

## ARTICLE OPEN



# Genome-wide meta-analysis of alcohol use disorder in East Asians

Hang Zhou<sup>1,2</sup>, Rasmon Kalayasiri<sup>3,4,5</sup>, Yan Sun<sup>6</sup>, Yaira Z. Nuñez<sup>1,2</sup>, Hong-Wen Deng<sup>7</sup>, Xiang-Ding Chen<sup>8</sup>, Amy C. Justice<sup>9,10</sup>, Henry R. Kranzler<sup>11,12</sup>, Suhua Chang<sup>13</sup>, Lin Lu<sup>6,13</sup>, Jie Shi<sup>13</sup>, Kittipong Sanichwankul<sup>14</sup>, Apiwat Mutirangura<sup>5</sup>, Robert T. Malison<sup>1,16</sup> and Joel Gelernter<sup>1,2,15</sup>✉

© The Author(s) 2022

Alcohol use disorder (AUD) is a leading cause of death and disability worldwide. Genome-wide association studies (GWAS) have identified ~30 AUD risk genes in European populations, but many fewer in East Asians. We conducted GWAS and genome-wide meta-analysis of AUD in 13,551 subjects with East Asian ancestry, using published summary data and newly genotyped data from five cohorts: (1) electronic health record (EHR)-diagnosed AUD in the Million Veteran Program (MVP) sample; (2) DSM-IV diagnosed alcohol dependence (AD) in a Han Chinese–GSA (array) cohort; (3) AD in a Han Chinese–Cyto (array) cohort; and (4) two AD Thai cohorts. The MVP and Thai samples included newly genotyped subjects from ongoing recruitment. In total, 2254 cases and 11,297 controls were analyzed. An AUD polygenic risk score was analyzed in an independent sample with 4464 East Asians (Genetic Epidemiology Research in Adult Health and Aging (GERA)). Phenotypes from survey data and ICD-9-CM diagnoses were tested for association with the AUD PRS. Two risk loci were detected: the well-known functional variant rs1229984 in *ADH1B* and rs3782886 in *BRAP* (near the *ALDH2* gene locus) are the lead variants. AUD PRS was significantly associated with days per week of alcohol consumption ( $\beta = 0.43$ ,  $SE = 0.067$ ,  $p = 2.47 \times 10^{-10}$ ) and nominally associated with pack years of smoking ( $\beta = 0.09$ ,  $SE = 0.05$ ,  $p = 4.52 \times 10^{-2}$ ) and ever vs. never smoking ( $\beta = 0.06$ ,  $SE = 0.02$ ,  $p = 1.14 \times 10^{-2}$ ). This is the largest GWAS of AUD in East Asians to date. Building on previous findings, we were able to analyze pleiotropy, but did not identify any new risk regions, underscoring the importance of recruiting additional East Asian subjects for alcohol GWAS.

*Neuropsychopharmacology* (2022) 47:1791–1797; <https://doi.org/10.1038/s41386-022-01265-w>

## INTRODUCTION

Globally, alcohol use disorders (AUD) are among the top causes of morbidity and mortality [1]. Numerous factors predispose to the risk of developing AUD. Genetic factors contribute substantial risk to the etiology of AUD [2], and the heritability has been estimated to be ~0.5 in twin studies [3]. Genome-wide association studies (GWAS) of AUD have been completed in multiple populations including European (EUR), African, Latin American and Asians [4–13]. To date, the largest GWAS of problematic alcohol use (PAU, a proxy for AUD) in 435,563 EUR subjects identified 29 independent risk variants [14]. In contrast, the largest GWAS of AUD in East Asians included less than 1% of this number: 3381 subjects (533 cases) [12]. Genetic architecture often differs between populations; polygenic risk prediction between populations, though sometimes useful, often is not transferable [15]. Thus, it is critically important that non-EUR populations be investigated

to permit inferences to be made about these ancestral populations, which represent the majority of the world's people [16, 17].

Because of the limited sample available and consequent lack of power for GWAS, little is known about the genetic architecture of AUD in East Asians. The most consistent loci identified are *ADH1B* (alcohol dehydrogenase 1b) and *ALDH2* (aldehyde dehydrogenase). Candidate studies of *ADH1B*\*rs1229984 and *ALDH2*\*rs671 in East Asians showed strong associations between these functional variants and alcohol dependence (AD) [18, 19]. The first GWAS of AD in a Chinese sample was conducted in 102 male cases and 212 male controls; rs3782886 in the *ALDH2* region was genome-wide significant [5] despite the very small sample size. The first GWAS of AD in Thai samples included 1045 subjects and identified rs149212747 in the *ALDH2* region as the lead variant [6]. The latest GWAS of AUD in a Chinese cohort identified both *ADH1B* and *ALDH2* genes as risk loci [12]. However, only a small

<sup>1</sup>Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. <sup>2</sup>Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA. <sup>3</sup>Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand. <sup>4</sup>Department of Psychiatry, King Chulalongkorn Memorial Hospital, Bangkok, Thailand. <sup>5</sup>Center for Excellence in Molecular Genetics of Cancer and Human Diseases, Department of Anatomy, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand. <sup>6</sup>National Institute on Drug Dependence, Peking University, Beijing, China. <sup>7</sup>Center for Biomedical Informatics and Genomics, School of Medicine, Tulane University, New Orleans, LA, USA. <sup>8</sup>Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, Hunan, China. <sup>9</sup>Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. <sup>10</sup>Center for Interdisciplinary Research on AIDS, Yale School of Public Health, New Haven, CT, USA. <sup>11</sup>University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>12</sup>Crescenz Veterans Affairs Medical Center, Philadelphia, PA, USA. <sup>13</sup>Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health (Peking University), National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital), Peking University, Beijing, China. <sup>14</sup>Suanprung Psychiatric Hospital, Chiang Mai, Thailand. <sup>15</sup>Departments of Genetics and Neuroscience, Yale University School of Medicine, New Haven, CT, USA. <sup>16</sup>Deceased: Robert T. Malison. ✉email: joel.gelernter@yale.edu

