

1 Genotype-stratified GWAS meta-analysis reveals novel loci associated with alcohol 2 consumption

3
4 Yuriko N. Koyanagi^{1#}, Masahiro Nakatochi^{2##*}, Hidemi Ito^{1,3}, Yumiko Kasugai^{4,5}, Akira Narita⁶,
5 Takahisa Kawaguchi⁷, Hiroaki Ikezaki^{8,9}, Asahi Hishida¹⁰, Megumi Hara¹¹, Toshiro Takezaki¹²,
6 Teruhide Koyama¹³, Haruo Mikami¹⁴, Sadao Suzuki¹⁵, Sakurako Katsuura-Kamano¹⁶,
7 Kiyonori Kuriki¹⁷, Yasuyuki Nakamura¹⁸, Kenji Takeuchi¹⁰, Atsushi Hozawa¹⁹, Kengo Kinoshita⁶,
8 Yoichi Sutoh²⁰, Kozo Tanno^{21,22}, Atsushi Shimizu^{20,23}, Isao Oze⁴, Yukino Kawakatsu⁴,
9 Yukari Taniyama¹, Issei Imoto²⁴, Yasuharu Tabara⁷, Meiko Takahashi⁷, Kazuya Setoh⁷,
10 Shiori Suzuki²⁵, Atsushi Goto^{25,26}, Ryoko Katagiri²⁵, Taiki Yamaji²⁵, Norie Sawada²⁵,
11 Shoichiro Tsugane^{25,27}, Kenji Wakai¹⁰, Masayuki Yamamoto⁶, Makoto Sasaki^{28,29},
12 Fumihiko Matsuda⁷, Motoki Iwasaki²⁵, Paul Brennan³⁰, Keitaro Matsuo^{4,5*}

- 13
14 1. Division of Cancer Information and Control, Department of Preventive Medicine, Aichi Cancer Center Research
15 Institute, Nagoya, Japan.
16 2. Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School
17 of Medicine, Nagoya, Japan.
18 3. Department of Descriptive Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya,
19 Japan.
20 4. Division of Cancer Epidemiology and Prevention, Department of Preventive Medicine, Aichi Cancer Center
21 Research Institute, Nagoya, Japan.
22 5. Department of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan.
23 6. Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai,
24 Japan.
25 7. Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.
26 8. Department of General Internal Medicine, Kyushu University Hospital, Fukuoka, Japan.
27 9. Department of Comprehensive General Internal Medicine, Faculty of Medical Sciences, Kyushu University,
28 Fukuoka, Japan.
29 10. Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan.
30 11. Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, Japan.
31 12. Department of International Island and Community Medicine, Kagoshima University Graduate School of
32 Medical and Dental Sciences, Kagoshima, Japan.
33 13. Department of Epidemiology for Community Health and Medicine, Kyoto Prefectural University of Medicine,
34 Kyoto, Japan.
35 14. Cancer Prevention Center, Chiba Cancer Center Research Institute, Chiba, Japan.
36 15. Department of Public Health, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.
37 16. Department of Preventive Medicine, Tokushima University Graduate School of Biomedical Sciences,
38 Tokushima, Japan.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

- 39 17. Laboratory of Public Health, Division of Nutritional Sciences, School of Food and Nutritional Sciences,
40 University of Shizuoka, Shizuoka, Japan.
- 41 18. Department of Public Health, Shiga University of Medical Science, Otsu, Japan.
- 42 19. Department of Preventive Medicine and Epidemiology, Tohoku Medical Megabank Organization, Tohoku
43 University, Sendai, Japan.
- 44 20. Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical
45 University, Iwate, Japan.
- 46 21. Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University, Iwate, Japan.
- 47 22. Division of Clinical Research and Epidemiology, Iwate Tohoku Medical Megabank Organization, Iwate Medical
48 University, Iwate, Japan.
- 49 23. Division of Biomedical Information Analysis, Institute for Biomedical Sciences, Iwate Medical University,
50 Iwate, Japan.
- 51 24. Aichi Cancer Center Research Institute, Nagoya, Japan.
- 52 25. Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan.
- 53 26. Department of Health Data Science, Graduate School of Data Science, Yokohama City University, Yokohama,
54 Japan.
- 55 27. National Institute of Health and Nutrition, National Institutes of Biomedical Innovation, Health and Nutrition,
56 Tokyo, Japan.
- 57 28. Division of Ultrahigh Field MRI, Institute for Biomedical Sciences, Iwate Medical University, Iwate, Japan.
- 58 29. Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan.
- 59 30. Genomic Epidemiology Branch, International Agency for Research on Cancer, Lyon, France.

60

61 #These authors contributed equally to this work.

62

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

64 Masahiro Nakatochi, PhD, Public Health Informatics Unit, Department of Integrated Health
65 Sciences, Nagoya University Graduate School of Medicine, 1-1-20 Daiko-Minami, Higashi-ku,
66 Nagoya, Aichi, 461-8673, Japan

67 Phone/Fax: +81-52-719-1923, E-mail: mnakatochi@met.nagoya-u.ac.jp

68

69 Keitaro Matsuo, MD, PhD, MSc, Division of Cancer Epidemiology and Prevention, Department of
70 Preventive Medicine, Aichi Cancer Center Research Institute, 1-1 Kanokoden Chikusa-ku, Nagoya,
71 Aichi, 464-8681, Japan

72 Phone: +81-52-764-2982, Fax: +81-52-763-5233, E-mail: kmatsuo@aichi-cc.jp

73

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

76

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

83 **Abstract**

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

94

95 Alcohol consumption is a major contributor to mortality and influences risk for various human
96 diseases and disorders¹. Even moderate consumption may have a substantial impact on mortality².
97 Indeed, the latest Global Burden of Disease study on alcohol use states that the level of consumption
98 should be reduced to zero to minimize health risk¹. Alcohol consumption has been considered a
99 heritable trait^{3,4}. The number of genetic studies on alcohol consumption is increasing⁵⁻¹⁴, and the
100 genetic variants that are consistently reproducible are those of genes encoding alcohol-metabolizing
101 enzymes^{5,7,8,10-14}. Ingested alcohol is predominantly metabolized to acetaldehyde through alcohol
102 dehydrogenase (ADH) enzymes, and aldehyde dehydrogenase (ALDH) enzymes further catalyze the
103 oxidation of acetaldehyde to acetate¹⁵. Notably, rs671 (c.1510G>A [p.Glu504Lys]), a functional
104 single nucleotide polymorphism (SNP) in the *ALDH2* gene which is highly prevalent in East
105 Asians¹⁶, is a strong and well-known determinant of alcohol consumption. Every previous genome-
106 wide association study (GWAS) in East Asians^{5,7,8,11,14} identified the strongest signals in the rs671
107 variant (or variants in strong linkage disequilibrium [LD] with rs671), ranging from $P < 1.0 \times 10^{-58}$ (n
108 = 2,834)⁵ to $P < 1.0 \times 10^{-4,740}$ ($n = 165,084$)¹⁴.

