

Genome-wide Meta-analysis of Alcohol Use Disorder in East Asians

Hang Zhou^{1,2}, Rasmon Kalayasiri^{3,4,5}, Yan Sun⁶, Yaira Z. Nuñez^{1,2}, Hong-Wen Deng⁷, Xiang-Ding Chen⁸, Amy C. Justice^{9,10}, Henry R. Kranzler^{11,12}, Suhua Chang¹³, Lin Lu^{6,13}, Jie Shi⁶, Kittipong Sanichwankul¹⁴, Apiwat Mutirangura¹⁵, Robert T. Malison^{1,‡}, Joel Gelernter^{1,2,16*}

¹Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA

²Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA

³Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

⁴Department of Psychiatry, King Chulalongkorn Memorial Hospital, Bangkok, Thailand

⁵Center for Excellence in Molecular Genetics of Cancer and Human Diseases, Department of Anatomy, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

⁶National Institute on Drug Dependence, Peking University, Beijing, China

⁷Center for Biomedical Informatics and Genomics, School of Medicine, Tulane University, New Orleans, LA, USA

⁸Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, Hunan, China

⁹Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

¹⁰Center for Interdisciplinary Research on AIDS, Yale School of Public Health, New Haven, CT, USA

¹¹University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

¹²Crescenz Veterans Affairs Medical Center, Philadelphia, PA, USA

¹³Peking 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

¹⁴Suanprung Psychiatric Hospital, Chiang Mai, Thailand

¹⁵Center of Excellence in Molecular Genetics of Cancer & Human Diseases, Department of Anatomy, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

¹⁶Departments of Genetics and Neuroscience, Yale University School of Medicine, New Haven, CT, USA

[‡]Deceased.

*Corresponding Author: Joel Gelernter, Department of Psychiatry, Yale School of Medicine, Veterans Affairs Connecticut Healthcare System, 116A2, 950 Campbell Ave, West Haven, CT 06516, USA. Phone: +1 (203) 494-6326 x3590; Fax: +1 (203) 937-4741; Email: joel.gelernter@yale.edu.

Running title: Alcohol use disorder in East Asians

Keywords: alcohol use disorder, East Asians, genome-wide association study, polygenic risk score

41 **Abstract**

42 BACKGROUND: Alcohol use disorder (AUD) is a leading cause of death and disability
43 worldwide. Genome-wide association studies (GWAS) have identified ~30 AUD risk genes in
44 European populations, but many fewer in East Asians.

45 METHODS: We conducted GWAS and genome-wide meta-analysis of AUD in 13,551 subjects
46 with East Asian ancestry, using published summary data and newly genotyped data from four
47 cohorts: 1) electronic health record (EHR)-diagnosed AUD in the Million Veteran Program (MVP)
48 sample; 2) DSM-IV diagnosed alcohol dependence (AD) in a Han Chinese-GSA (array) cohort;
49 3) AD in a Han Chinese-Cyto (array) cohort; and 4) two AD datasets in a Thai cohort. The MVP
50 and Thai samples included newly genotyped subjects from ongoing recruitment. In total, 2,254
51 cases and 11,297 controls were analyzed. An AUD polygenic risk score was analyzed in an
52 independent sample with 4,464 East Asians (Kaiser Permanente data from dbGaP).
53 Phenotypes from survey data and ICD-9-CM diagnoses were tested for association with the
54 AUD PRS.

55 RESULTS: Two risk loci were detected: the well-known functional variant rs1229984 in *ADH1B*
56 and rs3782886 in *BRAP* (near the *ALDH2* gene locus) are the lead variants. AUD PRS was
57 significantly associated with days per week of alcohol consumption ($\beta = 0.43$, $se = 0.067$, $p =$
58 2.47×10^{-10}) and nominally associated with pack years of smoking ($\beta = 0.09$, $se = 0.05$, $p =$
59 4.52×10^{-2}) and ever vs. never smoking ($\beta = 0.06$, $se = 0.02$, $p = 1.14 \times 10^{-2}$).