Received: 20 September 2021 Revised: 22 December 2021 Accepted: 29 December 2021  
Published online: 29 January 2022

**Table 1.** Sample characteristics.

Cohorts	Traits	N (%female)	# Cases	Mean (SD), Age	Country of recruitment
Thai METH–GSA	DSM-IV AD	532 (49.4)	127	26.6 (6.9)	Thailand
Thai METH–MEGA	DSM-IV AD	2370 (42.5)	794	34.7 (10.1)	Thailand
MVP–EAA	ICD-9/10 AUD	6955 (10.7)	701	53.4 (17.1)	United States
Han Chinese–GSA	DSM-IV AD	3381 (29.9)	533	34.2 (8.3)	China
Han Chinese–Cyto	DSM-IV AD	313 (0)	99	49.6 (14.7)	China
Total		13,551	2254		

Thai METH studies of the genetics of methamphetamine dependence in Thailand, GSA Global Screening Array, MEGA Multi-Ethnic Global Array, EAA East Asian American, Cyto Cyto12 array, AD alcohol dependence, AUD alcohol use disorder.

proportion of the variance was explicable by variants in these genes. Larger samples are required to identify more risk variants to provide a better understanding of the genetic architecture in Asian populations.

Here we conducted a GWAS that combined five datasets from previously published cohorts and newly genotyped subjects from Thai and MVP cohorts. In total, 13,551 subjects of East Asian ancestry were analyzed, including 2254 AUD cases. We then analyzed the resulting AUD PRS in an independent East Asian sample for associations with 26 phenotypes from surveys or ICD diagnoses. This GWAS of AUD is the largest to date in East Asians.

## METHODS AND MATERIALS

### Datasets

**Thai METH–GSA.** As described previously [6], subjects were recruited in two stages for studies of the genetics of methamphetamine dependence (Thai METH). For both stages, subjects were recruited in Bangkok and assessed using the Thai version of the Semi-Structured Assessment for Drug Dependence and Alcoholism [20]. The IRB protocols were approved by both the Chulalongkorn University (Thailand) IRB and the Yale University Human Research Protection Program. All subjects provided written informed consent prior to their research participation.

The first stage included methamphetamine users hospitalized between 2007 and 2011 for 4 months of residential drug treatment (Thai METH–GSA, Table 1) [21]. DNA samples were genotyped on the Illumina (San Diego, CA) Global Screening Array (GSA) which includes ~640 K SNPs. Among the 863 genotyped subjects, we removed those with sample genotype call rate <0.9, mismatched genotypic and phenotypic sex, or excess heterozygosity rate [6]. Unlike for our prior report, here we retained related subjects and applied linear mixed models (LMM) to correct for relatedness (see below). SNPs with genotype call rate  $\geq 0.95$ , minor allele frequency (MAF)  $\geq 0.01$ , and Hardy–Weinberg equilibrium (HWE)  $p$  value  $> 10^{-6}$  were kept for imputation. Imputation was done by IMPUTE2 [22] with 1000 Genome project phase 3 (1KG) data [23] as reference. SNPs with imputation INFO score  $\geq 0.8$ , best-guess genotype call rate  $\geq 0.95$ , MAF  $\geq 0.01$  and HWE  $p$  value  $> 10^{-6}$  were retained for association analyses. Principal component analysis (PCA) was performed for the remaining subjects using EIGENSOFT [24, 25]. In contrast with our previous study, here we used DSM-IV AD to define case status, rather than the DSM-IV AD criterion count to match the design in other cohorts. This yielded 127 cases and 405 were exposed controls. LMM implemented in GEMMA [26] were used to test association, with age, sex, and the first ten PCs as covariates.

**Thai METH–MEGA.** Second-stage subjects ( $N = 3,161$ ; the Thai METH–MEGA sample, Table 1) were recruited from 2015 to 2020 [6]. DNA samples were genotyped using the Illumina Multi-Ethnic Global Array (MEGA) which includes ~1.78 M SNPs. We removed subjects with sample genotype call rate <0.95, sex mismatch, excess heterozygosity rate, or that were duplicates. SNPs with genotype call rate  $\geq 0.95$ , or MAF  $\geq 0.01$ , or HWE  $p$  value  $> 10^{-6}$  were retained for imputation as with the Thai METH–GSA sample. The same imputation processes and post-imputation quality controls (QC) were applied. We included 794 cases and 1576 alcohol-exposed controls in the association analysis, which used GEMMA and age, sex and the first ten PCs as covariates.

**MVP–EAA.** The Million Veteran Program (MVP) is an ongoing observational cohort study and mega-biobank supported by the U.S. Department of Veterans Affairs [27, 28]. In October 2020, MVP released the latest genotype data (Release 4), which included 658,582 subjects. MVP subjects were genotyped using an Affymetrix Axiom Biobank Array with ~687 K markers. QC was first done by the MVP Release 4 Data Team and included the removal of duplicate DNA samples and those with sex mismatch, excessive heterozygosity, or a genotype call rate <0.985. We ran PCA for the MVP subjects with 1KG as the reference, Euclidean distances between each participant and the centers of the five reference populations were calculated using the first ten PCs, with each participant assigned to the nearest reference population. For subjects grouped as East Asian Americans (EAA), we ran a second PCA and removed outliers with PC scores  $> 6$  standard deviations from the mean on any of the ten PCs (as we did before [8]), yielding in 7364 EAAs. Imputation [22] was performed specifically for the EAAs using the 1KG as reference. SNPs with genotype call rate  $\geq 0.95$ , MAF  $\geq 0.01$ , HWE  $> 1 \times 10^{-6}$ , and imputation INFO  $\geq 0.8$  were retained for analysis. As for our prior study in EUR [14], subjects with  $\geq 2$  outpatient or  $\geq 1$  inpatient International Classification of Diseases (ICD) codes for AUD were defined as cases ( $N = 701$ , Table 1) and subjects with no AUD ICD code as controls ( $N = 6254$ ). BOLT-LMM [29] was used to correct for relatedness, with age, sex, and the first ten PCs as covariates.

