



OPEN

CUX2, *BRAP* and *ALDH2* are associated with metabolic traits in people with excessive alcohol consumption

I-Chun Chen^{1,2,3}, Po-Hsiu Kuo^{4,5}, Albert C. Yang^{6,7}, Shih-Jen Tsai^{7,8,9}, Tung-Hsia Liu¹⁰, Hung-Jen Liu^{3,11}✉, Tsuo-Hung Lan^{1,12}, Hong-Ming Chen¹³, Huang-Nan Huang¹³, Ren-Hua Chung¹⁴ & Yu-Li Liu^{10,15}✉

Molecular mechanisms that prompt or mitigate excessive alcohol consumption could be partly explained by metabolic shifts. This genome-wide association study aims to identify the susceptibility gene loci for excessive alcohol consumption by jointly measuring weekly alcohol consumption and γ -GT levels. We analysed the Taiwan Biobank data of 18,363 Taiwanese people, including 1945 with excessive alcohol use. We found that one or two copies of the G allele in rs671 (*ALDH2*) increased the risk of excessive alcohol consumption, while one or two copies of the C allele in rs3782886 (*BRAP*) reduced the risk of excessive alcohol consumption. To minimize the influence of extensive regional linkage disequilibrium, we used the ridge regression. The ridge coefficients of rs7398833, rs671 and rs3782886 were unchanged across different values of the shrinkage parameter. The three variants corresponded to posttranscriptional activity, including cut-like homeobox 2 (a protein coded by *CUX2*), Glu504Lys of acetaldehyde dehydrogenase 2 (a protein encoded by *ALDH2*) and Glu4Gly of BRCA1-associated protein (a protein encoded by *BRAP*). We found that Glu504Lys of *ALDH2* and Glu4Gly of *BRAP* are involved in the negative regulation of excessive alcohol consumption. The mechanism underlying the γ -GT-catalytic metabolic reaction in excessive alcohol consumption is associated with *ALDH2*, *BRAP* and *CUX2*. Further study is needed to clarify the roles of *ALDH2*, *BRAP* and *CUX2* in the liver–brain endocrine axis connecting metabolic shifts with excessive alcohol consumption.

The recommended level of low-risk alcohol consumption is < 100 g/week¹. Phenotypes of excessive alcohol consumption are expressed in several forms. Before development of alcohol use disorder, the condition, for example, may initiate with problematic drinking, which has a 2.1% prevalence in Asian countries^{2,3}. Excessive alcohol consumption creates a medical and social burden and is associated with alcohol-related liver diseases, public safety incidents, and trauma-related admissions to hospitals.

¹Department of Psychiatry, Taichung Veterans General Hospital, Taichung City, Taiwan. ²Ph.D. Program in Translational Medicine, National Chung Hsing University, Taichung City, Taiwan. ³Rong Hsing Research Center For Translational Medicine, National Chung Hsing University, Taichung City, Taiwan. ⁴Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan. ⁵Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan. ⁶Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. ⁷Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan. ⁸Division of Psychiatry, National Yang-Ming University, Taipei, Taiwan. ⁹Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan. ¹⁰Center for Neuropsychiatric Research, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 35053, Taiwan. ¹¹Institute of Molecular Biology, National Chung Hsing University, 145 Xingda Road, South District, Taichung City 402, Taiwan. ¹²Faculty of Medicine, National Yang-Ming University, Taipei, Taiwan. ¹³Department of Applied Mathematics, Tunghai University, Taichung City, Taiwan. ¹⁴Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taipei, Taiwan. ¹⁵Graduate Institute of Clinical Medical Science, China Medical University, Taichung City, Taiwan. ✉email: hjliu5257@nchu.edu.tw; ylliu@nhri.org.tw

The genetic architecture of alcohol consumption involves the genetic liability of alcohol use disorder, metabolism, risky behaviors and cognitive phenotypes⁴. The alcohol dehydrogenase 1B (*ADH1B*), alcohol-metabolizing acetaldehyde dehydrogenase 2 (*ALDH2*), β -Klotho (*KLB*), glucokinase regulator (*GCKR*), corticotropin releasing hormone receptor 1 (*CRHR1*), and cell adhesion molecule 2 (*CADM2*) show strong links to drinking behaviours^{4–6}. *GCKR* is associated with both alcohol consumption and alcohol use disorder⁴. The role of dopamine receptor D2 subtype (*DRD2*) has been confirmed and replicated in a large-scale genome-wide association study (GWAS)⁷.

Genes that act in pleiotropy across various systems (e.g., cardiovascular, adrenal, pancreatic and central nervous systems) form the genetic picture of excessive alcohol consumption. The alcohol-decreasing allele in *ADH1B* gene was associated with lower odds of coronary heart disease, and those SNPs significantly associated with alcohol consumption were associated with high-density lipoprotein cholesterol levels⁸. The largest study of GWAS on tobacco and alcohol uses involved 1.2 million individuals and uncovered genetic bearing of *ADH1B* and *GCKR*, suggesting that alcohol consumption is influenced by individual differences in one's ability to process calorie-rich alcoholic beverages⁹. Studies have replicated the *KLB/FGF21* interaction in the putative liver-brain axis^{10,11}; and notably, neuronal FGF21 senses metabolic changes in the peripheral tissues, resulting in homeostatic regulation of the liver-brain axis¹².

Alcohol is chemically bound to hydrophobic amino acids and hydrogen-bonding polar groups of channel proteins¹³, which drive “go pathways” and “stop pathways” in the intracellular level. The “go pathways” are signaling cascades that contribute to the transition from moderate to excessive alcohol intake, including activation of protein kinase A (PKA) and calcium/calmodulin-dependent protein kinase II, whereas the “stop pathways” keep alcohol intake in check, by upregulation of brain-derived neurotrophic factor (BDNF) and glial cell line-derived neurotrophic factor (GDNF)¹⁴.

Alcohol Use Disorder Identification Test (AUDIT) makes specific quantitative statements about alcohol consumption versus alcohol use disorder, while Alcohol Use Disorder Identification Test-Consumption (AUDIT-C) measure alcohol consumption. The AUDIT can be applied as a proxy measurement to increase sample sizes in a GWAS on alcohol use disorder¹⁵. Physiological biomarkers may be used to identify persistent and excessive alcohol consumption. The ethanol intake of excessive drinkers is reflected in the ratio of carbohydrate transferrin to transferrin¹⁶. The extent of elevated aminotransferase levels in the body is also helpful in detecting alcohol abuse¹⁷. Asymptomatic patients with alcoholic liver disease present serum levels of γ -glutamyl transpeptidase (γ -GT) doubling that of normal¹⁷. The γ -GT, which is involved in the metabolism of glutathione, is a major antioxidant in humans, and it is also a common biomarker used in studying alcohol use disorder^{18,19}.

