Lancet* **391**, 1513-1523 (2018).
3. Kaprio, J. *et al.* Genetic influences on use and abuse of alcohol: a study of 5638 adult Finnish twin brothers. *Alcohol Clin Exp Res* **11**, 349-56 (1987).
4. Heath, A.C., Meyer, J., Jardine, R. & Martin, N.G. The inheritance of alcohol consumption patterns in a general population twin sample: II. Determinants of consumption frequency and quantity consumed. *J Stud Alcohol* **52**, 425-33 (1991).
5. Baik, I., Cho, N.H., Kim, S.H., Han, B.G. & Shin, C. Genome-wide association studies identify genetic loci related to alcohol consumption in Korean men. *Am J Clin Nutr* **93**, 809-16 (2011).
6. Schumann, G. *et al.* Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc Natl Acad Sci U S A* **108**, 7119-24 (2011).
7. Takeuchi, F. *et al.* Confirmation of ALDH2 as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circ J* **75**, 911-8 (2011).
8. Yang, X. *et al.* Common variants at 12q24 are associated with drinking behavior in Han Chinese. *Am J Clin Nutr* **97**, 545-51 (2013).
9. Schumann, G. *et al.* KLB is associated with alcohol drinking, and its gene product beta-Klotho is necessary for FGF21 regulation of alcohol preference. *Proc Natl Acad Sci U S A* **113**, 14372-14377 (2016).
10. Clarke, T.K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol Psychiatry* **22**, 1376-1384 (2017).
11. Jorgenson, E. *et al.* Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Mol Psychiatry* **22**, 1359-1367 (2017).
12. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244 (2019).
13. Zhong, V.W. *et al.* A genome-wide association study of bitter and sweet beverage consumption. *Hum Mol Genet* **28**, 2449-2457 (2019).
14. Matoba, N. *et al.* GWAS of 165,084 Japanese individuals identified nine loci associated with dietary habits. *Nat Hum Behav* **4**, 308-316 (2020).
15. Druesne-Pecollo, N. *et al.* Alcohol and genetic polymorphisms: effect on risk of alcohol-related

cancer. *Lancet Oncol* **10**, 173-80 (2009).

16. Chen, C.H., Ferreira, J.C., Gross, E.R. & Mochly-Rosen, D. Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiol Rev* **94**, 1-34 (2014).
17. Marchitti, S.A., Brocker, C., Stagos, D. & Vasiliou, V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol* **4**, 697-720 (2008).
18. Koyanagi, Y.N. *et al.* Across-Site Differences in the Mechanism of Alcohol-Induced Digestive Tract Carcinogenesis: An Evaluation by Mediation Analysis. *Cancer Res* **80**, 1601-1610 (2020).
19. Chen, C.C. *et al.* Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *Am J Hum Genet* **65**, 795-807 (1999).
20. Matsuo, K. *et al.* Alcohol dehydrogenase 2 His47Arg polymorphism influences drinking habit independently of aldehyde dehydrogenase 2 Glu487Lys polymorphism: analysis of 2,299 Japanese subjects. *Cancer Epidemiol Biomarkers Prev* **15**, 1009-13 (2006).
21. Lai, C.L. *et al.* Dominance of the inactive Asian variant over activity and protein contents of mitochondrial aldehyde dehydrogenase 2 in human liver. *Alcohol Clin Exp Res* **38**, 44-50 (2014).
22. Wang, X., Sheikh, S., Saigal, D., Robinson, L. & Weiner, H. Heterotetramers of human liver mitochondrial (class 2) aldehyde dehydrogenase expressed in Escherichia coli. A model to study the heterotetramers expected to be found in Oriental people. *J Biol Chem* **271**, 31172-8 (1996).
23. Chang, J.S., Hsiao, J.R. & Chen, C.H. ALDH2 polymorphism and alcohol-related cancers in Asians: a public health perspective. *J Biomed Sci* **24**, 19 (2017).
24. Koyanagi, Y.N. *et al.* Development of a prediction model and estimation of cumulative risk for upper aerodigestive tract cancer on the basis of the aldehyde dehydrogenase 2 genotype and alcohol consumption in a Japanese population. *Eur J Cancer Prev* **26**, 38-47 (2017).
25. Hamajima, N. *et al.* Gene-environment Interactions and Polymorphism Studies of Cancer Risk in the Hospital-based Epidemiologic Research Program at Aichi Cancer Center II (HERPACC-II). *Asian Pac J Cancer Prev* **2**, 99-107 (2001).
26. Hamajima, N. & Group, J.M.S. The Japan Multi-Institutional Collaborative Cohort Study (J-MICC Study) to detect gene-environment interactions for cancer. *Asian Pac J Cancer Prev* **8**, 317-23 (2007).
27. Wakai, K. *et al.* Profile of participants and genotype distributions of 108 polymorphisms in a cross-sectional study of associations of genotypes with lifestyle and clinical factors: a project in the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study. *J Epidemiol* **21**, 223-35 (2011).
28. Tsugane, S. & Sawada, N. The JPHC study: design and some findings on the typical Japanese diet. *Jpn J Clin Oncol* **44**, 777-82 (2014).
29. Hozawa, A. *et al.* Study Profile of the Tohoku Medical Megabank Community-Based Cohort Study. *J Epidemiol* **31**, 65-76 (2021).