**Han Chinese–Cyto.** This first GWAS of AD, flushing response, and maximum daily drinks consumed in a Han Chinese family sample [5] used the Illumina Cyto12 array containing ~300 K SNPs (Table 1). Whereas the cohort was not imputed in the original report, we re-analyzed the data and imputed the SNPs for an AD GWAS. Subjects with genotype call rate <0.95, duplicated DNA samples, mismatched sex, or excessive heterozygosity were removed, resulting in 511 subjects for imputation. Imputation used IMPUTE2 and 1KG reference, SNPs with MAF <0.01, genotype call rate <0.95, HWE  $p$  value  $< 1 \times 10^{-6}$ , or imputation INFO <0.8 were removed from further analyses. Due to the drinking practices and characteristic of this particular population that result in few AD cases in females, only males were included in this analysis. After QC, 99 DSM-IV-diagnosed male AD cases and 214 male alcohol-exposed controls were analyzed using GEMMA to correct for relatedness, with age, sex and ten PCs as covariates.

**Han Chinese–GSA.** The second case-control AD GWAS in Han Chinese [12] included 533 cases and 2848 alcohol-exposed controls who were genotyped using the GSA array (Table 1). Here we used the summary statistics from previous study.

### Meta-analysis

Using association analyses or summary statistics for each of the five cohorts, effective sample-size-weighted meta-analysis was performed using METAL [30]. SNPs present in only one cohort or in less than 15% of the total samples were removed (6.8 million SNPs remained). To define lead variants, the meta-analysis summary data were clumped by LD with  $r^2 < 0.1$  in a 2500-kb window, using 1KG East Asians as the LD reference. For the two lead SNPs in the *ADH1B* gene region (rs1229984 and rs1814125), we performed conditional analysis [31] for rs1814125 conditioning on rs1229984 to test if its association is independent from rs1229984. Regional association plots were generated using LocusZoom v1.4 [32] with reference LD calculated from corresponding 1KG populations. We converted the effect sizes of lead SNPs from the LMM to odds ratios (OR) for comparison and further investigation of cohort heterogeneity [33].

**Table 2.** Tested phenotypes in GERA and association results with AUD PRS.

Traits	Distribution	Beta (SE)	p value
Alcohol use in days per week	1 = 2757, 2 = 603, 3 = 503, 4 = 159, 5 = 267 <sup>c</sup>	0.43 (0.07)	<b>2.47 × 10<sup>-10</sup></b>
Smoking in pack years	0 = 3232, 1 = 530, 2 = 306, 3 = 128, 4 = 44 <sup>d</sup>	0.09 (0.05)	<b>4.52 × 10<sup>-2</sup></b>
Ever vs. never smoked	1 = 1055, 0 = 3232	0.06 (0.02)	<b>1.14 × 10<sup>-2</sup></b>
Former vs. current smoker	1 = 924, 0 = 131	-0.04 (0.04)	2.50 × 10 <sup>-1</sup>
Physical activity	1 = 897, 2 = 958, 3 = 1181, 4 = 1323 <sup>e</sup>	0.02 (0.06)	7.97 × 10 <sup>-1</sup>
Health status	1 = 740, 2 = 1506, 3 = 1625, 4 = 476 <sup>f</sup>	0.03 (0.05)	5.23 × 10 <sup>-1</sup>
Disease or conditions			
Acute reaction to stress	1 = 275, 0 = 4189	-0.01 (0.01)	5.22 × 10 <sup>-1</sup>
Allergic rhinitis	1 = 1307, 0 = 3157	0.01 (0.03)	5.62 × 10 <sup>-1</sup>
Asthma	1 = 654, 0 = 3810	-0.01 (0.02)	7.21 × 10 <sup>-1</sup>
Cancer: any <sup>a</sup>	1 = 529, 0 = 3935	-0.00 (0.02)	9.36 × 10 <sup>-1</sup>
Cardiovascular disease: any <sup>b</sup>	1 = 688, 0 = 3776	-0.03 (0.02)	1.48 × 10 <sup>-1</sup>
Major depressive disorder	1 = 262, 0 = 4202	0.01 (0.01)	3.66 × 10 <sup>-1</sup>
Dermatophytosis	1 = 374, 0 = 4090	-0.01 (0.02)	4.87 × 10 <sup>-1</sup>
Type II diabetes	1 = 729, 0 = 3735	0.03 (0.02)	9.65 × 10 <sup>-2</sup>
Dyslipidaemia	1 = 2192, 0 = 2272	-0.02 (0.03)	5.17 × 10 <sup>-1</sup>
Hemorrhoids	1 = 716, 0 = 3748	0.01 (0.02)	6.19 × 10 <sup>-1</sup>
Hernia abdominopelvic cavity	1 = 177, 0 = 4287	0.00 (0.01)	7.54 × 10 <sup>-1</sup>
Hypertensive disease	1 = 2028, 0 = 2436	-0.00 (0.02)	9.52 × 10 <sup>-1</sup>
Insomnia	1 = 185, 0 = 4279	-0.01 (0.01)	2.30 × 10 <sup>-1</sup>
Iron deficiency anemias	1 = 118, 0 = 4346	0.00 (0.01)	9.14 × 10 <sup>-1</sup>
Irritable bowel syndrome	1 = 103, 0 = 4361	0.00 (0.01)	7.05 × 10 <sup>-1</sup>
Macular degeneration	1 = 130, 0 = 4334	-0.00 (0.01)	7.94 × 10 <sup>-1</sup>
Osteoarthritis	1 = 941, 0 = 3523	-0.01 (0.02)	6.72 × 10 <sup>-1</sup>
Osteoporosis	1 = 392, 0 = 4072	0.01 (0.01)	6.46 × 10 <sup>-1</sup>
Psychiatric disorder: any	1 = 433, 0 = 4031	-0.00 (0.02)	9.93 × 10 <sup>-1</sup>
Peripheral vascular disease	1 = 160, 0 = 4304	0.01 (0.01)	3.81 × 10 <sup>-1</sup>

If not specified for distribution, 1 is case and 0 is control. Traits with  $p$  value < 0.05 are labeled in bold font.