It has been suggested in a large-scale GWAS on alcohol consumption⁴ and also other conditions, that to control the effect of population stratification, one may analyze participants according to races and ethnicities. The population of Taiwan comprises 92.6% southern Han Chinese, 4.9% northern Han Chinese, and 1.9% aborigines of Southeast Asian and Austronesian descent²⁰. Genetic intermixing between these ethnic groups is rare, resulting in a genetically homogenous Taiwanese population of mostly Han Chinese descent. Our study aims to infer the susceptibility gene loci of excessive alcohol consumption by jointly measuring weekly alcohol consumption and γ -GT levels.

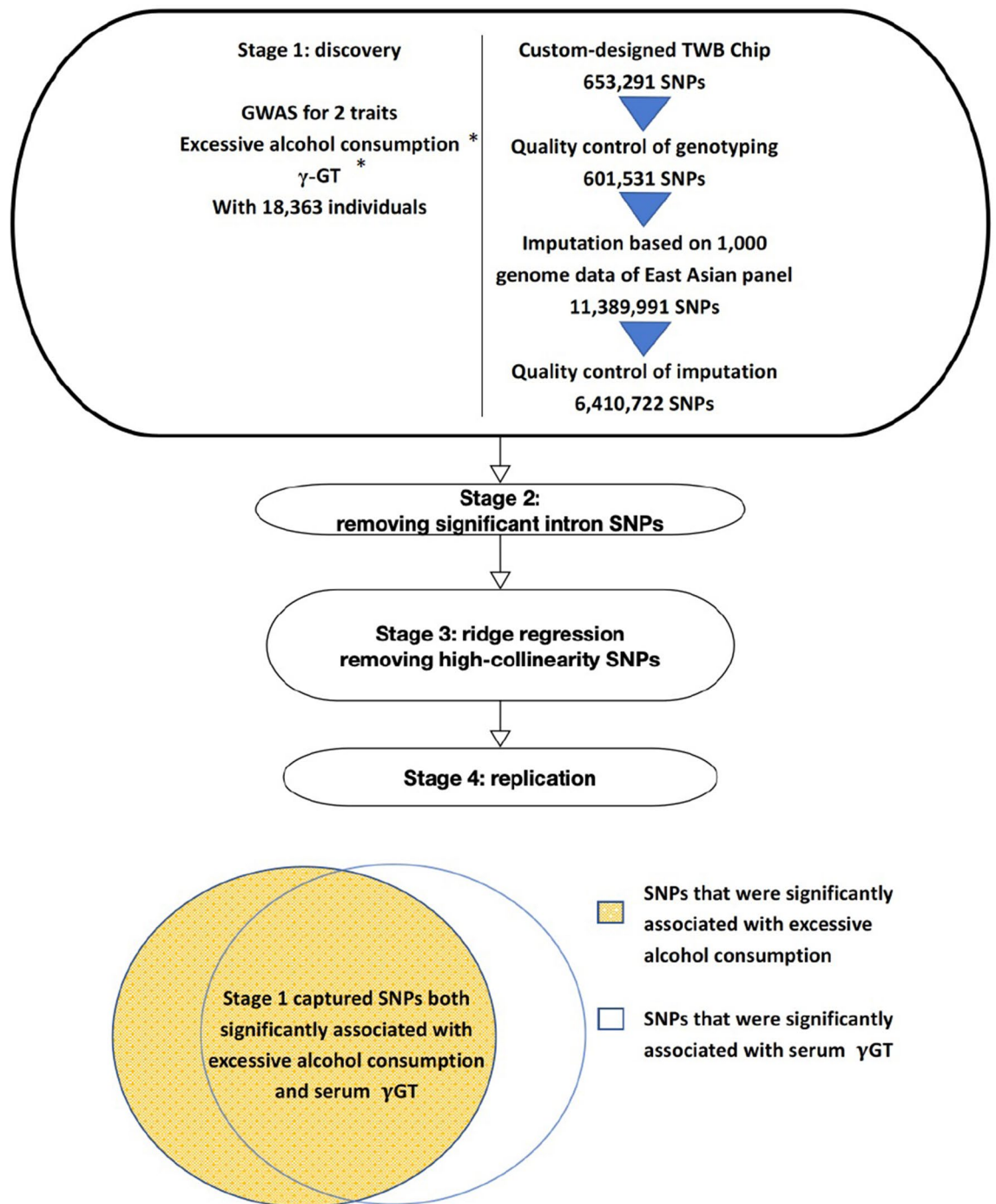
Results

We retrieved data on the whole-genome genotyping, and also the levels of serum γ -GT and medical history of the 18,363 people whose information had been held in the TWB. Of the participants, 9275 were women. “Excessive alcohol consumption” was defined as a weekly intake of alcoholic beverages with an equivalent of > 150 mL of alcohol for ≥ 6 months. To identify the phenotype of excessive alcohol consumption, we used serum γ -GT as an add-on trait (Fig. 1).

To plot the genetic ancestry of our cohort from Taiwan Biobank (TWB), we used principal component analysis (PCA), and results confirmed a reliable distribution (see Supplementary Fig. S1 online). In this cohort, 1945 participants (10.60%; 87.9% men) had excessive alcohol use (weekly intake of > 150 mL of alcohol for ≥ 6 months) (Table 1, see Supplementary Fig. S2 online). The average serum γ -GT level of those with excessive alcohol use was 46.15 ± 77.08 U/L, higher than those without (23.60 ± 25.71 U/L). Among excessive alcohol users, a significant correlation was found between alcohol consumption and serum γ -GT levels ($p < 1 \times 10^{-3}$).

There were 1794 SNPs significantly associated with excessive alcohol use ($p < 5 \times 10^{-8}$) (see Supplementary Table S1 online). The COJO analysis of GCTA suggests that there were 3 independent signals among these SNPs. LocusZoom plots for the 3 SNPs are shown in Supplementary Fig. S3 online. The plot of log quantile–quantile (Q-Q) p values suggested only a few systematic sources of spurious associations (Fig. 2). Because the QQ plot contains a wider range of the observed $-\log_{10} p$ values, we further applied LD Score regression (LDSC) to analyze polygenicity and other factors²¹. The estimated LDSC intercept was 1.0083 with a standard error of 0.0056. Furthermore, the genomic inflation factor (λ_{GC}) was also reported by LDSC. The value of λ_{GC} was estimated to be 1.0043. Both LDSC intercept and λ_{GC} are close to 1, suggesting no inflation had occurred in our analysis due to confounding factors. The inflation observed in the QQ plot could be driven by a few causal signals as suggested by the COJO analysis and the LocusZoom plots, while many SNPs close to the causal signals are in high linkage disequilibrium (LD) with the causal SNPs. Those SNPs observed corresponded to small p values most likely mapped to susceptibility risk loci for excessive alcohol use (Table 2).

We captured 1015 SNPs that were both significantly associated with excessive alcohol use and with γ -GT (FDR < 0.05) (Fig. 1; Table 2). These 1015 significant SNPs aggregated on chromosome 12 (Fig. 3). They were identified within a region of approximately 3.7 million bases located between *TRPV4* and *SDS* (chr12: 110238596–113944048) (see Supplementary Table S1 online). The strongest signal appeared at rs671 (*ALDH2*), where the codon change from the G allele to the A allele creates a missense variant and represents the translation



* Excessive alcohol consumption is defined as a weekly intake of alcoholic beverages with an equivalent of >150 mL of alcohol for longer than 6 months. Serum γ -GT is continuous data.