30. Funada, S. *et al.* Cross-Sectional Epidemiological Analysis of the Nagahama Study for Correlates of Overactive Bladder: Genetic and Environmental Considerations. *J Urol* **199**, 774-778 (2018).
31. Luczak, S.E., Glatt, S.J. & Wall, T.L. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians. *Psychol Bull* **132**, 607-21 (2006).
32. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* **88**, 586-98 (2011).
33. Gelernter, J. *et al.* Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry* **19**, 41-9 (2014).
34. Zuo, L. *et al.* NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk gene region specific for alcohol dependence in subjects of European descent. *Drug Alcohol Depend* **129**, 254-64 (2013).
35. Agrawal, A., Freedman, N.D. & Bierut, L.J. Genome-wide association studies of alcohol intake--a promising cocktail? *Am J Clin Nutr* **93**, 681-3 (2011).
36. Husemoen, L.L. *et al.* The association of ADH and ALDH gene variants with alcohol drinking habits and cardiovascular disease risk factors. *Alcohol Clin Exp Res* **32**, 1984-91 (2008).
37. Linneberg, A. *et al.* Genetic determinants of both ethanol and acetaldehyde metabolism influence alcohol hypersensitivity and drinking behaviour among Scandinavians. *Clin Exp Allergy* **40**, 123-30 (2010).
38. Way, M.J., Ali, M.A., McQuillin, A. & Morgan, M.Y. Genetic variants in ALDH1B1 and alcohol dependence risk in a British and Irish population: A bioinformatic and genetic study. *PLoS One* **12**, e0177009 (2017).
39. Jackson, B.C., Reigan, P., Miller, B., Thompson, D.C. & Vasiliou, V. Human ALDH1B1 polymorphisms may affect the metabolism of acetaldehyde and all-trans retinaldehyde--in vitro studies and computational modeling. *Pharm Res* **32**, 1648-62 (2015).
40. Larson, H.N., Weiner, H. & Hurley, T.D. Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase "Asian" variant. *J Biol Chem* **280**, 30550-6 (2005).
41. Oota, H. *et al.* The evolution and population genetics of the ALDH2 locus: random genetic drift, selection, and low levels of recombination. *Ann Hum Genet* **68**, 93-109 (2004).
42. Okada, Y. *et al.* Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nat Commun* **9**, 1631 (2018).
43. Stagos, D. *et al.* Aldehyde dehydrogenase 1B1: molecular cloning and characterization of a novel mitochondrial acetaldehyde-metabolizing enzyme. *Drug Metab Dispos* **38**, 1679-87 (2010).
44. Jackson, B.C. *et al.* Comparative genomics, molecular evolution and computational modeling of ALDH1B1 and ALDH2. *Chem Biol Interact* **202**, 11-21 (2013).
45. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell

types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).

46. Zhou, S.L. *et al.* Ethanol up-regulates fatty acid uptake and plasma membrane expression and export of mitochondrial aspartate aminotransferase in HepG2 cells. *Hepatology* **27**, 1064-74 (1998).
47. Takeshita, T., Yang, X. & Morimoto, K. The ALDH2 genotype, alcohol intake, and liver-function biomarkers among Japanese male workers. *Hum Genet* **106**, 589-93 (2000).
48. Nagata, N., Watanabe, N., Tsuda, M., Tsukamoto, H. & Matsuzaki, S. Relationship between serum levels of anti-low-density lipoprotein-acetaldehyde-adduct antibody and aldehyde dehydrogenase 2 heterozygotes in patients with alcoholic liver injury. *Alcohol Clin Exp Res* **23**, 24S-28S (1999).
49. Kwon, H.J. *et al.* Aldehyde dehydrogenase 2 deficiency ameliorates alcoholic fatty liver but worsens liver inflammation and fibrosis in mice. *Hepatology* **60**, 146-57 (2014).
50. Engs, R.C., Hanson, D.J., Gliksman, L. & Smythe, C. Influence of religion and culture on drinking behaviours: a test of hypotheses between Canada and the USA. *Br J Addict* **85**, 1475-82 (1990).
51. Dingler, F.A. *et al.* Two Aldehyde Clearance Systems Are Essential to Prevent Lethal Formaldehyde Accumulation in Mice and Humans. *Mol Cell* **80**, 996-1012 e9 (2020).
52. Suzuki, S. *et al.* Body mass index and colorectal cancer risk: A Mendelian randomization study. *Cancer Sci* **112**, 1579-1588 (2021).
53. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
54. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).
55. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-7 (2016).
56. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
57. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
58. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
59. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
60. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
61. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
62. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human

tissues. *Science* **369**, 1318-1330 (2020).

63. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5 (2005).

605 **Acknowledgements**

606 We are grateful to the participants and all staff members in this study. This study was supported by
607 the National Cancer Center Research and Development Fund (28-A-19 and 31-A-18). The
608 HERPACC Study was supported by Grants-in-Aid for Scientific Research from the Ministry of
609 Education, Science, Sports, Culture and Technology of Japan Priority Areas of Cancer (17015018),
610 Innovative Areas (221S0001), and JSPS KAKENHI Grants (JP16H06277 [CoBiA], JP26253041,
611 JP20K10463) and Grant-in-Aid for the Third Term Comprehensive 10-year Strategy for Cancer
612 Control from the Ministry of Health, Labour and Welfare of Japan. Yamagiwa-Yoshida Memorial
613 UICC International Cancer Study Grants supported the initial component of the alcohol consumption
614 GWAS within the HERPACC Study. The J-MICC Study was supported by Grants-in-Aid for
615 Scientific Research for Priority Areas of Cancer (No. 17015018) and Innovative Areas (No.
616 221S0001) and by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant (No.
617 16H06277 [CoBiA]) from the Japanese Ministry of Education, Culture, Sports, Science and
618 Technology. This work has also been supported in part by funding for the BioBank Japan Project
619 from the Japan Agency for Medical Research and Development since April 2015, and the Ministry of
620 Education, Culture, Sports, Science and Technology from April 2003 to March 2015. We thank Drs
621 Nobuyuki Hamajima and Hideo Tanaka for their work in initiating and organizing the J-MICC Study
622 as former principal investigators. The JPHC Study was supported by the National Cancer Center
623 Research and Development Fund (23-A-31 [toku], 26-A-2, 28-A-19, 29-A-4 and 31-A-18), the

Practical Research for Innovative Cancer Control (JP16ck0106095 and JP19ck0106266) from the Japan Agency for Medical Research and Development, and a Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan (from 1989 to 2010). The TMM Study was supported by grants from the Reconstruction Agency, from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), and from the Japan Agency for Medical Research and Development (AMED) [Grant no. JP20km0105001, JP20km0105002, JP20km0105003, and JP20km0105004]. The Nagahama Study was supported by JSPS, a Grant-in-Aid for Scientific Research (C), KAKENHI Grant Number JP17K07255 and JP17KT0125, and the Practical Research Project for Rare/Intractable Diseases from Japan Agency for Medical Research and Development, AMED, under Grant Numbers JP16ek0109070h0003, JP18kk0205008h0003, JP18kk0205001s0703 JP19ek0109283h0003, and JP19ek0109348h0002. The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review and approval of the article; or the decision to submit the article for publication. Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Ethical approval:

643 All studies were approved by their respective institutional review boards.

644

645 **Conflicts of interest:** The authors declare no potential conflicts of interest.

646

Figure legends

Figure 1. Manhattan plots of the GWAS of daily alcohol intake

The results for (a) rs671 wild-type homozygotes (GG), (b) rs671 heterozygotes (GA), (c) unstratified, and (d) interaction with rs671 are shown. The position on each chromosome (x-axis) and the observed $-\log_{10}(P \text{ value})$ (y-axis) of all tested genetic variants are shown. The solid red and gray lines indicate genome-wide and suggestive significance levels, respectively. Blue triangles represent loci containing SNPs with P values of $<1\times10^{-15}$.

Figure 2. Manhattan plots of the GWAS of drinking status

The results for (a) rs671 wild-type homozygotes (GG), (b) rs671 heterozygotes (GA), (c) unstratified, and (d) interaction with rs671 are shown. The position on each chromosome (x-axis) and the observed $-\log_{10}(P \text{ value})$ (y-axis) of all tested genetic variants are shown. The solid red and gray lines indicate genome-wide and suggestive significance level, respectively. Blue triangles represent loci containing SNPs with P values of $<1\times10^{-15}$.

Figure 3. Genomic loci reaching genome-wide significance in either analysis for association with daily alcohol intake

Direction of effects of identified variants other than rs671 is presented as a heatmap. Estimates with a single asterisk show suggestive significance ($P < 5.0 \times 10^{-6}$). Estimates with double asterisks show

genome-wide significance ($P < 5.0 \times 10^{-8}$). Lead SNP in each locus is highlighted with its estimates in bold. SNP, single nucleotide polymorphism; Ref, reference allele; Alt, alternative allele; SE, standard error; Het*P*, *P* value from test of heterogeneity.