109 Among ALDH isoforms, ALDH2 has by far the highest affinity for acetaldehyde ($K_m < 1$
110 μM) and is primarily responsible for its oxidation^{16,17}. Because the *ALDH2* rs671 variant inactivates
111 ALDH2 enzymatic activity, individuals who are heterozygous (GA) or homozygous (AA) for this
112 variant experience a rapid accumulation of blood acetaldehyde after alcohol ingestion¹⁶. This variant
113 thereby increases exposure to the unpleasant effects of acetaldehyde (e.g. flushing, headache,

114 palpitation, and nausea), which in consequence significantly reduces alcohol consumption and
115 thereby confers a protective effect against alcohol-induced carcinogenesis¹⁸. Conversely,
116 heterozygotes (GA) who drink alcohol experience increased susceptibility to carcinogenesis, in
117 particular for head and neck and esophageal cancers, due to higher concentrations of acetaldehyde,
118 one of the most likely carcinogens in alcohol¹⁵. With regard to variant homozygotes (AA), however,
119 these have rarely evidenced an increased cancer risk associated with alcohol, because they are unable
120 to oxidize acetaldehyde, a characteristic which is highly correlated with nondrinking^{19,20}. In contrast,
121 heterozygotes having 16-18% of normal enzyme activity^{21,22} show a broader range of alcohol
122 consumption. Overall, the highest risk group for alcohol-related cancers are heterozygotes^{23,24}, and
123 alcohol consumption among genotypes of rs671 shows distinct genetic heterogeneity.

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

132

133 **Results**

134 **Characteristics of study participants**

135 A total of 40,679 individuals were included in this GWAS meta-analysis of five Japanese cohorts,
136 namely the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC)²⁵
137 Study ($n = 4,958$), the Japan Multi-Institutional Collaborative Cohort (J-MICC)^{26,27} Study ($n =$
138 $13,236$), the Japan Public Health Center-based Prospective (JPHC)²⁸ Study ($n = 10,037$), the Tohoku
139 Medical Megabank Community-Based Cohort (TMM)²⁹ Study ($n = 7,857$), and the Nagahama
140 Prospective Cohort for Comprehensive Human Bioscience (Nagahama) Study³⁰ ($n = 4,591$), after
141 imputation and quality control of individual subject genotype data (Supplementary Information and
142 Supplementary Tables 1 and 2). Median self-reported daily alcohol intake, mean age of participants,
143 number of ever/never drinkers, and proportion of male participants were obtained from the studies
144 and are shown in Supplementary Table 1. The number of participants included in each analysis for
145 daily alcohol intake and drinking status was as follows: rs671 wild-type homozygotes (GG)-only
146 analysis, $n = 23,398$ and $24,514$; heterozygotes (GA)-only analysis, $n = 13,385$ and $13,848$;
147 unstratified analysis, $n = 39,077$ and $40,679$; and interaction analysis, $n = 36,783$ and $38,362$,
148 respectively (Supplementary Table 3). As the number of variant homozygotes (AA) was small ($n =$
149 $2,294$ for daily alcohol intake; $n = 2,317$ for drinking status) and included only 101 subjects in the
150 ever drinking group (Supplementary Table 3), association analysis in variant homozygotes only was
151 not conducted, and these subjects were excluded from the interaction analyses. Quantile-quantile

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

154

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

156 Major results of the genotype-stratified GWAS meta-analysis are summarized in Figures 1-4. In
157 wild-type homozygotes (GG), no genome-wide significant loci were detected for either daily alcohol
158 intake or drinking status (Figures 1a and 2a). In heterozygotes (GA), on the other hand, six and four
159 loci satisfied the genome-wide significance ($P < 5.0 \times 10^{-8}$) for daily alcohol intake and drinking
160 status, respectively (Figures 1b, 2b, 3 and 4). These included three loci that were previously
161 implicated in alcohol consumption, namely chromosome 4q23¹⁰⁻¹⁴, *ALDH1B1*¹⁴, and *ALDH1A1*¹⁴. Of
162 the three remaining loci, two loci (*GOT2* at 16q21, and *MYOM1-MYL12A* at 18p11.31) have not
163 been reported in previous GWASs of alcohol consumption, and one locus (chromosome 12q24.12) is
164 the same locus as rs671. Regional association plots for these novel loci are shown in Figure 5. The
165 lead SNPs are rs56884502 in chromosome 12q24.12 (*ALDH2*), rs73550818 in *GOT2*, and
166 rs572435541 in *MYOM1-MYL12A* for daily alcohol intake, and rs7978737 in chromosome 12q24.12
167 (*ACAD10-ALDH2*) for drinking status. The unstratified GWAS showed three hits (rs1260326 in
168 *GCKR* for daily alcohol intake; rs1229984 in *ADH1B* and rs671 in *ALDH2* for daily alcohol intake
169 and drinking status) (Figures 1c, 2c, 3 and 4), all of which have been previously reported^{5,7-14}.

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

172 A further analysis evaluating variant-rs671 interaction detected two loci (chromosome 4q23
173 and *ALDH1B1*) showing genome-wide significant interaction with rs671 (Figures 1d, 2d, 3 and 4 and
174 Supplementary Figures 3c and 4c). This is the first identification of the interactive effects of these
175 two loci on rs671, although that of rs1229984 in *ADH1B* at 4q23 on rs671 was suggested from a
176 meta-analysis of studies using candidate gene-based approaches³¹. Among the other loci reaching
177 genome-wide significance in either rs671-stratified or unstratified analysis, three loci (*ALDH1A1*,
178 chromosome 12q24.12, and *GOT2*) demonstrated interaction with rs671 with a suggestive
179 significance level ($P < 5.0 \times 10^{-6}$) for daily alcohol intake and/or drinking status (Figures 3 and 4).
180 Regarding the multiple hits on chromosome 4q23 observed through these analyses, rs1813977
181 (interaction), rs35333426 (for drinking status among heterozygotes), and rs10005290 (for daily
182 alcohol intake among heterozygotes) were in strong LD with a functional SNP
183 of *ADH1B* (rs1229984) (all r^2 and D' values are 0.71 and 1.00 in 1000 Genome Project phase 3-
184 Japanese in Tokyo (1KGP- JPT), respectively). We further applied a random effects model³² given
185 that estimates in *ALDH1B1* rs2228093 showed between-study heterogeneity (P values from test of
186 heterogeneity < 0.05) (Figures 3 and 4), but the results did not change substantially ($P = 8.34 \times 10^{-11}$
187 in the interaction analysis for daily alcohol intake; and $P = 4.38 \times 10^{-12}$ and $P = 1.62 \times 10^{-11}$ in the
188 analyses of heterozygotes only and interaction, respectively, for drinking status).