Figure 1. Overall study scheme. GWAS: genome-wide association study, TWB: Taiwan Biobank, SNP: single nucleotide polymorphism.

from glutamic acid to lysine in the sequence (see Supplementary Table S2 online). The SNP rs671 is in strong LD with rs4646776 (LD $r^2 = 0.998$), one of the three independent SNPs identified by the COJO analysis, suggesting that they are in the same LD block.

To find the possible causal variants of excessive alcohol consumption within this region, we further identified their coding variants. We used coding-synonymous SNPs, 5' untranslated region SNPs, missense SNPs, non-coding RNA elements in the 3' untranslated regions, cds-indels, and frameshift mutations to obtain 48 significant

	Excessive alcohol consumption ^a (N = 1945)		No excessive alcohol consumption (N = 16,405)	
	N; MEAN (S.D.)	N (%)	N; mean (S.D.)	N (%)
γ -GT ^b (U/L)	1158; 46.15 (77.08) ^c		9663; 23.60 (25.71) ^c	
Age	1945; 49.55 (10.31)		16,405; 48.72 (11.07)	
Sex				
Male		1710 (87.92%)		7369 (44.92%)
Female		235 (12.08%)		9036 (55.08%)

Table 1. γ -GT, age, and sex for the two groups differentiated by their alcohol consumption. S.D.: Standard deviation. ^aExcessive alcohol consumption is defined as a weekly intake > 150 mL of alcohol for > 6 months. ^b γ -GT: γ -glutamyl transpeptidase. ^cThere is a significant difference in serum γ -GT levels between these two groups ($p < 1 \times 10^{-3}$).

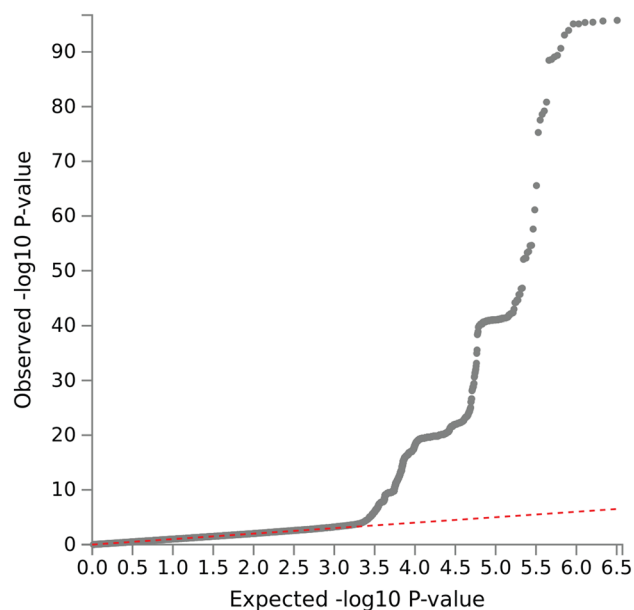


Figure 2. Q–Q plot of the SNP-based test for the drinking variable, adjusted for age, sex, and 10 PCs. Q–Q plot: quantile–quantile plot, SNP: single nucleotide polymorphism, PCs: principal components.

SNPs. Among these 48 SNPs, rs7398833 (*CUX2*), rs671 (*ALDH2*) and rs3782886 (*BRAP*) had ridge coefficients, $\hat{\beta}^{ridge}$, that remained unchanged across different values of the shrinkage parameter λ (Fig. 4). The rs7398833 (*CUX2*) is located in the 3' untranslated region (3'-UTR), where it post-translationally manipulates the stability of *CUX2*. The coding change from T to C allele at rs3782886 (*BRAP*) creates a missense variant, which leads to a coding change from glutamic acid to glycine in the translation of BRCA1-associated protein isoform 4.

In our participants, a strong LD was found between rs671 and rs3782886 ($r^2 = 0.98$) (see Supplementary Fig. S4 online). A significant haplotype was therefore associated with excessive alcohol consumption and it was comprised of both rs671 and rs3782886. The presence of a haplotype carrying the G allele of rs671 and T allele of rs3782886 (haplotype GT) showed an odds ratio (OR) of 2.49 (95% confidence interval CI 2.27–2.72) for excessive alcohol consumption, whereas a haplotype carrying A allele of rs671 and C allele of rs3782886 (haplotype AC) had an odds ratio (OR) of 0.4 (95% CI 0.37–0.44). Comparing levels of γ -GT between carriers with haplotype GT and those with haplotype AC, we found a differential increment of 2.42 ± 0.53 U/L ($p = 4.92 \times 10^{-6}$).

We performed conditional analysis to identify independent signals between rs671 (*ALDH2*) and rs3782886 (*BRAP*). We compared using a partial F-test, three models each with rs671 (*ALDH2*), with rs3782886 (*BRAP*), then with both rs671 (*ALDH2*) and rs3782886 (*BRAP*). The regression coefficients estimated were: rs671 (*ALDH2*), rs3782886 (*BRAP*) and rs7398833 (*CUX2*) were estimated as 3.54 (95% CI 1.06, 6.02) for model rs671 (*ALDH2*), -1.98 (95% CI $-3.38, -1.59$) for model rs3782886 (*BRAP*) and -3.64 (95% CI $-13.51, 6.24$) for model rs7398833 (*CUX2*). Regarding the direction of effects, one or two copies of G allele in rs671 (*ALDH2*) increased the risk of excessive alcohol consumption, while one or two copies of C allele in rs3782886 (*BRAP*) reduced the risk of excessive alcohol consumption. We found that the model that included both rs671 (*ALDH2*) and rs3782886 (*BRAP*) was significantly better with a significantly lower sum of squared error ($p < 0.01$) (see Supplementary Table S3 online).

Gene	Start	End	Minimum <i>p</i> value of drinking	Number of overlapping SNPs
<i>ALDH2</i>	112204691	112247782	1.73E-96	20
<i>ACAD10</i>	112123857	112194903	3.78E-96	31
<i>BRAP</i>	112079950	112123790	7.68E-96	24
<i>HECTD4</i>	112597992	112819896	1.19E-94	185
<i>PTPN11</i>	112856155	112947717	8.45E-94	11
<i>NAA25</i>	112464500	112546826	2.27E-91	71
<i>TRAFD1</i>	112563305	112591407	8.44E-90	16
<i>RPH3A</i>	113008184	113336686	6.35E-80	188
<i>RPL6</i>	112842994	112856642	5.46E-76	2
<i>CUX2</i>	111471828	111788358	2.76E-66	39
<i>MYL2</i>	111348623	111358526	1.05E-43	30
<i>CCDC63</i>	111284573	111345339	7.33E-43	65
<i>ATXN2</i>	111890018	112037480	2.60E-39	86
<i>TMEM116</i>	112369086	112450970	1.35E-29	117
<i>MAPKAPK5</i>	112279782	112334343	1.36E-25	100
<i>SH2B3</i>	111843752	111889427	9.87E-25	18
<i>ERP29</i>	112451120	112461255	3.54E-20	9
<i>IFT81</i>	110562140	110656602	2.47E-19	3