60 CONCLUSIONS: This is the largest GWAS of AUD in East Asians to date. Building on previous
61 findings, we were able to analyze pleiotropy, but did not identify any new risk regions,
62 underscoring the importance of recruiting additional East Asian subjects for alcohol GWAS.

63

64 **Introduction**

65 Globally, alcohol use disorders (AUD) are among the top causes of morbidity and
66 mortality [1]. Numerous factors predispose to the risk of developing AUD. Genetic factors
67 contribute substantial risk to the etiology of AUD [2], and the heritability has been estimated to
68 be ~0.5 in twin studies [3]. Genome-wide association studies (GWAS) of AUD have been
69 completed in multiple populations including European (EUR), African, Latin American and
70 Asians [4-11]. To date, the largest GWAS of problematic alcohol use (PAU, a proxy for AUD) in
71 435,563 EUR subjects identified 29 independent risk variants [12]. In contrast, the largest
72 GWAS of AUD in East Asians included less than 1% of this number: 3,381 subjects [13].
73 Genetic architecture often differs between populations; polygenic risk prediction between
74 populations, though sometimes useful, often is not transferable [14, 15]. Thus, it is critically
75 important that non-EUR populations be investigated to permit inferences to be made about
76 these ancestral populations, which represent the majority of the world's people [16, 17].

77 Because of the limited sample available and consequent lack of power for GWAS, little is
78 known about the genetic architecture of AUD in East Asians. The most consistent loci identified
79 are *ADH1B* (alcohol dehydrogenase 1b) and *ALDH2* (aldehyde dehydrogenase). Candidate
80 studies of *ADH1B**rs1229984 and *ALDH2**rs671 in East Asians showed strong association
81 between these functional variants and alcohol dependence [18, 19]. The first GWAS of alcohol
82 dependence (AD) in a Chinese sample was conducted in 102 male cases and 212 male
83 controls; rs3782886 in the *ALDH2* region was genome-wide significant (GWS) [5] despite the
84 very small sample size. The first GWAS of AD in Thai samples included 1,045 subjects and
85 identified rs149212747 in the *ALDH2* region as the lead variant [6]. The latest GWAS of AUD in
86 a Chinese cohort identified both *ADH1B* and *ALDH2* genes as risk loci [13]. However, only a
87 small proportion of the variance was explicable by variants in these genes. Larger samples are

88 required to identify more risk variants to provide a better understanding of the genetic
89 architecture in Asian populations.

90 Here we conducted a GWAS that combined 4 datasets from previously published
91 cohorts and newly genotyped subjects from Thai and MVP cohorts. In total, 13,551 subjects of
92 East Asian ancestry were analyzed, including 2,254 AUD cases. We then analyzed the resulting
93 AUD PRS in an independent East Asian sample for associations with 26 phenotypes from
94 surveys or ICD diagnoses. This GWAS of AUD is the largest to date in East Asians.

95

96 **Methods and Materials**

97 **Datasets**

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

105 The first stage included methamphetamine users hospitalized between 2007 and 2011
106 for 4 months of residential drug treatment (Thai METH-GSA, Table 1) [21]. DNA samples were
107 genotyped on the Illumina (San Diego, CA) Global Screening Array (GSA). Among the 863
108 genotyped subjects, we removed those with sample genotype call rate < 0.9, mismatched
109 genotypic and phenotypic sex, or excess heterozygosity rate [6]. Unlike for our prior report, here
110 we retained related subjects and applied linear mixed models to correct for relatedness (see
111 below). SNPs with genotype call rate ≥ 0.95 , minor allele frequency (MAF) ≥ 0.01 , or Hardy-

112 Weinberg Equilibrium (HWE) p-value $> 10^{-6}$ were kept for imputation. Imputation was done by
113 IMPUTE2 [22] with 1000 Genome project phase 3 data [23] as reference. SNPs with imputation
114 INFO score ≥ 0.8 were retained for association analyses. Principal component analysis (PCA)
115 was performed for the remaining subjects using EIGENSOFT [24, 25]. In contrast with our
116 previous study, here we used DSM-IV alcohol dependence (AD) to define case status, rather
117 than the DSM AUD criterion count to match the design in other cohorts. This yielded 127 cases
118 and 405 were exposed controls. Linear mixed models implemented in GEMMA [26] were used
119 to test association, with age, sex, and the first 10 PCs as covariates.