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

197

198 **Effect of a novel SNP within the same locus as rs671**

199 Figures 3 and 4 show the direction of effects of the identified variants other than rs671 under each
200 analysis. Notably, with regard to the novel SNP in chromosome 12q24.12 (*ALDH2* rs56884502,
201 T>A), the A allele of rs56884502, which was associated with decreasing daily alcohol intake in the
202 rs671 heterozygotes ($\beta = -0.217$), showed the opposite direction of effect in the unstratified analysis
203 ($\beta = 0.274$) (Figure 3). This apparently conflicting result was due to strong LD between rs56884502
204 and rs671. The 1KGP-JPT ($n = 104$) and our own direct genotyped data from subjects in the
205 HERPACC Study ($n = 96$) indicated that there were only three rs56884502-rs671 haplotypes,
206 namely T-G, A-G and T-A (Supplementary Table 5). The respective LD coefficients of r^2 and D'
207 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**

226 Of the detected novel variants, rs56884502 and rs7978737 in chromosome 12q24.12 and rs73550818
227 in *GOT2* were found to be eQTL using the GTEx database (Supplementary Table 8). rs56884502 A
228 allele and rs7978737 T allele are associated with decreased expression of *ALDH2* in multiple tissues.
229 rs73550818 A allele is associated with increased expression of *GOT2* in liver ($P = 1.0 \times 10^{-8}$).

230 Discussion

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

239 The failure of other loci to reach a genome-wide significant level in rs671 wild-type
240 homozygotes indicates that the rs671 GG genotype itself is strong enough to make a significant
241 contribution to determining the alcohol consumption phenotype in this population. This is consistent
242 with the observation that this phenotype has been resistant to gene discovery efforts in non-Asian
243 populations, where rs671 is often monomorphic³⁵. Further, the strongest signal for daily alcohol
244 intake in rs671 heterozygotes was observed in *ADH1B* ($P < 5.0 \times 10^{-26}$), followed by *ALDH1B1* ($P =$
245 7.2×10^{-14}) and then *ALDH1A1* ($P = 1.2 \times 10^{-10}$), all of which are associated with the concentration
246 of acetaldehyde. *ADH1B* is the predominant isoform involved in alcohol oxidation, whereas
247 *ALDH1B1* and *ALDH1A1* are the *ALDH* isoforms involved in acetaldehyde oxidation, with the
248 second (K_m 30 μ M) and third (K_m 50–180 μ M) highest affinities for acetaldehyde, respectively¹⁷.

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 GWAS¹⁴, 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 product³⁸ and catalytic inactivity³⁹ 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 rs671 wild-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.

268 The rs671 A allele is predicted to disrupt the structure of not only its own subunit but also its dimer
269 partner, reducing the stability of the tetramic structure of ALDH2 and resulting in a dramatic
270 reduction in enzyme activity⁴⁰. This East Asian-specific SNP is considered to be a relatively young
271 polymorphism⁴¹ and to have been under strong recent natural selection pressure in the Japanese
272 population⁴². On the other hand, the novel SNP at 12q24.12 of rs56884502 is a globally common
273 SNP located <40Kb distant to rs671 (Supplementary Table 4). Given this evidence and the two LD
274 measures of $r^2 < 0.1$ and $D' = 1.0$ for rs56884502 and rs671, we speculate that rs56884502 arose
275 prior to rs671, and that rs671 then arose on a different branch from rs56884502 in the rs56884502-
276 rs671 T-G haplotype without subsequent historic recombination, finally resulting in the three
277 haplotypes of T-G, A-G, and T-A. The protective effect of the rs56884502 A allele against alcohol
278 consumption observed in rs671 heterozygotes can therefore be regarded as a protective effect of the
279 rs56884502-rs671 T-A/A-G diplotype. Accordingly, we hypothesize that rs56884502 (or variants in
280 LD) is associated with reduced enzyme activity via an effect on the rs671 A allele located on the
281 opposite haplotype. However, rs56884502 is located in the intron of *ALDH2* (Supplementary Table
282 4) and the rs56884502 A allele was found to be associated with decreased expression of *ALDH2*
283 (Supplementary Table 8). Further, none of the other 72 SNPs that were in LD with rs56884502 and
284 showed genome-wide significance for daily alcohol intake are located within the coding region
285 (Supplementary Table 6), suggesting that while these SNPs may be potentially associated with
286 expression, they may have no direct effect on the protein structure or tetramer formation of ALDH2

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

294 Our genome-wide analysis indicates the interactive effect of rs2228093 in *ALDH1B1* with
295 rs671. *ALDH1B1* is another mitochondrial ALDH which shares 75% similarity with *ALDH2* at the
296 amino acid sequence level, and is predicted to form a homotetramer, similarly to *ALDH2*⁴³.
297 Although *ALDH1B1* is also able to oxidize acetaldehyde, individuals with the rs671 A allele are
298 reported not to exhibit a compensatory increase in *ALDH1B1* activity⁴⁴. Further, a bioinformatics
299 analysis predicted protein-protein interactions between *ALDH2* and *ALDH1B1*, indicating that
300 *ALDH2* and *ALDH1B1* subunits are likely to form heterotetramers⁴⁴. These findings suggest the
301 hypothesis that the rs671 A allele reduces the catalytic activity of *ALDH1B1*. They also explain the
302 present finding of gene-gene interaction between the rs671 A allele and rs2228093 T allele, both of
303 which are validated and predicted³⁹ to be associated with catalytic inactivity. Moreover, these
304 findings may further explain the limited and conflicting genetic evidence for rs2228093 on drinking

305 in European populations^{36,37}, in which rs2228093 is polymorphic (MAF = 0.15 in 1KGP EUR) but
306 rs671 is monomorphic.

307 The remaining newly identified loci associated with alcohol consumption in this study are
308 *GOT2* and *MYOMI-MYL12A*. The lead SNPs in these loci are located within the non-coding region
309 and the functional effects of these variants are unknown. However, the lead SNP in *GOT2*
310 (rs73550818, C>A) is located in the intron of *GOT2* and showed a protective effect against alcohol
311 consumption in heterozygotes; this may be a suitable target for future study, given that the
312 rs73550818 A allele was shown to be significantly associated with increased levels of aspartate
313 aminotransferase (AST), a biochemical marker for liver injury, in a previous GWAS of 134,154
314 Japanese individuals⁴⁵. *GOT2* encodes the mitochondrial isoform of glutamic-oxaloacetic
315 transaminase; this plays an important role in many processes, including amino acid metabolism,
316 long-chain fatty acid uptake, and the urea and tricarboxylic acid cycles⁴⁶. An *in vivo* study suggested
317 that increased mitochondrial AST among alcoholics is a consequence of the pharmacologic
318 upregulation of *GOT2* gene expression by ethanol, which further mediates fatty acid uptake,
319 resulting in alcoholic fatty liver⁴⁶. Considering that the rs73550818 A allele is associated with
320 increased expression of *GOT2* in the liver (Supplementary Table 8), this allele might be associated
321 with ethanol-induced liver injury. On the other hand, previous studies of rs671 showed significantly
322 lower AST in heterozygotes than in wild-type homozygotes among drinkers⁴⁷, even after adjustment
323 for alcohol intake⁷. An observational study of patients with alcoholic liver injury⁴⁸ and a study of

324 *Aldh2* knockout mice⁴⁹ suggested a protective effect of the rs671 A allele on ethanol-induced liver
325 injury. These findings suggest the opposite effects of the rs73550818 A and rs671 A alleles on
326 ethanol-induced liver injury. The mechanism of the suggested interaction between rs73550818 and
327 rs671 observed in our present study therefore warrants further investigation.