Table 2. Loci significantly associated with both excessive alcohol consumption and γ -GT. Adjusted for age, sex, and 10 principal components. *ALDH2*: aldehyde dehydrogenase 2, *ACAD10*: acyl-coenzyme A dehydrogenase family, member 10, *BRAP*: breast cancer 1-associated protein, *HECTD4*: HECT domain E3 ubiquitin protein ligase 4, *PTPN11*: protein tyrosine phosphatase non-receptor type 11, *NAA25*: N(alpha)-acetyltransferase 25, *TRAFD1*: TRAF-type zinc finger domain containing 1, *RPH3A*: rabphilin 3A, *RPL6*: ribosomal protein L6, *CUX2*: cut-like homeobox 2, *MYL2*: myosin, light polypeptide 2, *ATXN2*: ataxin 2, *CCDC63*: coiled-coil domain containing 63, *TMEM116*: transmembrane protein 116, *MAPKAPK5*: MAP kinase activated protein kinase 5, *SH2B3*: SH2B adaptor protein 3, *ERP29*: endoplasmic reticulum protein 29, *IFT81*: intraflagellar transport 81.

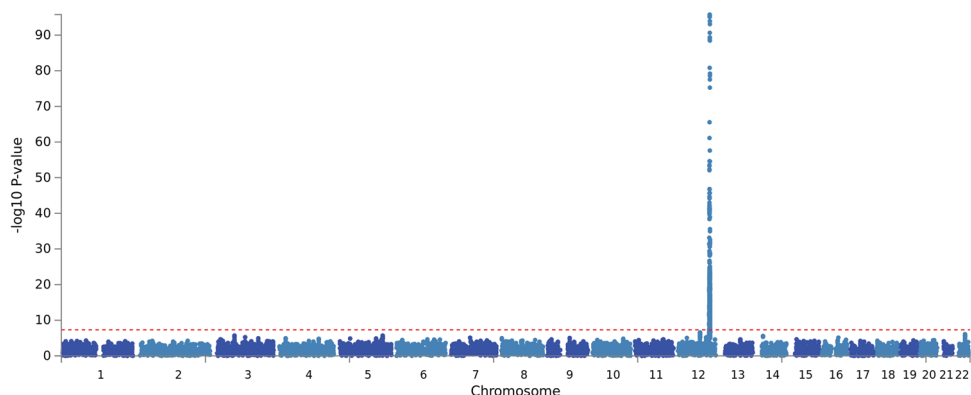


Figure 3. Manhattan plot of SNP-based test for the drinking variable, adjusted for age, sex, and 10 PCs. SNP: single nucleotide polymorphism, PCs: principal components.

We also performed gene set-based analyses using gene sets including *ALDH2*, *BRAP* and *CUX2*. The gene set-based analyses for metabolic traits among excessive alcohol drinkers generated results in Fig. 5. None of these genes showed significant fold enrichment ($FDR > 0.05$).

For access to replication study, we compared of results of this GWAS with the publicly available database on the GWAS ATLAS resource (<https://atlas.ctglab.nl/>)²², a post-GWAS fine-mapping study in individuals of Korean descent (459 with alcohol dependence, 455 controls) and a trans-population GWAS meta-analysis of AUDIT-C ($N = 274,424$)^{7,23}. A total of 45 GWASs were derived from the GWAS ATLAS resource (average sample size = 132,522). The multiple GWAS comparisons in the GWAS ATLAS resource grouped physically overlapping risk loci, and identified risk loci from 111599617 to 111705565 on chromosome 12. This region mapped to *BRAP* and *ALDH2*. The post-GWAS fine-mapping study on participants of Korean descent showed a genetic correlation between rs3782886 (*BRAP*) and alcohol dependence ($p = 9.94 \times 10^{-31}$), with the minor homozygote associating with lesser risk of alcohol consumption²³. This adheres to our finding that one or two copies of C allele in rs3782886 (*BRAP*) reduced the risk of excessive alcohol consumption. A trans-population GWAS meta-analysis of AUDIT, including 1410 cases of excessive alcohol consumption in the East Asian subgroup of the

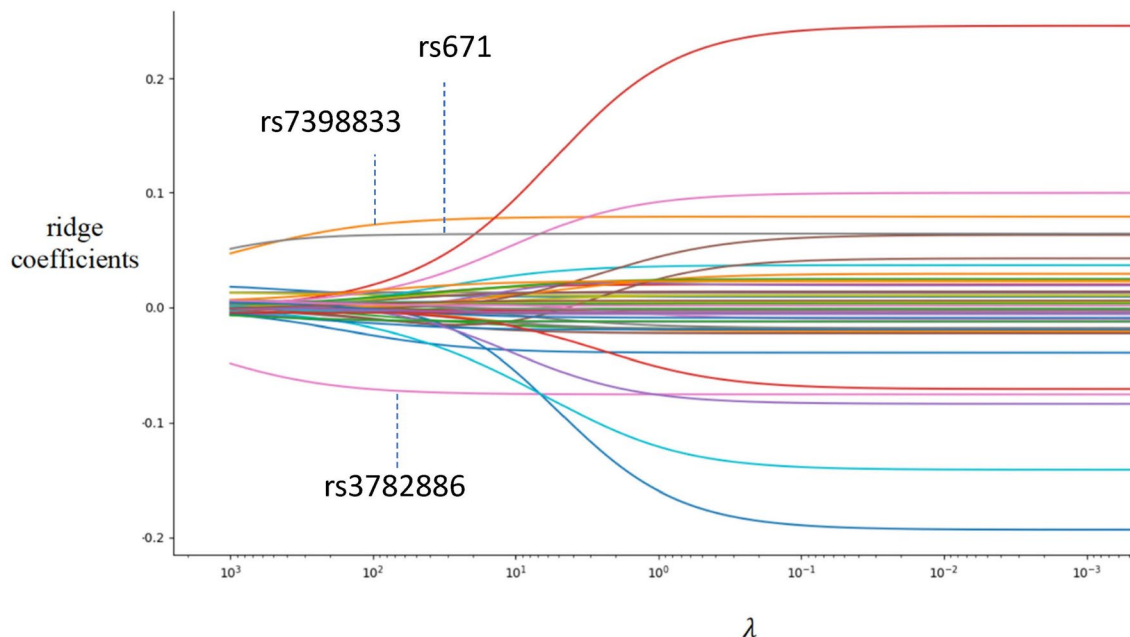


Figure 4. The X axis represents the weights, which are the ridge coefficients ($\hat{\beta}^{ridge}$) corresponding to 48 SNPs. The Y axis represents the shrinkage parameter λ , which controls the size of the coefficients and the amount of regularisation. Curves of the ridge coefficients as a function of regularisation. Note that rs7398833, rs671 and rs3782886 have $\hat{\beta}^{ridge}$ values that are maximal values away from zero and remain constant across different λ values. Those curves sharply alienated from X axis indicate dependent signals of linkage disequilibrium.