Figure 4. Genomic loci reaching genome-wide significance in either analysis for association with drinking status

Direction of effects of identified variants other than rs671 is presented as a heatmap (with colors indicating associated normalized ORs). Estimates with a single asterisk show suggestive significance ($P < 5.0 \times 10^{-6}$). Estimates with double asterisks show genome-wide significance ($P < 5.0 \times 10^{-8}$). Lead SNP in each locus is highlighted with its estimates in bold. SNP, single nucleotide polymorphism; Ref, reference allele; Alt, alternative allele; OR, odds ratio; 95% CI, 95% confidence interval; Het*P*, *P* value from test of heterogeneity.

Figure 5. Regional association plots of the identified novel regions

Regional association plots for (a) daily alcohol intake and (b) drinking status in rs671 heterozygotes are shown. The vertical axis indicates the $-\log_{10}$(*P* value) for the assessment of the association of each SNP with daily alcohol intake or drinking status. The black line represents a genome-wide significance threshold of 5.0×10^{-8}. The colors indicate the LD (r^2) between each lead SNP and neighboring SNPs based on the JPT population in the 1000 Genomes Project Phase 3.

Figure 1. Manhattan plots of the GWAS of daily alcohol intake

The results for (a) rs671 wild-type homozygotes (GG), (b) rs671 heterozygotes (GA), (c) unstratified, and (d) interaction with rs671 are shown. The position on each chromosome (x-axis) and the observed $-\log_{10}(P \text{ value})$ (y-axis) of all tested genetic variants are shown. The solid red and gray lines indicate genome-wide and suggestive significance levels, respectively. Blue triangles represent loci containing SNPs with P values of $<1 \times 10^{-15}$.