<sup>a</sup>Cancer: includes malignant tumors, neoplasms, lymphoma and sarcoma.

<sup>b</sup>Heart disease: includes ischemic heart disease, cardiac arrest, congestive heart failure, dysrhythmias, cardiomyopathy, aortic aneurysm, and cerebrovascular disease, but excludes PVD which is encompassed by the PVD variable.

<sup>c</sup>Days of alcohol intake per week, 1 is no days, 2 is 1 day, 3 is 2–4 days, 4 is 5–6 days, 5 is every day.

<sup>d</sup>Pack years for former or current smoker, 0 = 0, 1 < 10, 2 = 10–20, 3 = 20–30, 4 ≥ 30.

<sup>e</sup>Physical activity total metabolic equivalency of task (MET), 1 = first quartile, 0–173 for males and 0–74 for females, 2 = second quartile, 174–600 for males and 75–344 for females, 3 = third quartile, 601–1380 for males and 345–983 for females, 4 = fourth quartile, 1381+ for males and 984+ for females.

<sup>f</sup>Health status, 1 = excellent, 2 = very good, 3 = good, 4 = fair.

This method takes sample prevalence, effect size from LMM, and allele frequency as input. We also did meta-analyses for the lead SNPs using inverse variance-weighted meta-analysis using METAL and the converted log(OR) as input, for comparison.

### Polygenic risk scores

**Target dataset.** We requested and downloaded dbGaP (phs000788.v2.p3) data from the Kaiser Permanente Research Program on Genes, Environment, and Health Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. This large and ethnically diverse cohort contains genotype data from 5182 self-reported Asians using a custom Affymetrix Axiom array [34]. All subjects completed a broad written consent.

**Imputation.** Subjects with mismatched sex or genotype call rate < 0.95 were removed. The genomic build was transferred from 36 to 37 using LiftOver [35]. As we did for MVP, we ran PCA for the 5182 Asian subjects using the 1KG as reference, clustering them into different groups. A second PCA among Asians was used to remove outliers, resulting in 4464 genetically classified East Asians for imputation. Imputation was performed using IMPUTE2 and 1KG reference. SNPs with MAF < 0.01, genotype call rate < 0.95, HWE  $p$  value <  $1 \times 10^{-6}$ , or imputation INFO < 0.8 were removed from further analysis.

**Target phenotypes.** Two sources of phenotypes are included in this study. The first is survey data on physical observations, lifestyle and environment, including phenotypes such as BMI, general health, physical activity, alcohol use, smoking status and pack years. The second is ICD-9-CM disease and conditions measures. Participant were coded as cases if there were at least two diagnoses in a disease category. Binary phenotypes with less than 100 cases were removed from analyses. See Table 2 for details of the target phenotypes.

**Polygenic risk scoring and association.** PRS-CS [36] was used to infer posterior effect sizes of SNPs using GWAS summary statistics for AUD from this study, and an external East Asian LD reference panel (generated by the authors of PRS-CS using the 1KG East Asian reference). We used PLINK v1.9 [37] for polygenic risk scoring in the GERA East Asian samples. GEMMA was used to analyze associations between the PRS and target phenotypes, accounting for relatedness and correcting for age (in 5-year categories), sex and the first ten PCs. Bonferroni correction was applied such that associations with  $p$  value <  $0.05/26 = 1.92 \times 10^{-3}$  are considered significant.

### Additional downstream analyses

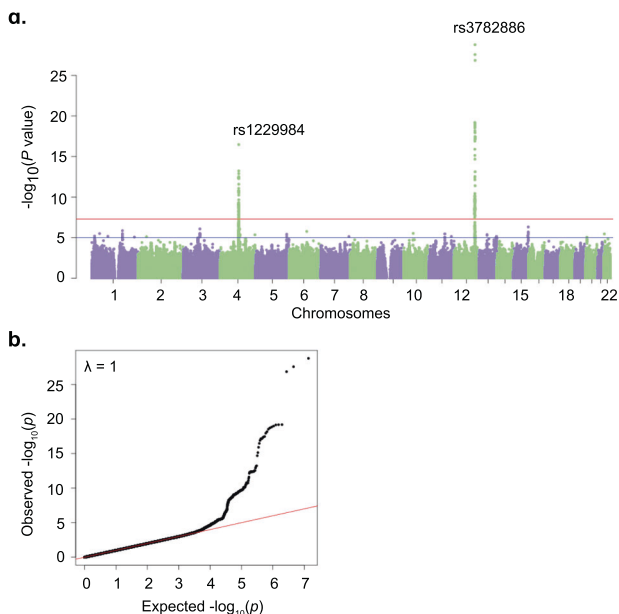
We used LD score regression [38] to estimate the SNP-based observed scale heritability of AUD using 1KG East Asians as the LD reference. We also

investigated the trans-ancestry genetic correlation between this study sample and PAU in EUR populations using Popcorn, a method that uses only summary-level data from GWAS while accounting for LD [39]. Trans-population meta-analysis between this study and PAU in EUR was conducted using METAL. Multi-trait analysis [40] was performed, which combined data from this study with excessive alcohol consumption defined as weekly intake >150 ml of alcohol for  $\geq 6$  months from the Taiwan Biobank [41].