GO category	Expected number	Observed number	Fold Enrichment	Overrepresentation	p-value	FDR
GO biological process						
ethanol oxidation	12	1	>100	+	1.87E-03	>0.05
ethanol catabolic process	12	1	>100	+	1.87E-03	>0.05
short-term memory	12	1	>100	+	1.87E-03	>0.05
primary alcohol catabolic process	15	1	>100	+	2.30E-03	>0.05
positive regulation of dendritic spine morphogenesis	20	1	>100	+	3.02E-03	>0.05
ethanol metabolic process	21	1	>100	+	3.16E-03	>0.05
positive regulation of excitatory postsynaptic potential	29	1	>100	+	4.31E-03	>0.05
positive regulation of dendrite morphogenesis	38	1	>100	+	5.60E-03	>0.05
positive regulation of dendrite spine development	44	1	>100	+	6.46E-03	>0.05
modulation of excitatory spine morphogenesis	45	1	>100	+	6.60E-03	>0.05
GO molecular function						
glyceraldehyde-3-phosphate dehydrogenase(NAD+)(non-phosphorylation) activity	6	1	>100	+	1.01E-03	>0.05
aldehyde dehydrogenase(NAD+) activity	16	1	>100	+	2.44E-03	>0.05
aldehyde dehydrogenase[NAD(P)+] activity	17	1	>100	+	2.59E-03	>0.05
nuclear localization sequence binding	26	1	>100	+	3.88E-03	>0.05
oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NAC	37	1	>100	+	5.46E-03	>0.05
oxidoreductase activity, acting on the aldehyde or oxo group of donors	44	1	>100	+	6.46E-03	>0.05
signal sequence binding	48	1	>100	+	7.03E-03	>0.05
NAD binding	56	1	>100	+	8.18E-03	>0.05
electron transfer activity	110	1	63.18	+	1.59E-03	>0.05
heterocyclic compound binding	6060	3	3.44	+	2.46E-03	>0.05
GO cellular component						
extracellular exosome	2098	2	6.63	+	2.84E-02	>0.05
extracellular vesicle	2119	2	6.56	+	2.89E-02	>0.05
extracellular organelle	2124	2	6.54	+	2.90E-02	>0.05
ubiquitin ligase complex	294	1	23.64	+	4.18E-02	>0.05
nuclear membrane	302	1	23.01	+	4.30E-02	>0.05

Figure 5. Results of gene set-based analysis for metabolic traits among people with excessive alcohol consumption. First column: The Gene Ontology (GO) category. Second column: The number of genes expected in this category. Third column: The observed number of genes that map to this GO category. Forth column: Fold Enrichment is the observed number divided by the expected number. If it is greater than 1, it indicates that the category is overrepresented. Fifth column: A plus sign indicates overrepresentation of this GO category. Sixth column: Cutoff is 0.05. Seventh column: The probability that the number of genes observed in this category occurred by chance.

274,424 individuals, found a risk locus in *BRAP*⁷. We found that rs3782886 (*BRAP*) and rs671 (*ALDH2*) were associated with pleiotropy across various systems including metabolic conditions (see Supplementary Fig. S5 online). These results do not manifest a replication of the initial findings, but it suggests that both *BRAP* and *ALDH2* influence metabolic traits²².

Discussion

Our main finding is that in excessive alcohol consumption, the γ -GT-catalytic reaction is associated with *ALDH2*, *BRAP* and *CUX2*. Both the A allele in rs671 (*ALDH2*), and the C allele in rs3782886 (*BRAP*) lowered risks of excessive alcohol consumption. These gene products acted as negative regulators on excessive alcohol consumption.

Our GWAS has several strengths. First, we developed a new approach for evaluating intermarker linkage disequilibrium. Conducting ridge regression led to the identification of significant SNPs. For complex traits like excessive alcohol consumption, strategies to elucidate polygenicity should be considered. Our strategy to tackle the polygenicity and linkage disequilibrium is the use of ridge regression, which has proven to efficiently identify genetic markers of complex genetic disorders^{24–26}. Like linkage disequilibrium score regression, ridge regression can help resolve SNPs in strong linkage disequilibrium²⁴. Second, we captured SNPs that are significantly associated with both excessive alcohol use and γ -GT. Diagnostic bias was reduced by exhibiting persistent phenotypes with higher alcohol consumption. Third, our use of a nationwide biobank provided statistical power of our tests greater than those of previous studies²⁷.

Our analysis of TWB revealed that 71% of participants carried the G allele at rs671 and 29% carried the A allele. As for rs3782886, 71% of subjects carried the T allele and 29% carried the C allele. In other East Asian populations, at rs671 83% of individuals have the G allele, and 17% have the A allele. In the American, African, European, and South Asian populations, in contrast, this frequency is 100% for the G allele and 0% for the A allele. For East Asian populations, the allele frequency at rs3782886 was 83% for the T allele and 17% for the C allele. However, for all the other population groups, the frequency is 100% for the T allele and 0% for the C allele. The strong linkage disequilibrium between rs671 and rs3782886 ($r^2 = 0.98$) as well as the higher proportion of haplotype AC in East Asian populations are the evidence for a race-specific haplotype.

The association of rs3782886 with excessive alcohol use should not be neglected simply due to high linkage disequilibrium with rs671, a well-documented single nucleotide variant encoding the alcohol-metabolism enzyme^{28–30}. The reasons against such negligence are as follows.

First, *BRAP* is associated with a risk of myocardial infarction and a phenotype of metabolic traits in Asian populations^{31,32}. *BRAP* is a risk locus for metabolic syndrome³². Metabolites associated with alcohol consumption are primarily involved in amino and fatty acid metabolism^{33,34}. During ethanol metabolism as well as NADH and acetyl-CoA build up, more acetyl-CoA generate more malonyl-CoA. For fatty acid metabolism, that leads to inhibition of catabolism and activation of synthesis. Studies of *Caenorhabditis elegans* have demonstrated that *BRAP2* (*BRAP* homolog) regulates the expression of proteins involved in lipid synthesis³⁵. During persistent and excessive alcohol consumption, it is clinically implicated to elucidate the mechanisms between *BRAP* and metabolism of amino acid, and fatty acid.

Second, *BRAP* is involved in cerebral cortical neurogenesis^{36,37}. For neural progenitor cells, cell signalling during the G1 phase of the cell cycle requires *BRAP*³⁷. *BRAP* regulates at the cellular level MAP kinase pathways and the ubiquitin system³⁸, which likely controls the cascade of protein turnover during neuronal differentiation. Given that *BRAP* is involved in cell differentiation of the central nervous system, its involvement in mechanisms of neurobiological changes during excessive alcohol consumption should be further explored.