328 This study has several strengths. First, most of the included cohorts were population-based
329 and included a large number of general Japanese individuals, and the possibility of selection bias is
330 likely small. Second, the study involved a single ethnic group with a similar religious and cultural
331 background, making it unlikely that these factors would bias the phenotype of alcohol
332 consumption⁵⁰. An important limitation is that data on alcohol consumption were self-reported.
333 Nevertheless, these data were collected at baseline survey using validated questionnaires or their
334 variants in all studies. Any misclassification bias is therefore likely to be non-differential, in which
335 case the validity of our observed associations is likely to hold.

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

343 generally termed gene-gene interaction or SNP-SNP interaction, and its existence has been identified
344 using the candidate approach⁵¹. GWASs examining interactions with environmental factors using a
345 statistical interaction term are not necessarily successful: in this study, the interaction term approach
346 was not effective despite use of a strong partner, rs671. Accordingly, we propose that hypothesis-
347 based genotype-stratified GWAS represents a promising new approach to discovery.

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

352

353 **Methods**

354 **Study subjects and genotyping**

355 We performed a genome-wide meta-analysis based on the Japanese Consortium of Genetic
356 Epidemiology studies (J-CGE)⁵² and the Nagahama Study³⁰, both of which comprised general
357 Japanese populations. The J-CGE consisted of the following Japanese population-based and hospital-
358 based studies: the HERPACC Study²⁵, the J-MICC Study^{26,27}, the JPHC Study²⁸, and the TMM
359 Study²⁹. Individual study descriptions and an overview of the characteristics of the study populations
360 are provided in the Supplementary Information and Supplementary Table 1. Data and sample
361 collection for the participating cohorts were approved by the respective research ethics committees.
362 All participating studies obtained informed consent from all participants by following the protocols
363 approved by their institutional ethical committees.

364

365 **Phenotype**

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

371

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 minimac3⁵⁵, 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, $\text{info} < 0.4$ 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, age², 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, age², 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>), SNPTTEST⁵⁷, or PLINK2⁵⁸.

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:

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

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, age², sex, and the first 10 principal components. The effect sizes of the interaction term,
403 $\beta_{interaction}$, 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:

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

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 PLINK2⁵⁸.
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 software⁶⁰ 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 $< 1.0 \times 10^{-300}$
421 $^{-300}$ was calculated with Rmpfr of the R package.

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 ANNOVAR⁶¹ 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 $< 5 \times 10^{-8}$. We also explored eQTLs in tissues considered relevant to daily alcohol intake and

429 drinking status using the GTEx v8 database⁶² 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- JPT⁵³ ($n = 104$) 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 software⁶³.

448 **References**

- 449 1. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories,
450 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **392**,
451 1015-1035 (2018).
- 452 2. Wood, A.M. *et al.* Risk thresholds for alcohol consumption: combined analysis of individual-
453 participant data for 599 912 current drinkers in 83 prospective studies. *Lancet* **391**, 1513-1523
454 (2018).
- 455 3. Kaprio, J. *et al.* Genetic influences on use and abuse of alcohol: a study of 5638 adult Finnish
456 twin brothers. *Alcohol Clin Exp Res* **11**, 349-56 (1987).
- 457 4. Heath, A.C., Meyer, J., Jardine, R. & Martin, N.G. The inheritance of alcohol consumption
458 patterns in a general population twin sample: II. Determinants of consumption frequency and
459 quantity consumed. *J Stud Alcohol* **52**, 425-33 (1991).
- 460 5. Baik, I., Cho, N.H., Kim, S.H., Han, B.G. & Shin, C. Genome-wide association studies identify
461 genetic loci related to alcohol consumption in Korean men. *Am J Clin Nutr* **93**, 809-16 (2011).
- 462 6. Schumann, G. *et al.* Genome-wide association and genetic functional studies identify autism
463 susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc Natl*
464 *Acad Sci U S A* **108**, 7119-24 (2011).
- 465 7. Takeuchi, F. *et al.* Confirmation of ALDH2 as a Major locus of drinking behavior and of its
466 variants regulating multiple metabolic phenotypes in a Japanese population. *Circ J* **75**, 911-8
467 (2011).
- 468 8. Yang, X. *et al.* Common variants at 12q24 are associated with drinking behavior in Han
469 Chinese. *Am J Clin Nutr* **97**, 545-51 (2013).
- 470 9. Schumann, G. *et al.* KLB is associated with alcohol drinking, and its gene product beta-Klotho
471 is necessary for FGF21 regulation of alcohol preference. *Proc Natl Acad Sci U S A* **113**, 14372-
472 14377 (2016).
- 473 10. Clarke, T.K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap
474 with other health-related traits in UK Biobank (N=112 117). *Mol Psychiatry* **22**, 1376-1384
475 (2017).
- 476 11. Jorgenson, E. *et al.* Genetic contributors to variation in alcohol consumption vary by
477 race/ethnicity in a large multi-ethnic genome-wide association study. *Mol Psychiatry* **22**, 1359-
478 1367 (2017).
- 479 12. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the
480 genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244 (2019).
- 481 13. Zhong, V.W. *et al.* A genome-wide association study of bitter and sweet beverage consumption.
482 *Hum Mol Genet* **28**, 2449-2457 (2019).
- 483 14. Matoba, N. *et al.* GWAS of 165,084 Japanese individuals identified nine loci associated with
484 dietary habits. *Nat Hum Behav* **4**, 308-316 (2020).
- 485 15. Druesne-Pecollo, N. *et al.* Alcohol and genetic polymorphisms: effect on risk of alcohol-related