120 **Thai METH – MEGA.** Second-stage subjects (N = 3,161; the Thai METH-MEGA sample, Table
121 1) were recruited from 2015 to 2020 [6]. DNA samples were genotyped using the Illumina Multi-
122 Ethnic Global Array (MEGA). We removed subjects with sample genotype call rate < 0.95 , sex
123 mismatch, excess heterozygosity rate, or that were duplicates. SNPs with genotype call rate \geq
124 0.95, or MAF ≥ 0.01 , or HWE p-value $> 10^{-6}$ were retained for imputation as with the Thai
125 METH-GSA sample. We included 794 cases and 1576 alcohol-exposed controls in the
126 association analysis, which used GEMMA and age, sex and the first 10 PCs as covariates.

127 **MVP – EAA.** The Million Veteran Program (MVP) is an ongoing observational cohort study and
128 mega-biobank supported by the U.S. Department of Veterans Affairs (VA) [27, 28]. In October
129 2020, MVP released the latest genotype data (Release 4), which included 658,582 subjects.
130 Quality control was first done by the MVP Release 4 Data Team and included the removal of
131 duplicate DNA samples and those with sex mismatch, excessive heterozygosity, or a genotype
132 call rate < 0.985 . We ran PCA for the MVP subjects with 1000 Genome Project phase 3 (1KG)
133 as the reference, defining genetic groups based on the top 10 PCs [8]. For subjects grouped as
134 East Asian Americans (EAA), we ran a second PCA and removed outliers with PC scores > 6
135 standard deviations from the mean on any of the 10 PCs, yielding in 7,364 EAAs. Imputation [22]
136 was performed specifically for the EAAs using the 1KG as reference. SNPs with genotype call

137 rate ≥ 0.95 , MAF ≥ 0.01 , HWE $> 1 \times 10^{-6}$, and imputation INFO ≥ 0.8 were retained for analysis.
138 As for our prior study in EUR [12], subjects with ≥ 2 outpatient or ≥ 1 inpatient International
139 Classification of Diseases (ICD) codes for AUD were defined as cases (N = 701, Table 1) and
140 subjects with no AUD ICD code as controls (N = 6,254). BOLT-LMM [29] was used to correct for
141 relatedness, with age, sex, and the first 10 PCs as covariates.

142 **Han Chinese – Cyto.** This first GWAS of alcohol dependence, flushing response, and
143 maximum daily drinks consumed in a Han Chinese family sample [5] used the Illumine Cyto12
144 array (Table 1). Whereas the cohort was not imputed in the original report, we re-analyzed the
145 data and imputed the SNPs for an AD GWAS. Subjects with genotype call rate < 0.95 ,
146 duplicated DNA samples, mismatched sex, or excessive heterozygosity were removed, resulting
147 in 511 subjects for imputation. Imputation used IMPUTE2 and 1KG reference, SNPs with minor
148 allele frequency < 0.01 , genotype call rate < 0.95 , HWE p-value $< 1 \times 10^{-6}$, or imputation INFO $<$
149 0.8 were removed from further analyses. Due to the drinking practices and characteristic of this
150 particular population that result in few AD cases in females, only males were included in this
151 analysis. After QC, 99 DSM-IV-diagnosed male AD cases and 214 male alcohol-exposed
152 controls were analyzed using GEMMA to correct for relatedness, with age, sex and 10 PCs as
153 covariates.

154 **Han Chinese – GSA.** The second case-control AD GWAS in Han Chinese [13] included 533
155 cases and 2,848 alcohol-exposed controls who were genotyped using the GSA array (Table 1).
156 Here we used the summary statistics from this study.