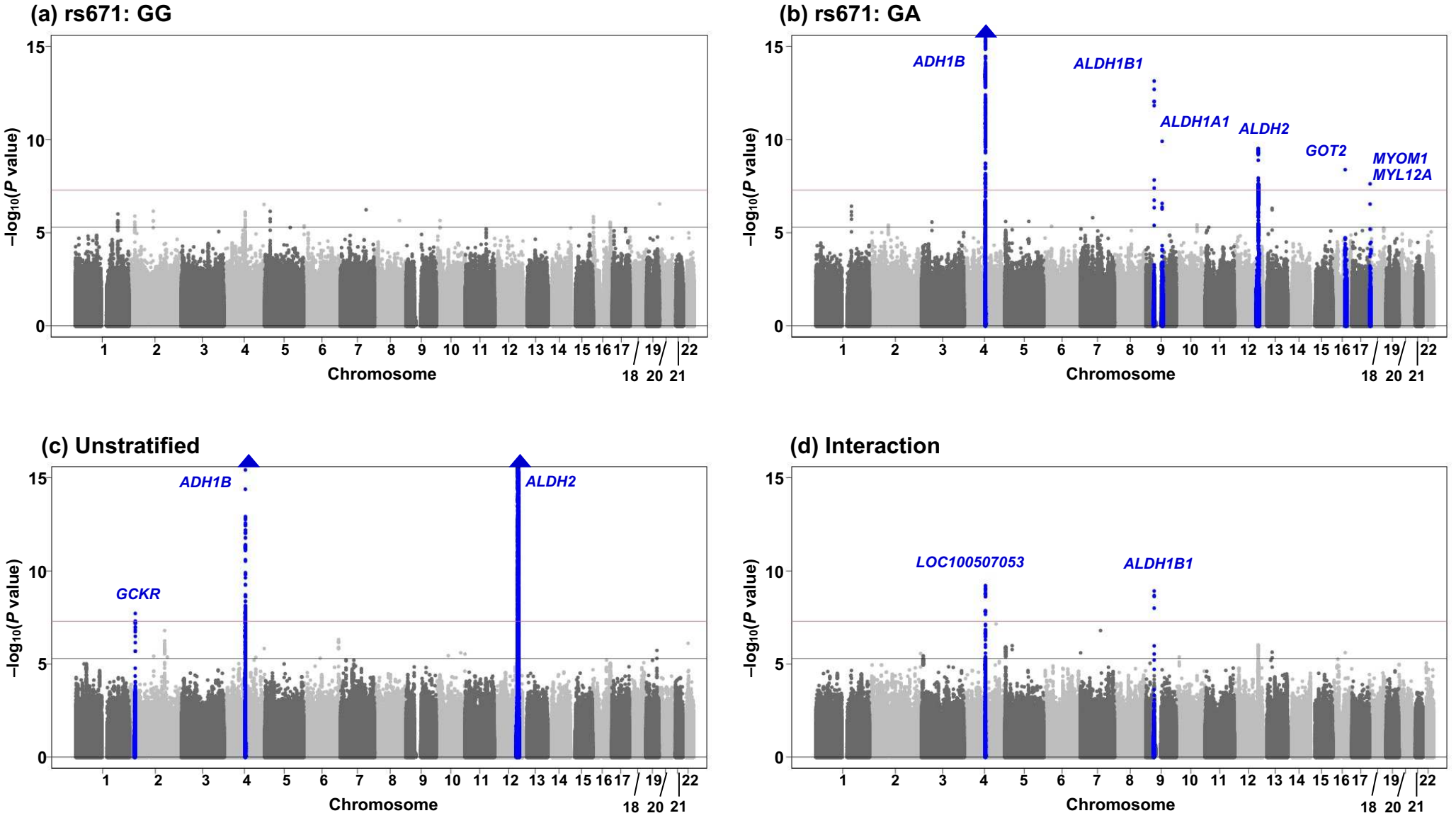


Figure 2. Manhattan plots of the GWAS of drinking status

The results for (a) rs671 wild-type homozygotes (GG), (b) rs671 heterozygotes (GA), (c) unstratified, and (d) interaction with rs671 are shown. The position on each chromosome (x-axis) and the observed $-\log_{10}(P \text{ value})$ (y-axis) of all tested genetic variants are shown. The solid red and gray lines indicate genome-wide and suggestive significance level, respectively. Blue triangles represent loci containing SNPs with P values of $<1 \times 10^{-15}$.