## RESULTS

### Genome-wide association and meta-analyses

As in our previous study of AUD in East Asians [12], in a meta-analysis here of 2254 cases and 11,297 controls, we confirmed two loci that were significantly associated with AUD (Table 1 and Fig. 1). One locus is on chromosome 4q23 and includes multiple alcohol dehydrogenase genes. After LD clumping, there are two lead SNPs in this locus. The first is rs1229984 (Arg48His,  $p = 3.35 \times 10^{-17}$ , Fig. 2a) in *ADH1B* (Alcohol Dehydrogenase 1B (Class I), Beta Polypeptide), the second is rs1814125 ( $p = 2.14 \times 10^{-10}$ ) near *ADH1C*. Conditional analysis indicated that rs1814125 is not independent from rs1229984. For comparison, we also looked up the association of rs1229984 in other populations. Rs1229984 is also associated with PAU in European populations [14] (Fig. 2b). In African Americans from MVP, rs122994 is nominally significantly associated with AUD while another coding variant, rs2066702, is the lead SNP [8] at *ADH1B* (Fig. 2c). Another locus is a long region with high LD on chromosome 12 for which there is positive selection in East Asians [42], which includes *ALDH2* (Aldehyde Dehydrogenase 2) and *BRAP* (BRCA1 associated protein) genes. The lead SNP is rs3782886 ( $p = 1.68 \times 10^{-29}$ ), a coding variant in the *BRAP* gene. The previously reported functional coding variant rs671 in *ALDH2* is the second most significant SNP ( $p = 2.70 \times 10^{-28}$ ); it is not independent from rs3782886. No other independent associations were detected in this study. There are allele frequency differences among cohorts for these two lead SNPs, and moderate heterogeneity of effect sizes (the converted ORs) detected by the IVW meta-analyses (Fig. S1). The



**Fig. 1 Association results for AUD meta-analyses.** **a** Manhattan plot for AUD,  $n_{\text{case}} = 2254$ ,  $n_{\text{control}} = 11,297$ . Effective sample size-weighted meta-analyses were performed using METAL. Red line indicates genome-wide significant after correction for multiple testing ( $p < 5 \times 10^{-8}$ ), blue line indicates suggestive significant ( $p < 1 \times 10^{-5}$ ). **b** QQ plot for AUD.

heterogeneity  $p$  values are  $6.17 \times 10^{-13}$  for rs1229984 and is  $2.63 \times 10^{-5}$  for rs3782886 by IVW meta-analysis, justifying the use of effective sample size-weighted meta-analysis.

### Polygenic risk score for AUD

We calculated PRS for AUD in an independent East Asian cohort from the GERA cohort. We tested 26 phenotypes from survey and ICD-9-CM diagnosed conditions for association with the AUD PRS (Table 2). As expected, AUD PRS is significantly associated with alcohol consumption as measured in days per week of drinking (beta = 0.43, SE = 0.067,  $p = 2.47 \times 10^{-10}$ ). Also, AUD PRS is nominally significantly associated with pack years of smoking (beta = 0.09, SE = 0.05,  $p = 4.52 \times 10^{-2}$ ) and ever vs. never smoking (beta = 0.06, SE = 0.02,  $p = 1.14 \times 10^{-2}$ ), but these associations did not survive Bonferroni correction. None of the other traits in this small target sample were associated with AUD PRS.

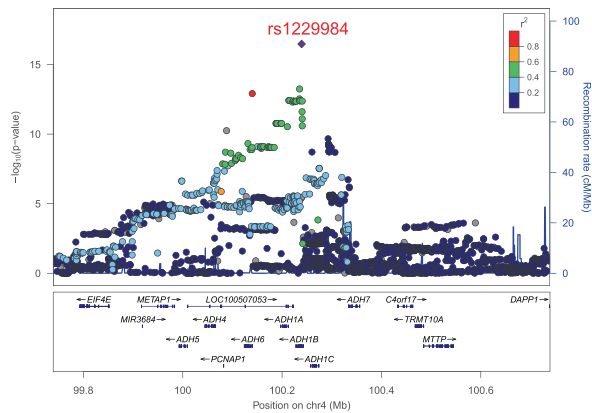
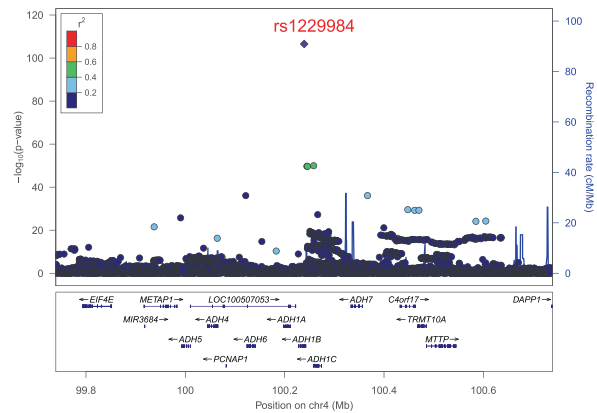
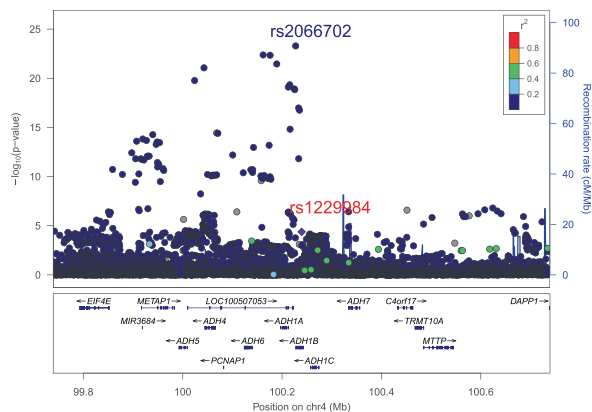
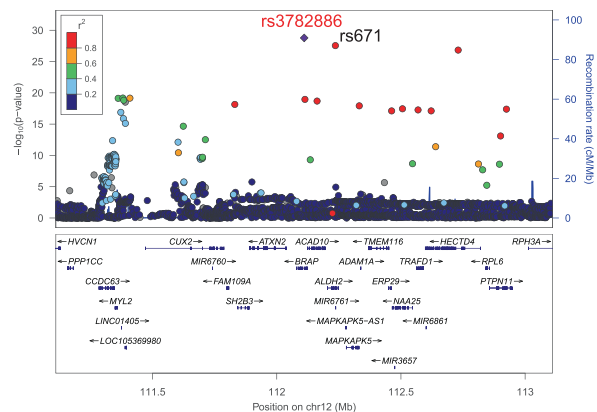
### Additional downstream analyses