Third, we argue that *BRAP* plays a role in the regulation of reactive oxygen species (ROS) during excessive alcohol consumption³⁹. Both alcohol metabolism by CYP2E1 and the reoxidation of NADH via the electron transport chain in the mitochondria generate more ROS⁴⁰. The *BRAP*/nuclear factor erythroid 2-related factor (Nrf2) signalling cascade responds to oxidative stress³⁵, suggesting *BRAP* regulates ROS during excessive alcohol consumption.

In European populations, other consistently replicated hits in GWASs of alcohol consumption include *KLB*, *FGF21*, and *GCKR*, which are also involved in metabolism. But these hits were not identified in our present study. Plausible explanations of the discrepancy are as follows. First, particularly in East Asians, *BRAP* gene plays the major role in excessive alcohol consumption trait. Second, the liver–brain endocrine axis for homeostatic regulation responds to excessive alcohol consumption via *FGF21*^{11,34}, of which both *KLB* and Nrf2 are substrates closely affected by the nature of diet and food preference^{12,41}. It remains unclear as to how *BRAP*/Nrf2 signalling links to energy use and nutrient use regarding metabolism. Functional analysis is required to determine the role of *BRAP*/Nrf2 signalling in the liver–brain endocrine axis during the metabolic shift of excessive alcohol consumption.

Here, we reported the novel locus rs7398833 (*CUX2*), which is a 3'-UTR variant that functionally locks or releases the poly-A tail⁴². This function likely maintains the stability of the *CUX2* protein and subcellularly localizes the *CUX2* protein⁴². Second, *CUX2* is expressed mostly in the brain and is involved in neuronal differentiation in the cortex, specifically acting at the progenitors of GABAergic or dopaminergic neurons⁴³. Alcohol is a ligand for both GABAergic and dopaminergic receptors. Further studies to verify the genetic correlation between rs7398833 (*CUX2*) and excessive alcohol consumption are necessary.

We selected loci that were associated with excessive alcohol consumption and elevated levels of γ -GT. The average levels of serum γ -GT, at 46.15 ± 77.08 U/L, fell within the range of those of excessive alcohol users ($n = 1945$) and are higher than the average level of all 18,363 participants (26.01 ± 35.69 U/L). The high standard deviation of γ -GT levels of the participants with excessive alcohol consumption in our study could reflect asymptomatic patients with alcohol-induced hepatitis¹⁷.

Alterations in the metabolic profiles of excessive alcohol drinkers involve vastly different systems, such as carbohydrates, lipids, and proteins. To move a step closer to the metabolic traits of people with excessive alcohol consumption, we may need to study targets other than γ -GT. Nonetheless, γ -GT catabolises biliary glutathione and expands the pool of amino acid precursors required for conjugation (glycine [directly] and taurine [through cysteine oxidation]), thus implicating the metabolism of amino acids⁴⁴. Additionally, γ -GT represents the impact of metabolic disease on vascular injury and atherosclerosis^{45,46}. In this aspect, our study showed that mechanisms underlying the γ -GT catalytic metabolic reaction among people with excessive alcohol consumption are associated with *ALDH2*, *BRAP* and *CUX2*.

Considering the impact of socioeconomic backgrounds, the living locations, income and education levels were incorporated in measurement of our study. Information of education levels had 0.08% missing data. The income information had 54.3% missing data, and interpretation subject to the lack of thereof. In population-based study, voluntary participation tends to attract individuals with higher education levels and socioeconomic status, as well as lower levels of problem drinking⁴. This trend complemented our study.

Our study has several limitations. First, we excluded significant intronic SNPs and used only significant exonic SNPs. The reason of why we excluded intronic variants was due to the limited sample size. The intronic signals that might be involved in alternative splicing and gene expression were therefore overlooked⁴³. As a result, intronic variants that convey a risk of excessive alcohol consumption were likely to be missed. Second, we defined “excessive alcohol consumption” according to the criterion of a weekly intake of > 150 mL of alcohol for > 6 months. The types of beverages consumed were unclear. Low-risk alcohol use of < 100 g/week is equivalent to 7.1 cans of beer (350 mL each, 5% alcohol content) or 1.3 bottles of wine (750 mL, 13%). Our definition of excessive alcohol consumption was stricter than that employed in the literature. However, in the Taiwan Biobank one cannot identify how many of the excessive users had an alcohol use disorder diagnosis. Third, out of 18,363 Taiwanese subjects, 1945 (~ 10%) were defined as cases, and 16,418 participants (~ 90%) were defined as controls in this case–control study. In addition, there was a sex imbalance in this sample. We addressed the limitation of case–control imbalance. In future work, SAIGE (Scalable and Accurate Implementation of Generalized mixed model) could be used to account for sample imbalance⁴⁷. Nonetheless, the PCA plot for the genetic ancestry of this TWB cohort revealed that the distribution had no obvious deviation (see Supplementary Fig. S1 online). Fourth, the majority of individuals from eastern Taiwan and the outlying islands live in rural townships. Supplementary Fig. S6 online shows that the prevalence of excessive alcohol consumption is likely to be different among individuals from northern, central, southern and eastern Taiwan. Those on the outlying islands had higher frequencies of excessive alcohol consumption. Owing to the small sample size from the outlying islands, we did not correct these islanders. Lastly, our findings did not provide directionality of causality (metabolism vs. alcoholism). One way to clarify this issue is to use Mendelian randomisations in future studies.

In conclusion, we developed an alternative strategy for overcoming extensive regional linkage disequilibrium. We uncovered Glu504Lys of *ALDH2* and Glu4Gly of *BRAP*, which are involved in the negative regulation of excessive alcohol consumption. The mechanism underlying the γ -GT catalytic metabolic reaction in excessive alcohol consumption is associated with *ALDH2*, *BRAP* and *CUX2*. Further studies are needed to determine the roles of *ALDH2*, *BRAP* and *CUX2* in the liver-brain endocrine axis upon the metabolic shift with excessive alcohol consumption.

Methods

Study participants. Data were taken from the TWB, which were random samples of Taiwanese people aged 30 to 70 years old with no history of cancer. Information analyzed was related to genomic data and lifestyle^{48,49}. Lifestyle factors included current tobacco use and cigarette smoking, weekly exercise activity of ≥ 3 times, each ≥ 30 min. We measured medical history containing the following conditions: gout, hypertension, hyperlipidaemia, stroke, diabetes mellitus, peptic ulcer, irritable bowel syndrome, migraine, gastric-oesophageal reflux syndrome, depressive disease, bipolar disorder, and schizophrenia. Using posters, brochures, websites, and audio and video media, we recruited TWB participants from 27 outreach centres in the rural and urban townships in Taiwan (see Supplementary Fig. S6 online). All participants signed informed consent forms. This study was approved by the Ethics Review Committees of National Taiwan University Hospital (project number: 201506095RINC).