157

158 **Meta-analysis**

159 Using association analyses or summary statistics for each of the four cohorts, effective sample-
160 size weighted meta-analysis was performed using METAL [30]. SNPs present in only one cohort

161 or in less than 15% of the total samples were removed (6.8 million SNPs remained). To define
162 lead variants, the meta-analysis summary data were clumped by LD with $r^2 < 0.1$ in a 2500-kb
163 window, using 1KG East Asians as the LD reference. Additional conditional analyses [31] were
164 performed to define independent variants when more than one lead variant was present in the
165 same locus (e.g., the chr4 ADH gene region).

166

167 **Polygenic Risk Scores**

168 **Target dataset.** We requested and downloaded from dbGaP (phs000788.v2.p3) data from the
169 Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) Genetic
170 Epidemiology Research on Adult Health and Aging (GERA) cohort. This large and ethnically
171 diverse cohort genotype data from 5,182 self-reported Asians using a custom Affymetrix Axiom
172 array [32]. All subjects completed a broad written consent.

173 **Imputation.** Subjects with mismatched sex or genotype call rate < 0.95 were removed. The
174 genomic build was transferred from 36 to 37 using LiftOver [33]. As we did for MVP, we ran
175 PCA for the 5,182 Asian subjects using the 1KG as reference, clustering them into different
176 groups. A second PCA among Asians was used to remove outliers, resulting in 4,464
177 genetically classified East Asians for imputation. Imputation was performed using IMPUTE2 and
178 1KG reference, SNPs with minor allele frequency < 0.01 , genotype call rate < 0.95 , HWE p-
179 value $< 1 \times 10^{-6}$, or imputation INFO < 0.8 were removed from further analysis.

180 **Target phenotypes.** Two sources of phenotypes are included in this study. The first is survey
181 data on physical observations, lifestyle and environment, including phenotypes such as BMI,
182 general health, physical activity, alcohol use, smoking status and pack years. The second is
183 ICD-9-CM. disease and conditions measures. Participant were coded as cases if there were at

184 least two diagnoses in a disease category. Binary phenotypes with less than 100 cases were
185 removed from analyses. See Table 2 for details of the target phenotypes.

186 **Polygenic risk scoring and association.** PRS-CS [34] was used to infer posterior effect sizes
187 of SNPs using GWAS summary statistics for AUD from this study, and an external East Asian
188 LD reference panel (generated by the authors of PRS-CS using the 1KG phase3 East Asian
189 reference). We used PLINK v1.9 [35] for polygenic risk scoring in the target East Asian samples.
190 GEMMA was used analyze associations between the PRS and target phenotypes, accounting
191 for relatedness and correcting for age (in 5-year categories), sex and the first 10 PCs.

192

193 **Additional downstream analyses**

194 We used LD score regression [36] to estimate the SNP-based heritability of AUD using 1KG
195 East Asians as the LD reference. We also investigated the trans-ancestry genetic correlation
196 between this study sample and PAU in EUR populations using popcorn, a method that uses
197 only summary-level data from GWAS while accounting for LD [37]. Trans-population meta-
198 analysis between this study and PAU in EUR was conducted using METAL. Multi-trait analysis
199 [38] was performed, which combined data from this study with excessive alcohol consumption
200 defined as weekly intake > 150 mL of alcohol for ≥ 6 months from the Taiwan Biobank [39]

201

202 **Results**

203 **Genome-wide association and meta-analyses**

204 As in our previous study of AUD in East Asians [13], in a meta-analysis here of 2,254 cases and
205 11,297 controls, we confirmed two loci that were significantly associated with AUD (Table 1,
206 Figure 1). One locus is on chromosome 4q23 and includes multiple alcohol dehydrogenase

207 genes. After LD clumping, there are two lead SNPs in this locus. The first is rs1229984
208 (Arg48His, $p = 3.35 \times 10^{-17}$) in *ADH1B* (Alcohol Dehydrogenase 1B (Class I), Beta Polypeptide),
209 the second is rs1814125 ($p = 2.14 \times 10^{-10}$) near *ADH1C*. Conditional analysis indicated that
210 rs1814125 is not independent from rs1229984. Another locus is a long region with high LD on
211 chromosome 12 for which there is positive selection in East Asians, which covers *ALDH2*
212 (Aldehyde Dehydrogenase 2) and *BRAP* (BRCA1 associated protein) genes. The lead SNP is
213 rs3782886 ($p = 1.68 \times 10^{-29}$), a coding variant in the *BRAP* gene. No other associations were
214 detected in this study.