- 486 cancer. *Lancet Oncol* **10**, 173-80 (2009).
- 487 16. Chen, C.H., Ferreira, J.C., Gross, E.R. & Mochly-Rosen, D. Targeting aldehyde dehydrogenase
488 2: new therapeutic opportunities. *Physiol Rev* **94**, 1-34 (2014).
- 489 17. Marchitti, S.A., Brocker, C., Stagos, D. & Vasiliou, V. Non-P450 aldehyde oxidizing enzymes:
490 the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol* **4**, 697-720 (2008).
- 491 18. Koyanagi, Y.N. *et al.* Across-Site Differences in the Mechanism of Alcohol-Induced Digestive
492 Tract Carcinogenesis: An Evaluation by Mediation Analysis. *Cancer Res* **80**, 1601-1610 (2020).
- 493 19. Chen, C.C. *et al.* Interaction between the functional polymorphisms of the alcohol-metabolism
494 genes in protection against alcoholism. *Am J Hum Genet* **65**, 795-807 (1999).
- 495 20. Matsuo, K. *et al.* Alcohol dehydrogenase 2 His47Arg polymorphism influences drinking habit
496 independently of aldehyde dehydrogenase 2 Glu487Lys polymorphism: analysis of 2,299
497 Japanese subjects. *Cancer Epidemiol Biomarkers Prev* **15**, 1009-13 (2006).
- 498 21. Lai, C.L. *et al.* Dominance of the inactive Asian variant over activity and protein contents of
499 mitochondrial aldehyde dehydrogenase 2 in human liver. *Alcohol Clin Exp Res* **38**, 44-50
500 (2014).
- 501 22. Wang, X., Sheikh, S., Saigal, D., Robinson, L. & Weiner, H. Heterotetramers of human liver
502 mitochondrial (class 2) aldehyde dehydrogenase expressed in *Escherichia coli*. A model to
503 study the heterotetramers expected to be found in Oriental people. *J Biol Chem* **271**, 31172-8
504 (1996).
- 505 23. Chang, J.S., Hsiao, J.R. & Chen, C.H. ALDH2 polymorphism and alcohol-related cancers in
506 Asians: a public health perspective. *J Biomed Sci* **24**, 19 (2017).
- 507 24. Koyanagi, Y.N. *et al.* Development of a prediction model and estimation of cumulative risk for
508 upper aerodigestive tract cancer on the basis of the aldehyde dehydrogenase 2 genotype and
509 alcohol consumption in a Japanese population. *Eur J Cancer Prev* **26**, 38-47 (2017).
- 510 25. Hamajima, N. *et al.* Gene-environment Interactions and Polymorphism Studies of Cancer Risk
511 in the Hospital-based Epidemiologic Research Program at Aichi Cancer Center II (HERPACC-
512 II). *Asian Pac J Cancer Prev* **2**, 99-107 (2001).
- 513 26. Hamajima, N. & Group, J.M.S. The Japan Multi-Institutional Collaborative Cohort Study (J-
514 MICC Study) to detect gene-environment interactions for cancer. *Asian Pac J Cancer Prev* **8**,
515 317-23 (2007).
- 516 27. Wakai, K. *et al.* Profile of participants and genotype distributions of 108 polymorphisms in a
517 cross-sectional study of associations of genotypes with lifestyle and clinical factors: a project
518 in the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study. *J Epidemiol* **21**, 223-35
519 (2011).
- 520 28. Tsugane, S. & Sawada, N. The JPHC study: design and some findings on the typical Japanese
521 diet. *Jpn J Clin Oncol* **44**, 777-82 (2014).
- 522 29. Hozawa, A. *et al.* Study Profile of the Tohoku Medical Megabank Community-Based Cohort
523 Study. *J Epidemiol* **31**, 65-76 (2021).

- 524 30. Funada, S. *et al.* Cross-Sectional Epidemiological Analysis of the Nagahama Study for
525 Correlates of Overactive Bladder: Genetic and Environmental Considerations. *J Urol* **199**, 774-
526 778 (2018).
- 527 31. Luczak, S.E., Glatt, S.J. & Wall, T.L. Meta-analyses of ALDH2 and ADH1B with alcohol
528 dependence in Asians. *Psychol Bull* **132**, 607-21 (2006).
- 529 32. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis
530 of genome-wide association studies. *Am J Hum Genet* **88**, 586-98 (2011).
- 531 33. Gelernter, J. *et al.* Genome-wide association study of alcohol dependence: significant findings
532 in African- and European-Americans including novel risk loci. *Mol Psychiatry* **19**, 41-9 (2014).
- 533 34. Zuo, L. *et al.* NKAIN1-SERINC2 is a functional, replicable and genome-wide significant risk
534 gene region specific for alcohol dependence in subjects of European descent. *Drug Alcohol*
535 *Depend* **129**, 254-64 (2013).
- 536 35. Agrawal, A., Freedman, N.D. & Bierut, L.J. Genome-wide association studies of alcohol
537 intake--a promising cocktail? *Am J Clin Nutr* **93**, 681-3 (2011).
- 538 36. Husemoen, L.L. *et al.* The association of ADH and ALDH gene variants with alcohol drinking
539 habits and cardiovascular disease risk factors. *Alcohol Clin Exp Res* **32**, 1984-91 (2008).
- 540 37. Linneberg, A. *et al.* Genetic determinants of both ethanol and acetaldehyde metabolism
541 influence alcohol hypersensitivity and drinking behaviour among Scandinavians. *Clin Exp*
542 *Allergy* **40**, 123-30 (2010).
- 543 38. Way, M.J., Ali, M.A., McQuillin, A. & Morgan, M.Y. Genetic variants in ALDH1B1 and
544 alcohol dependence risk in a British and Irish population: A bioinformatic and genetic study.
545 *PLoS One* **12**, e0177009 (2017).
- 546 39. Jackson, B.C., Reigan, P., Miller, B., Thompson, D.C. & Vasiliou, V. Human ALDH1B1
547 polymorphisms may affect the metabolism of acetaldehyde and all-trans retinaldehyde--in vitro
548 studies and computational modeling. *Pharm Res* **32**, 1648-62 (2015).
- 549 40. Larson, H.N., Weiner, H. & Hurley, T.D. Disruption of the coenzyme binding site and dimer
550 interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase "Asian"
551 variant. *J Biol Chem* **280**, 30550-6 (2005).
- 552 41. Oota, H. *et al.* The evolution and population genetics of the ALDH2 locus: random genetic
553 drift, selection, and low levels of recombination. *Ann Hum Genet* **68**, 93-109 (2004).
- 554 42. Okada, Y. *et al.* Deep whole-genome sequencing reveals recent selection signatures linked to
555 evolution and disease risk of Japanese. *Nat Commun* **9**, 1631 (2018).
- 556 43. Stagos, D. *et al.* Aldehyde dehydrogenase 1B1: molecular cloning and characterization of a
557 novel mitochondrial acetaldehyde-metabolizing enzyme. *Drug Metab Dispos* **38**, 1679-87
558 (2010).
- 559 44. Jackson, B.C. *et al.* Comparative genomics, molecular evolution and computational modeling
560 of ALDH1B1 and ALDH2. *Chem Biol Interact* **202**, 11-21 (2013).
- 561 45. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell

- 562 types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
- 563 46. Zhou, S.L. *et al.* Ethanol up-regulates fatty acid uptake and plasma membrane expression and
564 export of mitochondrial aspartate aminotransferase in HepG2 cells. *Hepatology* **27**, 1064-74
565 (1998).
- 566 47. Takeshita, T., Yang, X. & Morimoto, K. The ALDH2 genotype, alcohol intake, and liver-
567 function biomarkers among Japanese male workers. *Hum Genet* **106**, 589-93 (2000).
- 568 48. Nagata, N., Watanabe, N., Tsuda, M., Tsukamoto, H. & Matsuzaki, S. Relationship between
569 serum levels of anti-low-density lipoprotein-acetaldehyde-adduct antibody and aldehyde
570 dehydrogenase 2 heterozygotes in patients with alcoholic liver injury. *Alcohol Clin Exp Res* **23**,
571 24S-28S (1999).
- 572 49. Kwon, H.J. *et al.* Aldehyde dehydrogenase 2 deficiency ameliorates alcoholic fatty liver but
573 worsens liver inflammation and fibrosis in mice. *Hepatology* **60**, 146-57 (2014).
- 574 50. Engs, R.C., Hanson, D.J., Gliksman, L. & Smythe, C. Influence of religion and culture on
575 drinking behaviours: a test of hypotheses between Canada and the USA. *Br J Addict* **85**, 1475-
576 82 (1990).
- 577 51. Dingler, F.A. *et al.* Two Aldehyde Clearance Systems Are Essential to Prevent Lethal
578 Formaldehyde Accumulation in Mice and Humans. *Mol Cell* **80**, 996-1012 e9 (2020).
- 579 52. Suzuki, S. *et al.* Body mass index and colorectal cancer risk: A Mendelian randomization study.
580 *Cancer Sci* **112**, 1579-1588 (2021).
- 581 53. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74
582 (2015).
- 583 54. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease
584 and population genetic studies. *Nat Methods* **10**, 5-6 (2013).
- 585 55. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-
586 7 (2016).
- 587 56. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method
588 for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
- 589 57. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for
590 genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
- 591 58. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
592 datasets. *Gigascience* **4**, 7 (2015).
- 593 59. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004
594 (1999).
- 595 60. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide
596 association scans. *Bioinformatics* **26**, 2190-1 (2010).
- 597 61. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from
598 high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
- 599 62. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human