SNP-based heritability of AUD was estimated to be 0.11 (SE = 0.07), which is not a significant estimate (probably due to the limited sample size). Genetic correlation between AUD in East Asian and PAU in European samples is 0.62 (SE = 0.23,  $p = 8.42 \times 10^{-3}$ ), showing moderate trans-ancestry genetic correlation. None of the trans-population meta-analysis between this study and PAU in EUR or multi-trait analysis with excessive alcohol consumption from Taiwan Biobank identified additional association signals.

## DISCUSSION

We collected data from 13,551 subjects with East Asian ancestry to conduct the largest meta-analysis to date for an alcohol-related trait in this population (quadruple the previous largest reported sample). We detected association signals at the *ADH1B* and *ALDH2* loci with substantially stronger statistical significance than has been seen previously, but did not identify any novel risk loci. This is mostly consistent with observations from other studies, where GWAS of alcohol-related traits with sample sizes in this range are generally underpowered to detect multiple replicable variants [4, 7, 43–46]. In EUR and African-Ancestry populations, the first and strongest associations detected have been at *ADH1B*. The *ALDH2* association is, to this date, unique to East Asians (rs671, a well-known functional *ALDH2* variant, is apparently unique to certain Asian populations [19]). It is a common issue for complex traits like AUD that many variants contribute to the heritability, each with a small effect size [47, 48]. Missing ancestral diversity in human genetic studies is a critical issue and recruitment of non-EUR subjects is crucial to addressing this disparity [16, 17]. The identification of *ALDH2*, which as noted is unique to Asians, exemplifies that there are differences in the genetic architecture of AUD between Asians and, for example, EUR, making well-powered investigations in this population an important scientific issue. Beyond identifying *ADH1B* and *ALDH2* with greater statistical significance than previous studies, the present investigation extends prior findings in several ways, including by examining the utility of the AUD PRS derived from this meta-analysis in an independent cohort of 4,464 East Asians and testing the association between the AUD PRS and alcohol, smoking, and other traits.

The two genes implicated—*ADH1B* and *ALDH2*—are involved in ethanol metabolism [48]. *ADH1B* encodes an alcohol dehydrogenase that oxidizes alcohol to acetaldehyde, which is then oxidized to acetate by aldehyde dehydrogenases, including that encoded by *ALDH2*. This is the major metabolic pathway for ethanol metabolism but other genes are involved as well. For example, in the first step, *ADH1C*, *ADH4* and *ADH7*, which map to the same chromosome 4 gene cluster as *ADH1B*, encode proteins that perform similar biological functions under certain conditions,

**a.** rs1229984 in this study.**b.** rs1229984 in European populations**c.** rs2066702 in African Americans**d.** rs3782886 in this study

**Fig. 2 Regional Manhattan plots for the top SNPs. a** Regional plot for rs1229984 in East Asians. **b** Regional plot for rs1229984 in European populations in a previous study (Zhou et al. [14]). **c** Regional plot for rs1229984 in African Americans from a previous MVP study (Kranzler et al. [8]) where rs1229984 is nominally significant, rs2066702 is the lead variant. In total, 500 kb in the upstream and downstream of rs1229984 were presented in a–c. **d** Regional plot for rs3782886 in East Asians. Given the high LD in this region, 1 Mb from both sides were extended.

*ALDH1A1* and *ALDH1B1* similarly encode proteins with roles that are sometimes overlapping with that of *ALDH2* [49]. Given the importance of other genes in the metabolic pathway, lead variants in genes other than *ADH1B*\*rs1229984 (EUR and Asian) and *ALDH2*\*rs671 (Asian) have been reported [6, 8, 12, 14]. Some of these associations are supported by conditional analyses [8, 14], and some appear to be variants that “hitchhike” with rs1229984 or rs671 due to their strong LD. Here, conditional analyses identified only one lead variant at each locus: rs1229984 ( $p = 3.35 \times 10^{-17}$ ) in the *AHD1B* region and rs3782886 ( $p = 1.68 \times 10^{-29}$ ) in the *ALDH2* region. The high LD between rs3782886 and rs671 ( $r^2 = 0.98$ ) makes it difficult to distinguish the real causal variant, though biochemical analysis favors rs671 (reviewed in [48]), which is nearly a null variant. A single copy of the rs671\*T allele renders the aldehyde dehydrogenase protein product nearly inactive and it is also more rapidly degraded, which causes flushing in East Asians and other associated symptoms that are protective against heavy drinking and AUD [50].