215

216 **Polygenic risk score for AUD**

217 We calculated PRS for AUD in an independent East Asian cohort from the GERA cohort. We
218 tested 26 phenotypes from survey and ICD-9-CM diagnosed conditions for association with the
219 AUD PRS (Table 2). As expected, AUD PRS is significantly associated with alcohol
220 consumption as measured in days per week of drinking (beta = 0.43, se = 0.067, $p = 2.47 \times 10^{-10}$).
221 Also, AUD PRS is nominally significantly associated with pack years of smoking (beta = 0.09, se
222 = 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
223 these associations did not survive multiple testing correction. None of the other traits in this
224 small target sample were associated with AUD PRS. None of the additional downstream
225 analyses provided significant results nor did they identify additional association signals.

226

227 **Discussion**

228 We collected data from 13,551 subjects with East Asian ancestry to conduct the largest
229 meta-analysis to date for an alcohol-related trait in this population (quadruple the previous
230 largest reported sample). We detected association signals at the *ADH1B* and *ALDH2* loci with

231 substantially stronger statistical significance than has been seen previously, but did not identify
232 any novel risk loci. This is mostly consistent with observations from other studies, where GWAS
233 of alcohol-related traits with sample sizes in this range are generally underpowered to detect
234 multiple replicable variants [4, 7, 40-43]. In EUR and African-Ancestry populations, the first and
235 strongest associations detected have been at *ADH1B*. The *ALDH2* association is, to this date,
236 unique to Asians. It is a common issue for complex traits like AUD that many variants contribute
237 to the heritability, each with a small effect size [44, 45]. Missing ancestral diversity in human
238 genetic studies is a critical issue and recruitment of non-EUR subjects is crucial to addressing
239 this disparity [16, 17]. The identification of *ALDH2*, which as noted is unique to Asians,
240 exemplifies that there are differences in the genetic architecture of AUD between Asians and,
241 for example, EUR, making well-powered investigations in this population an important scientific
242 issue. Beyond identifying *ADH1B* and *ALDH2* with greater statistical significance than previous
243 studies, the present investigation extends prior findings in several ways, including by examining
244 the utility of the AUD PRS derived from this meta-analysis in an independent cohort of 4,464
245 East Asians and testing the association between the AUD PRS and alcohol, smoking, and other
246 traits.

247 The two genes implicated – *ADH1B* and *ALDH2* – are involved in ethanol metabolism
248 [45]. *ADH1B* encodes an alcohol dehydrogenase that oxidizes alcohol to acetaldehyde, which is
249 then oxidized to acetate by aldehyde dehydrogenases, including that encoded by *ALDH2*. This
250 is the major metabolic pathway for ethanol metabolism but other genes are involved as well. For
251 example, in the first step, *ADH1C*, *ADH4* and *ADH7*, which map to the same chromosome 4
252 gene cluster as *ADH1B*, encode proteins that perform similar biological functions under certain
253 conditions, *ALDH1A1* and *ALDH1B1* similarly encode proteins with roles that are sometimes
254 overlapping with that of *ALDH2* [46]. Given the importance of other genes in the metabolic
255 pathway, lead variants in genes other than *ADH1B**rs1229984 (EUR and Asian) and

256 *ALDH2**rs671 (Asian) have been reported [6, 8, 12, 13]. Some of these associations are
257 supported by conditional analyses [8, 12], and some appear to be variants that "hitchhike" with
258 rs1229984 or rs671 due to their strong LD. Here, conditional analyses identified only one lead
259 variant at each locus: rs1229984 ($p = 3.35 \times 10^{-17}$) in the *AHD1B* region and rs3782886 ($p =$
260 1.68×10^{-29}) in the *ALDH2* region. The high LD between rs3782886 and rs671 ($r^2 = 0.98$) makes
261 it difficult to distinguish the real causal variant, though biochemical analysis favors rs671
262 (reviewed in [45]), which is nearly a null variant.