- 600 tissues. *Science* **369**, 1318-1330 (2020).
- 601 63. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and
602 haplotype maps. *Bioinformatics* **21**, 263-5 (2005).
- 603
- 604

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

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

641

642 **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

647 **Figure legends**

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

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

654

655 **Figure 2. Manhattan plots of the GWAS of drinking status**

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

661

662 **Figure 3. Genomic loci reaching genome-wide significance in either analysis for association**

663 **with daily alcohol intake**

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

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

669

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

672 Direction of effects of identified variants other than rs671 is presented as a heatmap (with colors
673 indicating associated normalized ORs). Estimates with a single asterisk show suggestive significance
674 ($P < 5.0 \times 10^{-6}$). Estimates with double asterisks show genome-wide significance ($P < 5.0 \times 10^{-8}$).

675 Lead SNP in each locus is highlighted with its estimates in bold. SNP, single nucleotide
676 polymorphism; Ref, reference allele; Alt, alternative allele; OR, odds ratio; 95% CI, 95% confidence
677 interval; Het P , P value from test of heterogeneity.

678

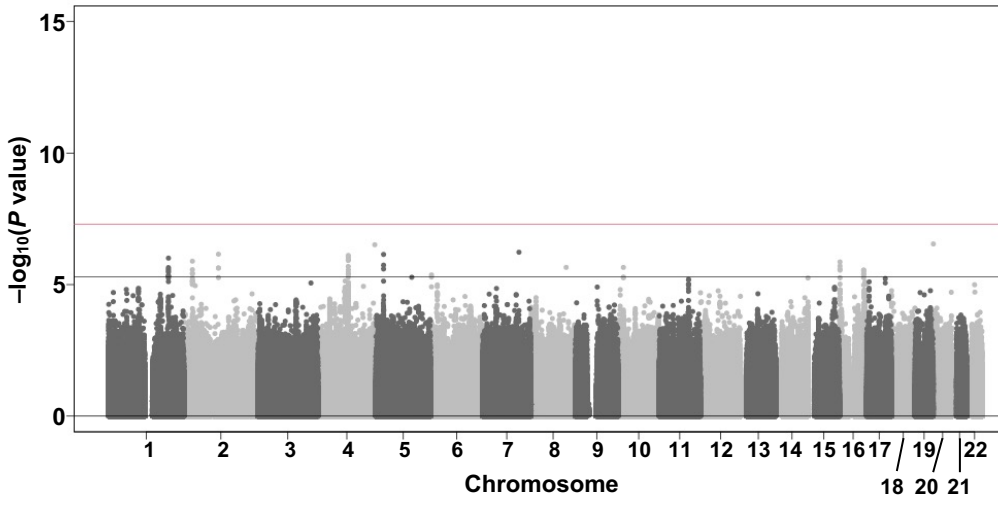
679 **Figure 5. Regional association plots of the identified novel regions**

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

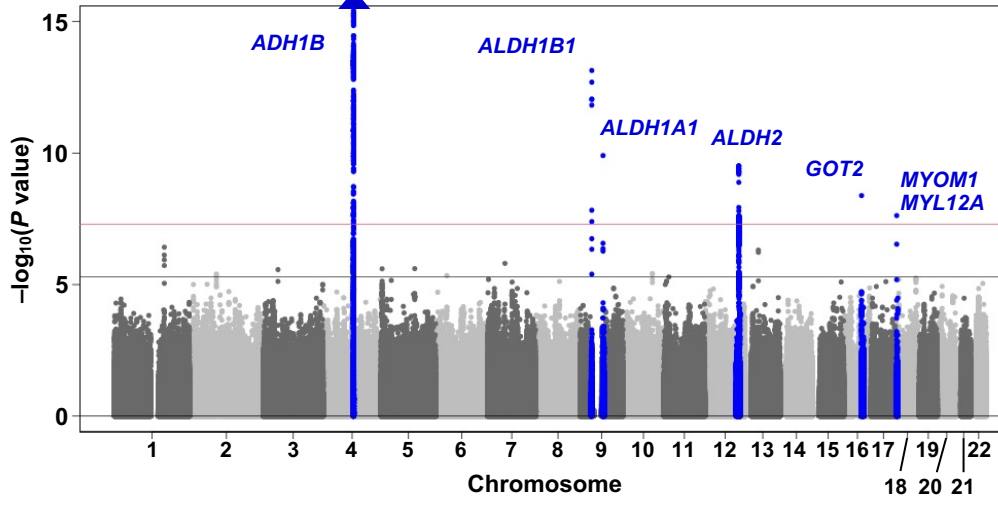
Figure 1. Manhattan plots of the GWAS of daily alcohol intake

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 $<1 \times 10^{-15}$.