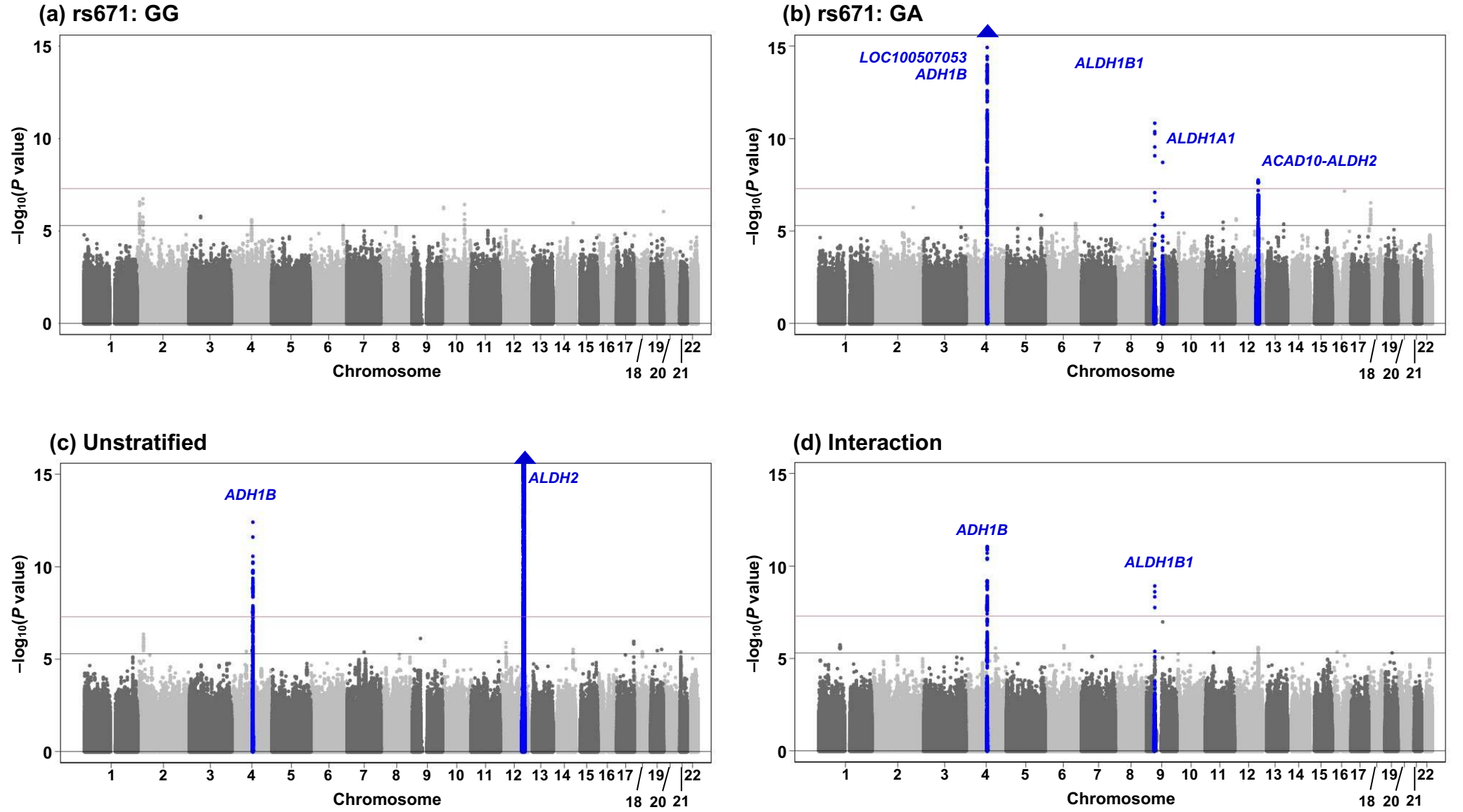


Figure 3. Genomic loci reaching genome-wide significance in either analysis for association with daily alcohol intake
Direction of effects of identified variants other than rs671 is presented as a heatmap. Estimates with a single asterisk show suggestive significance ($P < 5.0 \times 10^{-6}$). Estimates with double asterisks show genome-wide significance ($P < 5.0 \times 10^{-8}$). Lead SNP in each locus is highlighted with its estimates in bold. SNP, single nucleotide polymorphism; Ref, reference allele; Alt, alternative allele; freq., frequency; SE, standard error; Het*P*, *P* value from test of heterogeneity.

SNP *Gene*	Locus	Position	Function	Ref/Alt	Alt freq. GG GA All	***β*** **(SE)** ***P*** I^2, Het*P* GG	 GA	 Unstratified	 Interaction
rs1260326 *GCKR*	2p23.3	27730940	nonsynonymous	T/C	0.445 0.441 0.443	0.077* (0.016) 1.29×10^{-6} 33.4, 0.14	0.064 (0.02) 1.38×10^{-3} 22.2, 0.24	**0.074** (0.013)** 1.93×10^{-8} 49.1, 0.04	-0.021 (0.027) 4.40×10^{-1} 4.9, 0.40
rs1813977 *LOC100507053*	4q23	100216151	ncRNA_intronic	A/G	0.826 0.827 0.827	-0.068 (0.021) 1.17×10^{-3} 0, 0.97	-0.286** (0.027) 6.88×10^{-27} 59.3, 0.01	-0.128** (0.017) 1.77×10^{-13} 0, 0.52	**-0.219** (0.035)** 6.20×10^{-10} 33.6, 0.14
rs10005290 *ADH1B*	4q23	100229410	intronic	A/C	0.830 0.830 0.831	-0.066 (0.021) 1.65×10^{-3} 0, 0.98	**-0.287** (0.027)** 4.99×10^{-27} 58.7, 0.01	-0.127** (0.018) 3.00×10^{-13} 0, 0.46	-0.218** (0.036) 8.88×10^{-10} 29.2, 0.18
rs1229984 *ADH1B*	4q23	100239319	nonsynonymous	T/C	0.263 0.260 0.261	0.085 (0.019) 5.18×10^{-6} 0, 0.90	0.251** (0.024) 3.44×10^{-26} 62.6, 0.004	**0.133** (0.016)** 1.05×10^{-17} 0, 0.51	0.161* (0.031) 2.51×10^{-7} 35.3, 0.13
rs2228093 *ALDH1B1*	9p13.2	38396002	nonsynonymous	C/T	0.341 0.352 0.345	0.008 (0.019) 6.55×10^{-1} 43.4, 0.07	**-0.169** (0.023)** 7.23×10^{-14} 41.1, 0.08	-0.063 (0.015) 2.99×10^{-5} 0, 0.80	**-0.184** (0.03)** 1.20×10^{-9} 65.4, 0.002
rs8187929 *ALDH1A1*	9q21.13	75540504	nonsynonymous	T/A	0.032 0.034 0.033	0.014 (0.045) 7.48×10^{-1} 21.8, 0.24	**0.351** (0.055)** 1.24×10^{-10} 0, 0.72	0.142 (0.037) 1.11×10^{-4} 16.3, 0.29	0.315 (0.074) 2.03×10^{-5} 0, 0.74
rs56884502 *ALDH2*	12q24.12	112207300	intronic	T/A	0.202 0.106 0.158	-0.004 (0.02) 8.54×10^{-1} 28.3, 0.18	**-0.217** (0.035)** 3.05×10^{-10} 18.7, 0.27	0.274** (0.018) 2.44×10^{-53} 51.9, 0.03	-0.204* (0.042) 1.04×10^{-6} 41.9, 0.08
rs671 *ALDH2*	12q24.12	112241766	nonsynonymous	G/A	- - 0.233			**-1.197** (0.014)** 9.83×10^{-1524} 95.8, 3.10×10^{-41}	
rs73550818 *GOT2*	16q21	58764855	intronic	C/A	0.512 0.497 0.504	0.021 (0.016) 2.08×10^{-1} 5.4, 0.39	**-0.121** (0.021)** 4.13×10^{-9} 9.5, 0.36	-0.03 (0.014) 2.49×10^{-2} 31.2, 0.16	-0.128* (0.027) 2.47×10^{-6} 1.6, 0.42
rs572435541 *MYOM1/MYL12A*	18p11.31	3238841	intergenic	C/G	0.016 0.017 0.016	0.031 (0.072) 6.64×10^{-1} 0, 0.88	**0.474** (0.085)** 2.38×10^{-8} 0, 0.58	0.129 (0.059) 2.84×10^{-2} 0, 1.00	0.277 (0.118) 1.88×10^{-2} 0, 0.93