263 Rs3782886 in the *BRAP* gene (breast cancer suppressor protein (BRCA1)-associated
264 protein) has been associated with many traits in East Asians, include alcohol-related traits [39,
265 47], myocardial infarction [48], and a biochemical trait – alanine aminotransferase level [49].
266 Some or all these associations could be due to the high LD with rs671 (as in this study), or
267 reflect effects on activity of the metabolic pathway or cerebral cortical neurogenesis (argued in
268 [39]). We would suggest that the different lead variants (rs671 or rs3782886) in this high LD
269 region could reflect uncertainty introduced by different SNP arrays, imputation processes,
270 association analyses, or random variation in comparatively small samples. More data are
271 needed to ascertain the true causal variant despite the previous support and mechanistic appeal
272 of rs671.

273 We used additional analyses to explore the genetic architecture of AUD in East Asians.
274 The SNP-based heritability estimate was very low with a large standard error (SE), indicating a
275 lack of statistical power. Trans-ancestry genetic correlation between this study and PAU in EUR
276 populations is 0.61 (se = 0.23, $p = 0.10$). The trans-population meta-analysis in which PAU in
277 EUR was added did not detect any novel signals. Multi-trait analysis combining this study
278 sample and excessive alcohol consumption from the Taiwan Biobank also identified no novel
279 variants. Thus, additional study samples of European ancestry are needed to provide adequate
280 power for GWAS of AUD in East Asians.

281 Since it is a genetically complex trait, we expect that there are many variants that
282 contribute to the genetic risk of AUD, consistent with findings in EUR [12]. Polygenic risk score
283 analysis is a powerful tool for the application of GWAS results to investigate associations with
284 traits of interest, which has been used widely in studies to test the association with AUD or
285 related phenotypes in target cohorts [7, 8, 12, 13]. Here, we analyzed AUD PRS from our meta-
286 analysis in an independent East Asian cohort from GERA, a US cohort collected to facilitate
287 research on the genetic and environmental factors that affect health and disease [32]. We
288 tested the association between AUD PRS and 26 phenotypes in 4,464 subjects of East Asian
289 ancestry. AUD PRS was significantly associated with alcohol consumption as measured using
290 days of drinking per week (see Table 2), and nominally significantly associated with pack years
291 of smoking and ever vs. never smoking, consistent with the shared genetic architecture of AUD
292 and alcohol and (possibly) smoking traits in East Asians. These same, or closely similar,
293 associations, have been well established in EUR [8, 12]. There was no association detected
294 between AUD PRS and other diseases or conditions in this study.

295 This study has limitations, the most important of which is the sample size, which despite
296 being the largest reported so far for East Asian provides limited statistical power. Second, the
297 phenotypes among the different study samples are not identical, with AUD diagnosed as ICD-
298 9/10 codes in MVP and DSM-IV AD in other cohorts. This analytic approach is supported by the
299 high genetic correlation between AUD and AD in EUR, which is estimated to approach 1.0 [12].
300 Third, some cohorts used alcohol-exposed controls, and others used unscreened controls (i.e.,
301 the MVP). Controls with demonstrated exposure to alcohol are ideal, but such exposure is
302 commonplace in all the populations studied. Finally, although all of the cohorts are of East Asian
303 ancestry, there are population differences among cohorts that increase heterogeneity and
304 reduce power. These include cultural or environmental differences that affect trait prevalence
305 (e.g., drinking practices), and geographical differences that introduce genetic differences.