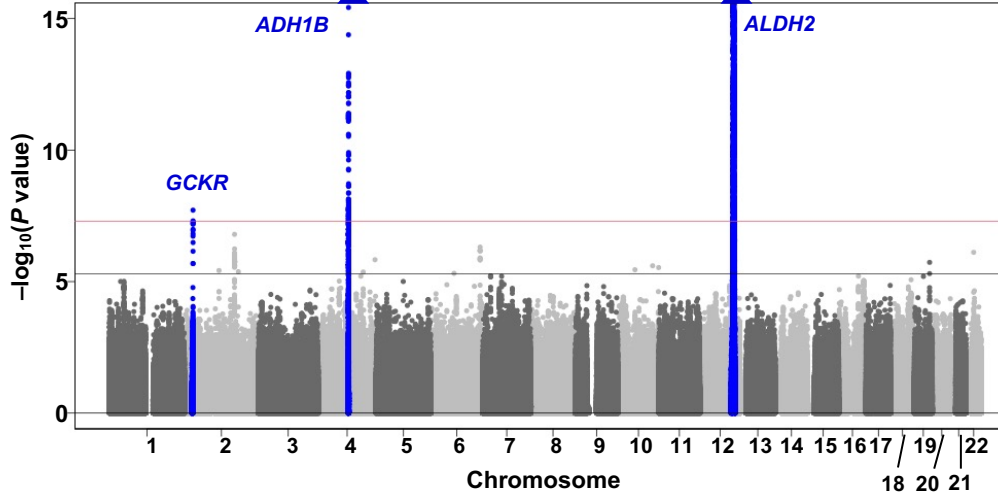
(a) rs671: GG



(b) rs671: GA



(c) Unstratified



(d) Interaction

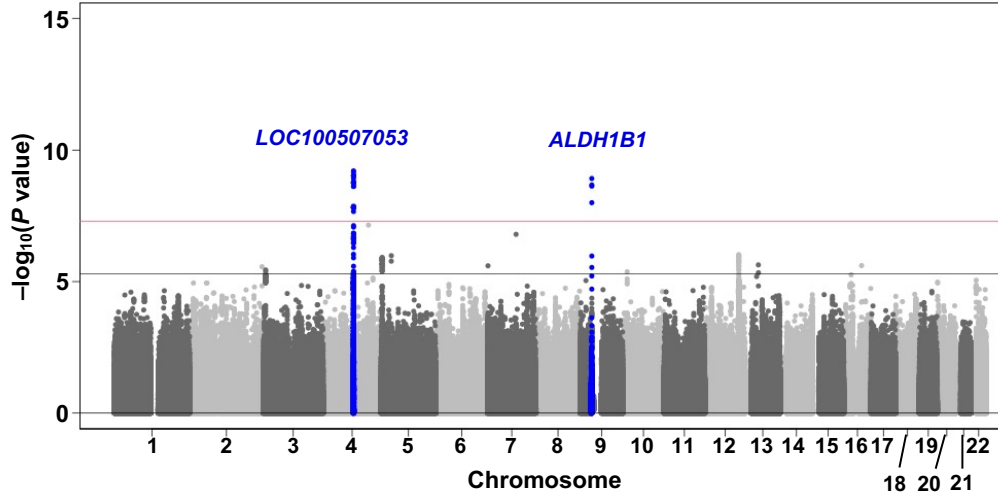
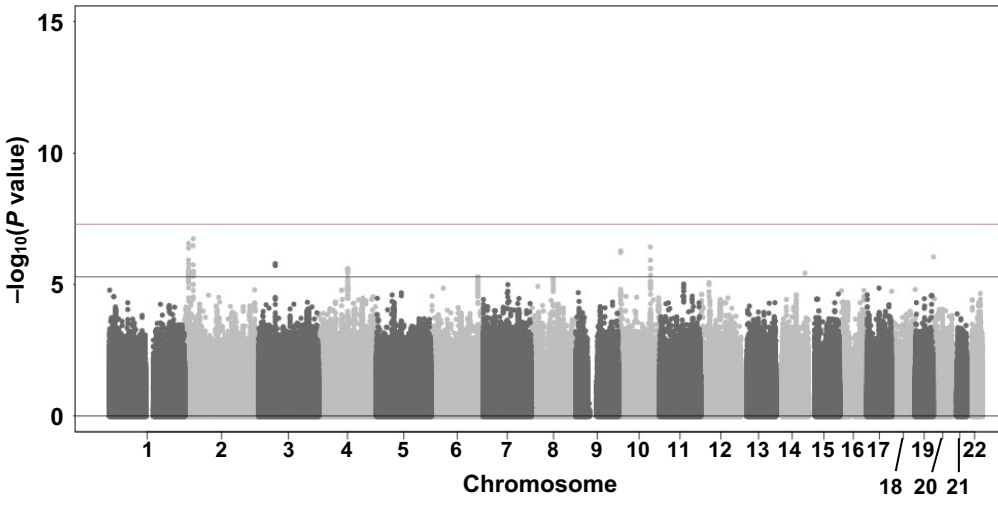


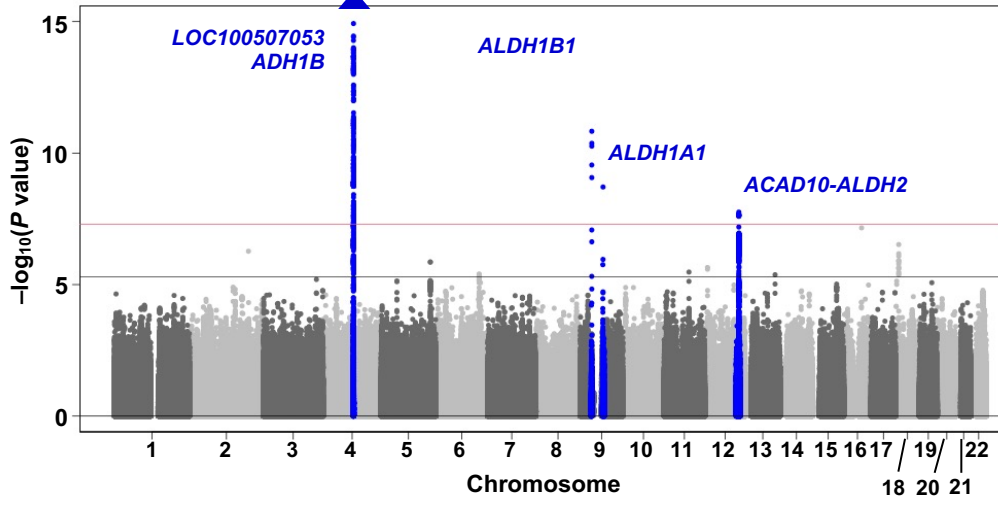
Figure 2. Manhattan plots of the GWAS of drinking status

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 $<1 \times 10^{-15}$.