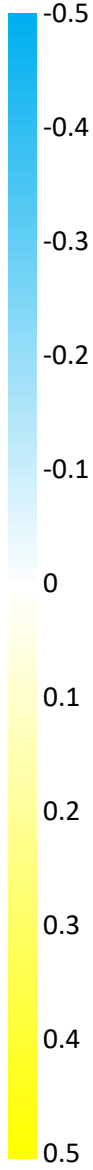
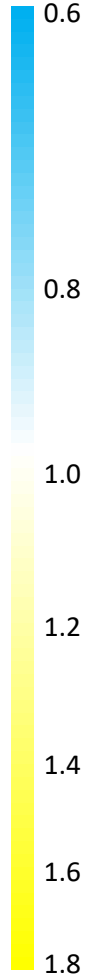


Figure 4. Genomic loci reaching genome-wide significance in either analysis for association with drinking status

Direction of effects of identified variants other than rs671 is presented as a heatmap (with colors indicating associated normalized ORs). Estimates with a single asterisk show suggestive significance ($P < 5.0 \times 10^{-6}$). Estimates with double asterisks show genome-wide significance ($P < 5.0 \times 10^{-8}$). Lead SNP in each locus is highlighted with its estimates in bold. SNP, single nucleotide polymorphism; Ref, reference allele; Alt, alternative allele; freq., frequency; OR, odds ratio; 95% CI, 95% confidence interval; Het*P*, *P* value from test of heterogeneity.

SNP *Gene*	Locus	Position	Function	Ref/Alt	Alt freq. GG GA All	OR (95%CI) *P* I^2, Het*P* GG	GA	Unstratified	Interaction
rs35333426 *LOC100507053 /ADH1B*	4q23	100224036	intergenic	G/A	0.827 0.831 0.829	0.93 (0.88, 0.99) 2.52×10^{-2} 0, 0.76	**0.68** (0.63, 0.74)** 3.10×10^{-21} 10.1, 0.35	0.87** (0.83, 0.90) 5.70×10^{-11} 0, 0.76	0.70** (0.63, 0.77) 1.35×10^{-11} 0, 0.67
rs10005290 *ADH1B*	4q23	100229410	intronic	A/C	0.830 0.833 0.832	0.94 (0.88, 1.00) 4.26×10^{-2} 0, 0.79	0.68** (0.63, 0.74) 4.17×10^{-21} 8.0, 0.37	0.87** (0.83, 0.91) 1.92×10^{-10} 0, 0.79	**0.69** (0.62, 0.77)** 8.92×10^{-12} 0, 0.73
rs1229984 *ADH1B*	4q23	100239319	nonsynonymous	T/C	0.263 0.257 0.260	1.10 (1.04, 1.16) 7.27×10^{-4} 0, 0.85	1.37** (1.28, 1.47) 9.81×10^{-19} 0, 0.50	**1.15** (1.11, 1.20)** 3.91×10^{-13} 0, 0.94	1.32** (1.20, 1.45) 6.30×10^{-9} 0, 0.58
rs2228093 *ALDH1B1*	9p13.2	38396002	nonsynonymous	C/T	0.337 0.350 0.342	1.02 (0.96, 1.07) 5.69×10^{-1} 34.6, 0.13	**0.80** (0.75, 0.85)** 1.48×10^{-11} 57.7, 0.01	0.93 (0.89, 0.96) 4.17×10^{-5} 0, 1.00	**0.76** (0.70, 0.83)** 1.21×10^{-9} 69.8, 0.0005
rs8187929 *ALDH1A1*	9q21.13	75540504	nonsynonymous	T/A	0.032 0.034 0.032	0.94 (0.83, 1.07) 3.63×10^{-1} 21.3, 0.25	**1.63** (1.39, 1.92)** 1.94×10^{-9} 0, 0.49	1.16 (1.06, 1.27) 1.45×10^{-3} 0, 0.50	1.80* (1.45, 2.23) 1.06×10^{-7} 0, 0.50
rs7978737 *ACAD10/ALDH2*	12q24.12	112196611	intergenic	C/T	0.199 0.112 0.160	1.00 (0.95, 1.06) 9.21×10^{-1} 22.6, 0.23	**0.75** (0.67, 0.83)** 1.76×10^{-8} 39.2, 0.10	1.42** (1.36, 1.49) 2.62×10^{-53} 34.3, 0.13	0.74* (0.66, 0.84) 2.53×10^{-6} 32.1, 0.15
rs671 *ALDH2*	12q24.12	112241766	nonsynonymous	G/A	- - 0.238			**0.16** (0.15, 0.17)** 5.60×10^{-1038} 52.3, 0.03	