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

311

312 **Acknowledgements and Disclosures**

313 This research used data from the Million Veteran Program, which was supported by funding
314 from the Department of Veterans Affairs Office of Research and Development, Million Veteran
315 Program Grant #I01BX003341. This publication does not represent the views of the Department
316 of Veterans Affairs or the United States Government. Supported also by NIH (NIDA) R01
317 DA037974 (JG), and a NARSAD Young Investigator Grant from the Brain & Behavior Research
318 Foundation (HZ). Data from the Kaiser Permanente Research Program on Genes, Environment,
319 and Health (RPGEH) were accessed from dbGaP (phs000788.v2.p3).

320 Disclosure: Dr. Kranzler is a member of a scientific advisory board for Dicerna Pharmaceuticals
321 and a consultant to Sophrosyne Pharmaceuticals and Sobrera Pharmaceuticals and is a
322 member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials
323 Initiative, which was supported in the last three years by AbbVie, Alkermes, Ethypharm, Indivior,
324 Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. Drs. Kranzler and
325 Gelernter are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-
326 guided dosing of opioid agonists," filed January 24, 2018.

327

328

329 **Table 1. Sample characteristics.**

Cohorts	Traits	N (%female)	Mean (SD), Age	# Cases
Thai METH – GSA	DSM-IV AD	532 (49.4)	26.6 (6.9)	127
Thai METH – MEGA	DSM-IV AD	2,370 (42.5)	34.7 (10.1)	794
MVP – EAA	ICD 9/10 AUD	6,955 (10.7)	53.4 (17.1)	701
Han Chinese – GSA	DSM-IV AD	3,381 (29.9)	34.2 (8.3)	533
Han Chinese – Cyto	DSM-IV AD	313 (0)	49.6 (14.7)	99
Total		13,551		2,254

330

331

332

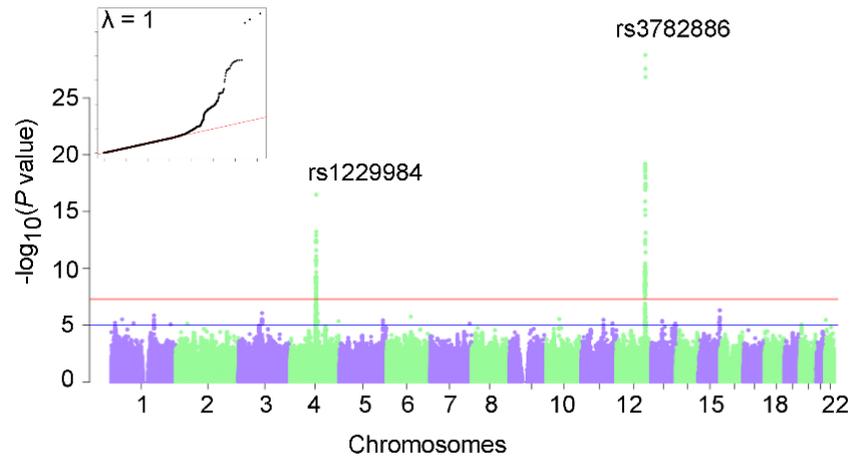
333 **Table 2. Tested phenotypes in GERA and association results with AUD PRS.** If not
 334 specified for distribution, 1 is case and 0 is control.

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

335
 336 Note: ^aCancer: includes malignant tumors, neoplasms, lymphoma and sarcoma. ^bHeart disease:
 337 includes ischemic heart disease, cardiac arrest, congestive heart failure, dysrhythmias,
 338 cardiomyopathy, aortic aneurysm, and cerebrovascular disease, but excludes PVD which is
 339 encompassed by the PVD variable. ^cDays of alcohol intake per week, 1 is no days, 2 is 1 day, 3
 340 is 2-4 days, 4 is 5-6 days, 5 is every day. ^dPack-years for former or current smoker, 0 = 0, 1 <
 341 10, 2 = 10-20, 3 = 20-30, 4 ≥ 30. ^ePhysical activity total metabolic equivalency of task (MET), 1
 342 = first quartile, 0-173 for males and 0-74 for females, 2 = second quartile, 174-600 for males
 343 and 75-344 for females, 3 = third quartile, 601-1380 for males and 345-983 for females, 4 =
 344 fourth quartile, 1381+ for males and 984+ for females. ^fHealth status, 1 = excellent, 2 = very
 345 good, 3 = good, 4 = fair.