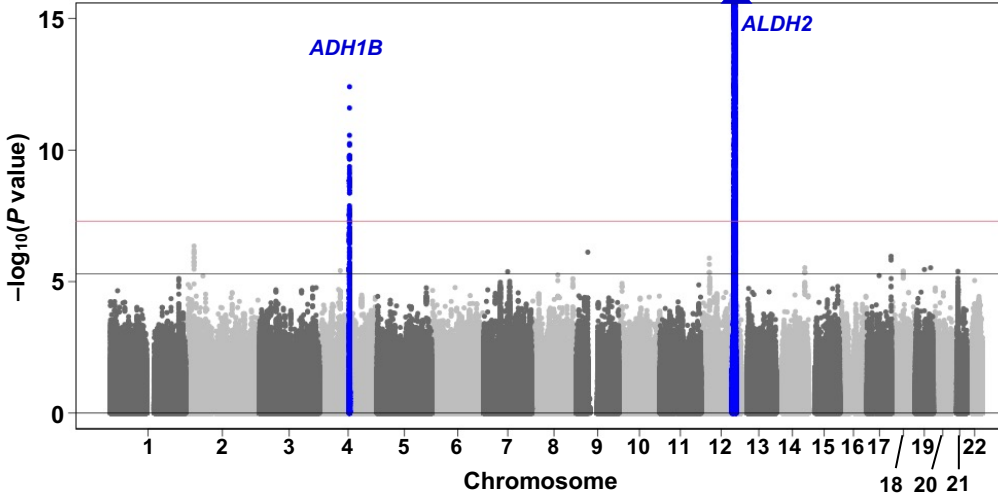
(a) rs671: GG



(b) rs671: GA



(c) Unstratified



(d) Interaction

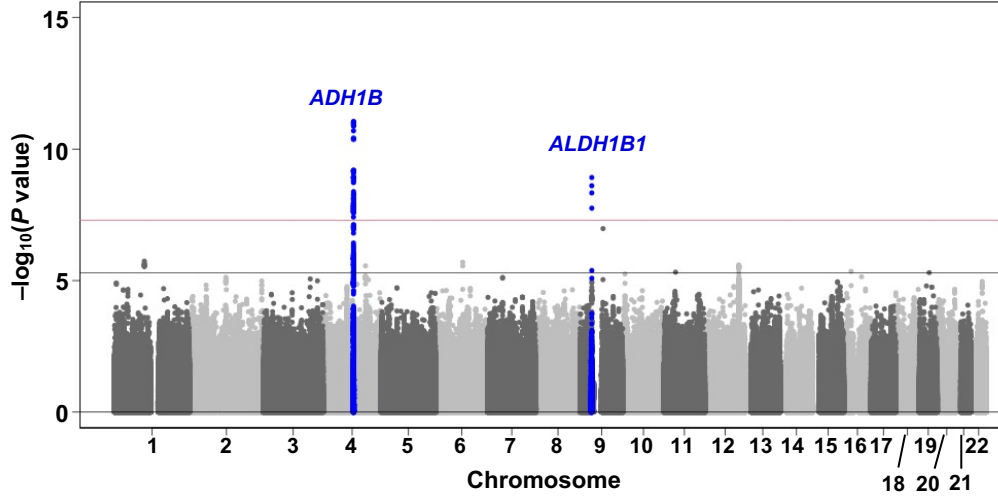


Figure 3. Genomic loci reaching genome-wide significance in either analysis for association with daily alcohol intake
 Direction of effects of identified variants other than rs671 is presented as a heatmap. Estimates with a single asterisk show suggestive significance ($P < 5.0 \times 10^{-6}$). Estimates 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; HetP, P value from test of heterogeneity.

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

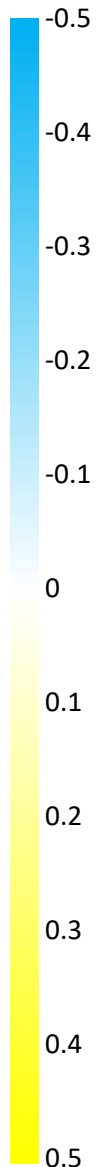
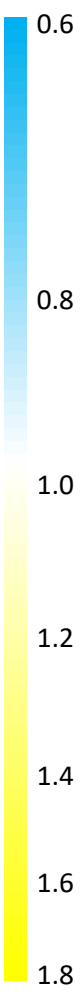


Figure 4. Genomic loci reaching genome-wide significance in either analysis for association with drinking status

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; HetP, P value from test of heterogeneity.

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



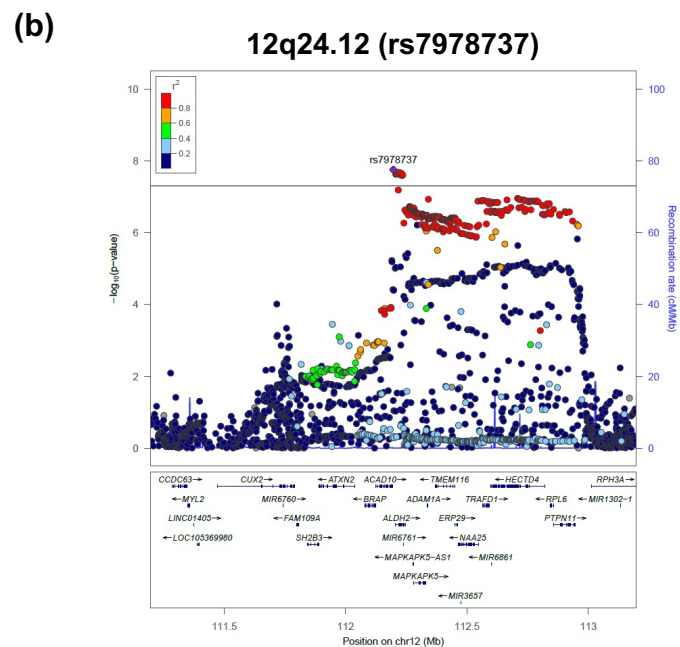
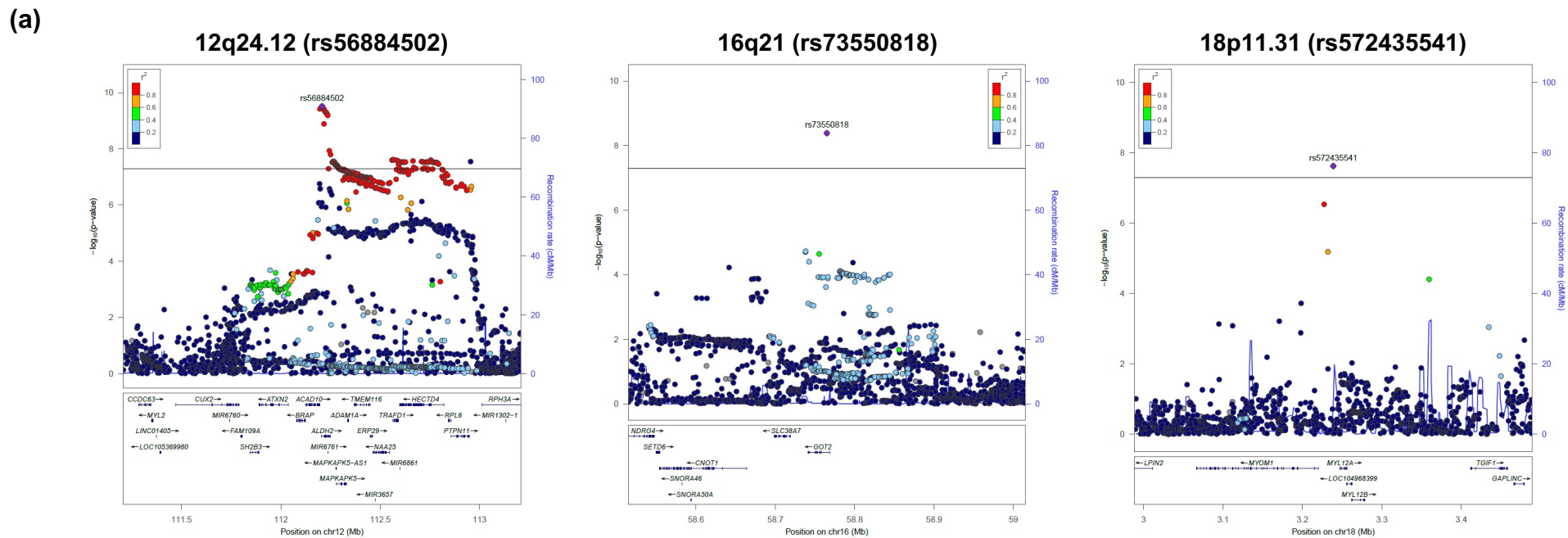


Figure 5. Regional association plots of the identified novel regions Regional association plots for (a) daily alcohol intake and (b) drinking status in rs671 heterozygotes are shown. The vertical axis indicates the $-\log_{10}(P \text{ value})$ for the assessment of the association of each SNP with daily 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.