Rs3782886 in the *BRAP* gene (breast cancer suppressor protein (BRCA1)-associated protein) has been associated with many traits in East Asians, include alcohol-related traits [41, 51], myocardial infarction [52], and a biochemical trait—alanine aminotransferase level [53]. Some or all these associations could be due to the high LD with rs671 (as in this study), or reflect effects on activity of the metabolic pathway or cerebral cortical neurogenesis (argued in [41]). We would suggest that the different lead variants (rs671 or rs3782886) in this high LD region could reflect uncertainty introduced by different SNP arrays, imputation processes,

association analyses, or random variation in comparatively small samples. More data are needed to ascertain the true causal variant (or variants) despite the previous support and mechanistic appeal of rs671.

We used additional analyses to explore the genetic architecture of AUD in East Asians. The SNP-based heritability estimate was very low with a large standard error (SE), indicating a lack of statistical power. Moderate genetic correlation ( $r_g = 0.61$ ,  $SE = 0.23$ ,  $p = 8.42 \times 10^{-3}$ ) was detected between the main meta-analysis of this study and PAU in EUR populations, indicating shared genetic architecture across ancestries. However, the trans-population meta-analysis in which PAU in EUR was added did not detect any novel signals, probably due to the limited power in this study. Multi-trait analysis combining this study sample and excessive alcohol consumption from the Taiwan Biobank also identified no novel variants. Thus, additional study samples of East Asian ancestry are needed to provide adequate power for GWAS of AUD in East Asians.

Since it is a genetically complex trait, we expect that there are many variants that contribute to the genetic risk of AUD, consistent with findings in EUR [14]. Polygenic risk score analysis is a powerful tool for the application of GWAS results to investigate associations with traits of interest, which has been used widely in studies to test the association with AUD or related phenotypes in target cohorts [7, 8, 12, 14]. Here, we analyzed AUD PRS from our meta-analysis in an independent East Asian cohort from GERA, a US cohort collected to facilitate research on the genetic and environmental factors that affect health and disease [34]. We tested the

association between AUD PRS and 26 phenotypes in 4464 subjects of East Asian ancestry. AUD PRS was significantly associated with alcohol consumption as measured using days of drinking per week (see Table 2), and nominally significantly associated with pack years of smoking and ever vs. never smoking, consistent with the shared genetic architecture of AUD and alcohol and (possibly) smoking traits in East Asians. These same, or closely similar, associations, have been well established in EUR [8, 14]. There was no association detected between AUD PRS and other diseases or conditions in this study.

This study has limitations, the most important of which is the sample size, which despite being the largest reported so far for East Asian provides limited statistical power. Second, the phenotypes among the different study samples are not identical, with AUD diagnosed as ICD-9/10 codes in MVP and DSM-IV AD in other cohorts. This analytic approach is supported by the high genetic correlation between AUD and AD in EUR, which is estimated to approach 1.0 [14]. Third, some cohorts used alcohol-exposed controls, and others used unscreened controls (i.e., the MVP). Controls with demonstrated exposure to alcohol are ideal, but such exposure is commonplace in all the populations studied. Finally, although all of the cohorts are of East Asian ancestry, there are population differences among cohorts that increase heterogeneity and reduce power [54, 55]. These include cultural or environmental differences that affect trait prevalence (e.g., drinking practices), and geographical differences that introduce genetic differences (Fig. S1).

In conclusion, we conducted a GWAS of AUD in 13,551 East Asian subjects, in which we confirmed the two previously known risk loci and applied the AUD PRS in an independent cohort. Despite a large increment in sample size over the previous largest Asian-population GWAS, the power remains an important limitation. Accordingly, we will continue to recruit more East Asian subjects for alcohol studies and urge other investigators to do the same.

## DATA AVAILABILITY

Summary statistics are available on dbGaP (<https://www.ncbi.nlm.nih.gov/gap>) under study accession (phs001672).

## REFERENCES

- GBD Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390:1151–210.
- Hart AB, Kranzler HR. Alcohol dependence genetics: lessons learned from genome-wide association studies (GWAS) and post-GWAS analyses. *Alcohol Clin Exp Res*. 2015;39:1312–27.
- Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol Med*. 2015;45:1061–72.
- Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, et al. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry*. 2014;19:41–9.
- Quillen EE, Chen XD, Almasy L, Yang F, He H, Li X, 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*. 2014;165B:103–10.
- Gelernter J, Zhou H, Nuñez YZ, Mutirangura A, Malison RT, Kalayasiri R. Genome-wide association study of alcohol dependence and related traits in a Thai population. *Alcohol Clin Exp Res*. 2018;42:861–8.
- Walters RK, Polimanti R, Johnson EC, McClintick JN, Adams MJ, Adkins AE, et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci*. 2018;21:1656–69.
- Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun*. 2019;10:1499.
- Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, et al. ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol Psychiatry*. 2012;17:445–50.
- Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, et al. Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry*. 2009;66:773–84.
- Frank J, Cichon S, Treutlein J, Ridinger M, Mattheisen M, Hoffmann P, et al. Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict Biol*. 2012;17:171–80.
- Sun Y, Chang S, Wang F, Sun H, Ni Z, Yue W, et al. Genome-wide association study of alcohol dependence in male Han Chinese and cross-ethnic polygenic risk score comparison. *Transl Psychiatry*. 2019;9:249.
- Gelernter J, Polimanti R. Genetics of substance use disorders in the era of big data. *Nat Rev Genet*. 2021;22:712–29.
- Zhou H, Sealock JM, Sanchez-Roige S, Clarke TK, Levey DF, Cheng Z, et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat Neurosci*. 2020;23:809–18.
- Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100:635–49.
- Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell*. 2019;179:589–603.
- Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell*. 2019;177:26–31.
- Li D, Zhao H, Gelernter J. Strong association of the alcohol dehydrogenase 1B gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. *Biol Psychiatry*. 2011;70:504–12.
- Li D, Zhao H, Gelernter J. Strong protective effect of the aldehyde dehydrogenase gene (ALDH2) 504lys (\*2) allele against alcoholism and alcohol-induced medical diseases in Asians. *Hum Genet*. 2012;131:725–37.
- Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, et al. Diagnostic reliability of the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend*. 2005;80:303–12.
- Kalayasiri R, Verachai V, Gelernter J, Mutirangura A, Malison RT. Clinical features of methamphetamine-induced paranoia and preliminary genetic association with DBH-1021C→T in a Thai treatment cohort. *Addiction*. 2014;109:965–76.
- Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44:955–9.
- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–9.
- Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2:e190.
- Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods*. 2014;11:407–9.
- Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214–23.
- Hunter-Zinck H, Shi Y, Li M, Gorman BR, Ji SG, Sun N, et al. Measuring genetic variation in the multi-ethnic Million Veteran Program (MVP). *bioRxiv*. 2020. <https://doi.org/10.1101/2020.01.06.896613>
- Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet*. 2018;50:906–8.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*. 2010;26:2190–1.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82.
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–7.
- Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of summary statistics from linear mixed model association on all-or-none traits to odds ratio. *Genetics*. 2018;208:1397–408.
- Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics*. 2015;200:1285–95.
- Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006;34:D590–8.
- Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10:1776.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.

38. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J. Schizophrenia Working Group of the Psychiatric Genomics Consortium. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47:291–5.
39. Brown BC, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye CJ, Price AL, Zaitlen N. Transethnic genetic-correlation estimates from summary statistics. *Am J Hum Genet.* 2016;99:76–88.
40. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet.* 2018;50:229–237.
41. Chen IC, Kuo PH, Yang AC, Tsai SJ, Liu TH, Liu HJ, et al. CUX2, BRAP and ALDH2 are associated with metabolic traits in people with excessive alcohol consumption. *Sci Rep.* 2020;10:18118.
42. Oota H, Pakstis AJ, Bonne-Tamir B, Goldman D, Grigorenko E, Kajuna SL, et al. The evolution and population genetics of the ALDH2 locus: random genetic drift, selection, and low levels of recombination. *Ann Hum Genet.* 2004;68:93–109.
43. Schumann G, Coin LJ, Lourdasamy A, Charoen P, Berger KH, Stacey D, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc Natl Acad Sci USA.* 2011;108:7119–24.
44. Sanchez-Roige S, Fontanillas P, Elson SL, 23andMe Research Team, Gray JC, de Wit H, et al. Genome-wide association study of alcohol use disorder identification test (AUDIT) scores in 20 328 research participants of European ancestry. *Addict Biol.* 2019;24:121–31.
45. Jorgenson E, Thai KK, Hoffmann TJ, Sakoda LC, Kvale MN, Banda Y, et al. Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Mol Psychiatry.* 2017;22:1359–67.
46. Schumann G, Liu C, O'Reilly P, Gao H, Song P, Xu B, 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.* 2016;113:14372–7.
47. Zhang Y, Qi G, Park JH, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet.* 2018;50:1318–26.
48. Edenberg HJ, McClintick JN. Alcohol dehydrogenases, aldehyde dehydrogenases, and alcohol use disorders: a critical review. *Alcohol Clin Exp Res.* 2018;42:2281–97.
49. Vasilidou V, Pappa A, Estey T. Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. *Drug Metab Rev.* 2004;36:279–99.
50. Crabb DW, Edenberg HJ, Bosron WF, Li TK. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2(2) allele is dominant. *J Clin Invest.* 1989;83:314–6.
51. Kim JW, Choe YM, Shin JG, Park BL, Shin HD, Choi IG, et al. Associations of BRAP polymorphisms with the risk of alcohol dependence and scores on the alcohol use disorders identification test. *Neuropsychiatr Dis Treat.* 2019;15:83–94.
52. Ozaki K, Sato H, Inoue K, Tsunoda T, Sakata Y, Mizuno H, et al. SNPs in BRAP associated with risk of myocardial infarction in Asian populations. *Nat Genet.* 2009;41:329–33.
53. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet.* 2010;42:210–5.
54. Hugo Pan-Asian SNP Consortium, Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, et al. Mapping human genetic diversity in Asia. *Science.* 2009;326:1541–5.
55. Xu S, Yin X, Li S, Jin W, Lou H, Yang L, et al. Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am J Hum Genet.* 2009;85:762–74.

## ACKNOWLEDGEMENTS

This research used data from the Million Veteran Program, which was supported by funding from the Department of Veterans Affairs Office of Research and

Development, Million Veteran Program Grant #101BX003341 (ACJ, HRK). This publication does not represent the views of the Department of Veterans Affairs or the United States Government. Supported also by NIH (NIDA) R01 DA037974 (JG, RTM), (NCI) R21 CA252916 (HZ), and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (HZ). Data from the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) were accessed from dbGaP (phs000788.v2.p3). We thank the authors of the Taiwan Biobank excessive alcohol consumption study for sharing summary statistics.

## AUTHOR CONTRIBUTIONS

HZ and JG take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: HZ and JG. Acquisition of data: all authors. Analysis, or interpretation of data: HZ and JG. Drafting of the manuscript: HZ. Critical revision of the manuscript for important intellectual content: HZ, JG, and HRK. Obtained funding: HRK, ACJ, JG, RTM, and HZ. Final approval of the version to be published: all authors. Administrative, technical, or material support: YZN and JG. Study supervision: JG.

## COMPETING INTERESTS

HRK is a member of a scientific advisory board for Dicerna Pharmaceuticals and a consultant to Sophrosyne Pharmaceuticals and Sobrera Pharmaceuticals and is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. HRK and JG are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018. The remaining authors have nothing to disclose.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41386-022-01265-w>.

**Correspondence** and requests for materials should be addressed to Joel Gelernter.

**Reprints and permission information** is available at <http://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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022