Genotyping. In the TWB, whole-genome genotyping was conducted on DNA extracted from blood samples using a QIAamp DNA blood kit, according to the manufacturer’s instructions (Qiagen, Valencia, CA, USA). The qualitative information of the extracted genomic DNA was visualised using agarose gel electrophoresis, and quantitative properties were measured by spectrophotometry. Samples were genotyped with a custom-designed Affymetrix Axiom Genome-Wide Array Plate, which contained 653,291 SNPs. To reach genotyping call-rate of 0.95, SNP and sample quality control thresholds were used in PLINK, a whole-genome data analysis toolset (MIND > 0.05). The identity state was set at 0.4 for each pair of individuals based on the average proportion of alleles shared at the genotyped SNPs. Those SNPs not following the Hardy–Weinberg equilibrium (with cut-off $p > 1 \times 10^{-6}$) or rare variants with minor allele frequencies ($< 1 \times 10^{-3}$) were pruned. In total, 601,531 SNPs remained after the exclusion. Imputation was conducted with the Michigan Imputation Server (<https://imputationserver.sph.umich.edu>) using 1000G phase 3 v 5 as a reference panel. Eagle v 2.3 was used for phasing, and the EAS population was used for quality control. We imputed 11,389,991 variants of the TWB data based on the East Asian panel of the 1000 Genomes dataset. For imputation quality control, the criteria considered were an imputation quality score of > 0.8 and minor allele frequency of > 0.01. Finally, 6,410,722 variants successfully passed the two quality control stages.

Statistical analyses. Based on information related to lifestyle, medical history, and the genotypes of 6,410,722 SNPs, we used the principal component analysis to extract 10 principal components for modelling the data. Multivariate logistic regression was used to calculate the odds ratio and p value for each SNP, and the model comprised of age, sex, and 10 principle components. We used the additive model to determine genotype risks. The false detection rate (FDR) was calculated to overcome effects of multiple tests. To determine the number of independent signals, the cut-off of FDR is less than 0.05.

Intermarker linkage disequilibrium is possibly caused by distance proximity and the coexpression of genes. If n is the number of significant SNPs, there are C_2^n possible pairs with intermarker linkage disequilibrium. The ordinary least squares approach results in hypercollinearity when a full set of significant SNPs is included in the multivariate regression model. To solve the hypercollinearity problem, we used the ridge regression. Ridge regression minimises a penalty-augmented loss function and obtains the optimisation parameters $\hat{\beta}^{ridge}$.

$$\hat{\beta}^{ridge} = \underset{\beta \in R}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2,$$

where $\|\beta\|_2 = \sqrt{\beta_0^2 + \beta_1^2 + \dots + \beta_p^2}$ and λ is the shrinkage parameter that controls the size of coefficients and amount of regularisation. As λ approaches zero, the least square solutions are obtained; as λ approaches infinity, the ridge coefficients $\hat{\beta}^{ridge} = 0$ are obtained. The result is a constant (intercept-only) model. We selected the SNPs for which $\hat{\beta}^{ridge}$ was stable across different λ values.

Statistical analyses were conducted using R, Python open-source programming languages, FUMA GWAS (<https://fuma.ctglab.nl/>), LDSC software (<https://github.com/bulik/ldsc>), Plink version 1.90, the Multiple GWAS comparison and PheWAS of the GWAS ATLAS resource (<https://atlas.ctglab.nl/>), HAPLOVIEW version 4.2, and standard SAS software.

Gene-set based analysis. To map the most significant genes to particular clusters of biological mechanisms, we conducted gene list analysis. The Gene Ontology (GO) terms were used for functional annotation. We performed gene-list analysis by using PANTHER software and tools⁵⁰. The list of significant genes was uploaded directly on the homepage of the GO website (geneontology.org/docs/go-enrichment-analysis). Hypergeometric distribution was applied to test whether the overrepresentation of a GO term occurred significantly more often than chance. Hypergeometric distribution and binomial test were applied to test whether the overrepresentation of a GO term occurred significantly more often than chance. Cut-off of p value is < 0.05 . Fold enrichment was defined as the number of significant genes in the list divided by the expected number of genes in a particular GO category⁵⁰.

Ethics approval. The study abided the Declaration of Helsinki. This study was approved by the Ethics Review Committees of National Taiwan University Hospital (project number: 201506095RINC).

Data availability

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation, to any qualified researchers.

Received: 13 October 2019; Accepted: 12 October 2020

Published online: 22 October 2020

References

- 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).
- Chen, W. J. *et al.* Differences in prevalence, socio-behavioral correlates, and psychosocial distress between club drug and hard drug use in Taiwan: Results from the 2014 national survey of substance use. *Int. J. Drug Policy* **48**, 99–107 (2017).
- Cheng, H. G., Deng, F., Xiong, W. & Phillips, M. R. Prevalence of alcohol use disorders in mainland China: A systematic review. *Addiction* **110**, 761–774 (2015).
- Sanchez-Roige, S., Palmer, A. A. & Clarke, T. K. Recent efforts to dissect the genetic basis of alcohol use and abuse. *Biol. Psychiatry* **87**, 609–618 (2020).
- Edenberg, H. J., Gelernter, J. & Agrawal, A. Genetics of alcoholism. *Curr. Psychiatry Rep.* **21**, 26 (2019).
- Deak, J. D., Miller, A. P. & Gizer, I. R. Genetics of alcohol use disorder: A review. *Curr. Opin. Psychol.* **27**, 56–61 (2019).
- Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, 1499. <https://doi.org/10.1038/s41467-019-09480-8> (2019).
- Larsson, S. C., Burgess, S., Mason, A. M. & Michaëlsson, K. Alcohol Consumption and cardiovascular disease: A Mendelian Randomization Study. *Circ. Genom. Precis. Med.* <https://doi.org/10.1161/CIRCGEN.119.002814> (2020).
- 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).
- 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).
- 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. USA* **113**, 14372–14377 (2016).
- Matsui, S. *et al.* Neuronal SIRT1 regulates macronutrient-based diet selection through FGF21 and oxytocin signalling in mice. *Nat. Commun.* **9**, 4604. <https://doi.org/10.1038/s41467-018-07033-z> (2018).
- Aryal, P., Dvir, H., Choe, S. & Slesinger, P. A. A discrete alcohol pocket involved in GIRK channel activation. *Nat. Neurosci.* **12**, 988–995 (2009).
- Ron, D. & Barak, S. Molecular mechanisms underlying alcohol-drinking behaviours. *Nat. Rev. Neurosci.* **17**, 576–591 (2016).
- Sanchez-Roige, S. *et al.* Genome-Wide association study meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) in two population-based cohorts. *Am. J. Psychiatry* **176**, 107–118 (2018).