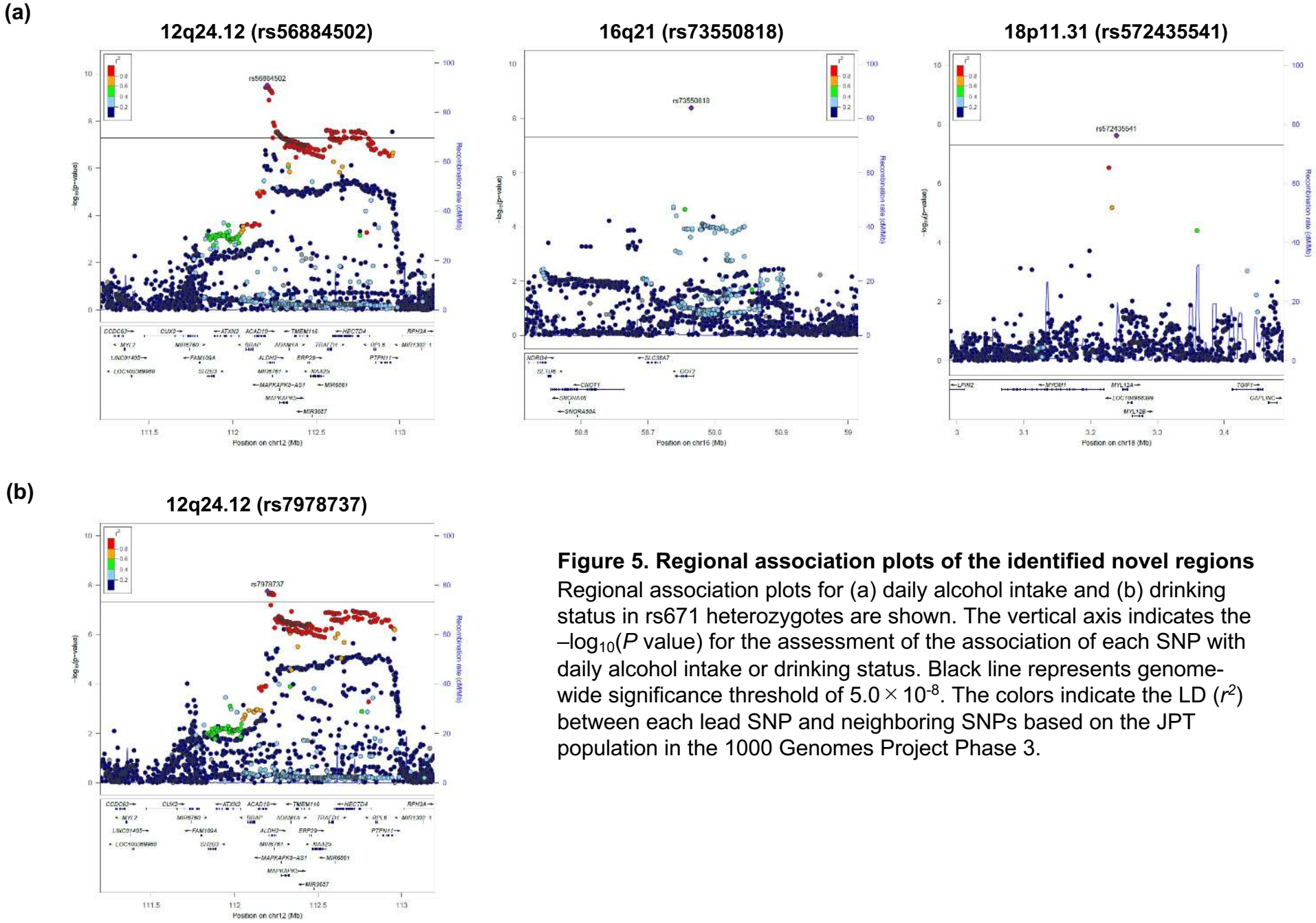


Figure 5. Regional association plots of the identified novel regions
Regional association plots for (a) daily alcohol intake and (b) drinking status in rs671 heterozygotes are shown. The vertical axis indicates the $-\log_{10}(P \text{ value})$ for the assessment of the association of each SNP with daily alcohol intake or drinking status. Black line represents genome-wide significance threshold of 5.0×10^{-8}. The colors indicate the LD (r^2) between each lead SNP and neighboring SNPs based on the JPT population in the 1000 Genomes Project Phase 3.