16. Zuhlendorf, A. *et al.* It is not always alcohol abuse—A transferrin variant impairing the CDT test. *Alcohol Alcohol*. **51**, 148–153 (2016).
17. Pratt, D. S. & Kaplan, M. M. Evaluation of abnormal liver-enzyme results in asymptomatic patients. *N. Engl. J. Med.* **342**, 1266–1271 (2000).
18. Spoto, B., D'Arrigo, G., Tripepi, G., Bolignano, D. & Zoccali, C. Serum gamma-glutamyltransferase, oxidized LDL and mortality in the elderly. *Aging Clin. Exp. Res.* <https://doi.org/10.1007/s40520-019-01391-4> (2019).
19. Zhang, H., Forman, H. J. & Choi, J. Gamma-glutamyl transpeptidase in glutathione biosynthesis. *Methods Enzym.* **401**, 468–483 (2005).
20. Chen, C. H. *et al.* Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan biobank project. *Hum. Mol. Genet.* **25**, 5321–5331 (2016).
21. Bulik-Sullivan, B. K. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
22. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
23. Kim, J. W. *et al.* Associations of BRAP polymorphisms with the risk of alcohol dependence and scores on the Alcohol Use Disorders Identification Test. *Neuropsychiatr. Dis. Treat.* **15**, 83–94 (2019).
24. de Vlaming, R. & Groenen, P. J. The current and future use of ridge regression for prediction in quantitative genetics. *Biomed. Res. Int.* **2015**, 143712. <https://doi.org/10.1155/2015/143712> (2015).
25. Romagnoni, A., Jegou, S., Van Steen, K., Wainrib, G. & Hugot, J. P. Comparative performances of machine learning methods for classifying Crohn Disease patients using genome-wide genotyping data. *Sci. Rep.* **9**, 10351. <https://doi.org/10.1038/s41598-019-46649-z> (2019).
26. Endelman, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**, 250–255 (2011).
27. Quillen, E. E. *et al.* ALDH2 is associated to alcohol dependence and is the major genetic determinant of “daily maximum drinks” in a GWAS study of an isolated rural Chinese sample. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **165b**, 103–110 (2014).
28. 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**, 911–918 (2011).
29. 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**, 809–816 (2011).
30. Chang, B. *et al.* Association between aldehyde dehydrogenase 2 Glu504Lys polymorphism and alcoholic liver disease. *Am. J. Med. Sci.* **356**, 10–14 (2018).
31. Ozaki, K. *et al.* SNPs in BRAP associated with risk of myocardial infarction in Asian populations. *Nat. Genet.* **41**, 329–333 (2009).
32. Avery, C. L. *et al.* A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains. *PLoS Genet.* **7**, e1002322 (2011).
33. Harada, S. *et al.* Metabolomic profiling reveals novel biomarkers of alcohol intake and alcohol-induced liver injury in community-dwelling men. *Environ. Health Prev. Med.* **21**, 18–26 (2016).
34. Yang, Z. *et al.* Serum metabolomic profiling identifies key metabolic signatures associated with pathogenesis of alcoholic liver disease in humans. *Hepatology*. **3**, 542–557 (2019).
35. Hu, Q., D'Amora, D. R., MacNeil, L. T., Walhout, A. J. M. & Kubiseski, T. J. The caenorhabditis elegans oxidative stress response requires the NHR-49 transcription factor. *G3* **8**, 3857–3863 (2018).
36. Lancot, A. A., Peng, C. Y., Pawlisz, A. S., Joksimovic, M. & Feng, Y. Spatially dependent dynamic MAPK modulation by the Nde1-Lis1-Brp complex patterns mammalian CNS. *Dev. Cell* **25**, 241–255 (2013).
37. Lancot, A. A. *et al.* Loss of brap results in premature G1/S phase transition and impeded neural progenitor differentiation. *Cell Rep.* **20**, 1148–1160 (2017).
38. Shoji, S., Hanada, K., Ohsawa, N. & Shirouzu, M. Central catalytic domain of BRAP (RNF52) recognizes the types of ubiquitin chains and utilizes oligo-ubiquitin for ubiquitylation. *Biochem. J.* **474**, 3207–3226 (2017).
39. Zakhari, S. Alcohol metabolism and epigenetics changes. *Alcohol Res. Curr. Rev.* **35**, 6–16 (2013).
40. George, A. K., Behera, J., Kelly, K. E., Zhai, Y. & Tyagi, N. Hydrogen sulfide, endoplasmic reticulum stress and alcohol mediated neurotoxicity. *Brain Res. Bull.* **130**, 251–256 (2017).
41. Parira, T. *et al.* Trichostatin A shows transient protection from chronic alcohol-induced Reactive Oxygen Species (ROS) production in human monocyte-derived dendritic cells. *J. Alcohol Drug Depend.* **6**, 316 (2018).
42. Szostak, E. & Gebauer, F. Translational control by 3'-UTR-binding proteins. *Brief. Funct. Genom.* **12**, 58–65 (2013).
43. Nefzger, C. M. *et al.* Lmx1a allows context-specific isolation of progenitors of GABAergic or dopaminergic neurons during neural differentiation of embryonic stem cells. *Stem Cells* **30**, 1349–1361 (2012).
44. Chen, M. F. *et al.* Preventive effect of YGDEY from Tilapia fish skin gelatin hydrolysates against alcohol-induced damage in HepG2 cells through ROS-mediated signaling pathways. *Nutrients* **11**, E392. <https://doi.org/10.3390/nu11020392> (2019).
45. Bradley, R. D. *et al.* Associations between gamma-glutamyltransferase (GGT) and biomarkers of atherosclerosis: The Multi-ethnic Study of Atherosclerosis (MESA). *Atherosclerosis* **233**, 387–393 (2014).
46. Choi, K. M. *et al.* Implication of liver enzymes on incident cardiovascular diseases and mortality: A nationwide population-based cohort study. *Sci. Rep.* **8**, 3764. <https://doi.org/10.1038/s41598-018-19700-8> (2018).
47. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
48. Lin, W. Y., Huang, C. C., Liu, Y. L., Tsai, S. J. & Kuo, P. H. Polygenic approaches to detect gene–environment interactions when external information is unavailable. *Brief. Bioinform.* **20**, 2236–2252 (2019).
49. Lin, E. *et al.* Effects of circadian clock genes and environmental factors on cognitive aging in old adults in a Taiwanese population. *Oncotarget* **8**, 24088–24098 (2017).
50. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).

Acknowledgements

This work was supported by grants from the following institutes: the Ministry of Economic Affairs, Taiwan (SBIR Grant S099000280249-154; EL), Taipei Veterans General Hospital, Taiwan (Grants VGHUST103-G1-4-1, V105C-008, and V105E17-002-MY2-1; SJT), the Ministry of Science and Technology (MOST 106-2320-B-400-012), and the National Health Research Institutes in Taiwan (NP108, 109-SP-04, NP-108-PP-06, and NP-109-PP-07). We thank Ms. Jun Ru Wei and Mr. Ya-Chin Lee for data collection and analysis. We sincerely thank Prof. Yi-Mei Lin for intellectual contributions.

Author contributions

I.-C.C. contributed to drafting of the manuscript and statistical analysis. P.-H.K., A.C.Y. and S.-J.T. contributed to acquisition of data and obtaining funding. T.-H.L. contributed to the statistical analysis. H.-J.L. and T.-H.L.

contributed to critical revision of the manuscript for important intellectual content. H.-M.C., H.-N.H. and R.-H.C. assisted with the analysis and interpretation of the data. Y.-L.L. contributed to the conception and design.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-75199-y>.

Correspondence and requests for materials should be addressed to H.-J.L. or Y.-L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020