346

347 **Figure 1. Manhattan plot of the meta-analysis of AUD in 13,551 East Asians.**



348

349 References

- 350 1. G. B. D. Causes of Death Collaborators, *Global, regional, and national age-sex specific mortality*
351 *for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study*
352 *2016*. *Lancet*, 2017. **390**(10100): p. 1151-1210.
- 353 2. Hart, A.B. and H.R. Kranzler, *Alcohol Dependence Genetics: Lessons Learned From Genome-Wide*
354 *Association Studies (GWAS) and Post-GWAS Analyses*. *Alcohol Clin Exp Res*, 2015. **39**(8): p. 1312-
355 27.
- 356 3. Verhulst, B., M.C. Neale, and K.S. Kendler, *The heritability of alcohol use disorders: a meta-*
357 *analysis of twin and adoption studies*. *Psychol Med*, 2015. **45**(5): p. 1061-72.
- 358 4. Gelernter, J., et al., *Genome-wide association study of alcohol dependence: significant findings in*
359 *African- and European-Americans including novel risk loci*. *Mol Psychiatry*, 2014. **19**(1): p. 41-9.
- 360 5. Quillen, E.E., et al., *ALDH2 is associated to alcohol dependence and is the major genetic*
361 *determinant of "daily maximum drinks" in a GWAS study of an isolated rural Chinese sample*. *Am*
362 *J Med Genet B Neuropsychiatr Genet*, 2014. **165B**(2): p. 103-10.
- 363 6. Gelernter, J., et al., *Genomewide Association Study of Alcohol Dependence and Related Traits in*
364 *a Thai Population*. *Alcohol Clin Exp Res*, 2018. **42**(5): p. 861-868.
- 365 7. Walters, R.K., et al., *Transancestral GWAS of alcohol dependence reveals common genetic*
366 *underpinnings with psychiatric disorders*. *Nat Neurosci*, 2018. **21**(12): p. 1656-1669.
- 367 8. Kranzler, H.R., et al., *Genome-wide association study of alcohol consumption and use disorder in*
368 *274,424 individuals from multiple populations*. *Nat Commun*, 2019. **10**(1): p. 1499.
- 369 9. Bierut, L.J., et al., *ADH1B is associated with alcohol dependence and alcohol consumption in*
370 *populations of European and African ancestry*. *Mol Psychiatry*, 2012. **17**(4): p. 445-50.
- 371 10. Treutlein, J., et al., *Genome-wide association study of alcohol dependence*. *Arch Gen Psychiatry*,
372 2009. **66**(7): p. 773-84.
- 373 11. Frank, J., et al., *Genome-wide significant association between alcohol dependence and a variant*
374 *in the ADH gene cluster*. *Addict Biol*, 2012. **17**(1): p. 171-80.
- 375 12. Zhou, H., et al., *Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals*
376 *yields insights into biology and relationships with other traits*. *Nat Neurosci*, 2020. **23**(7): p. 809-
377 818.
- 378 13. Sun, Y., et al., *Genome-wide association study of alcohol dependence in male Han Chinese and*
379 *cross-ethnic polygenic risk score comparison*. *Transl Psychiatry*, 2019. **9**(1): p. 249.
- 380 14. Martin, A.R., et al., *Human Demographic History Impacts Genetic Risk Prediction across Diverse*
381 *Populations*. *Am J Hum Genet*, 2017. **100**(4): p. 635-649.
- 382 15. Gelernter, J. and R. Polimanti, *Genetics of substance use disorders in the era of big data*. *Nature*
383 *Reviews Genetics*, 2021.
- 384 16. Peterson, R.E., et al., *Genome-wide Association Studies in Ancestrally Diverse Populations:*
385 *Opportunities, Methods, Pitfalls, and Recommendations*. *Cell*, 2019. **179**(3): p. 589-603.
- 386 17. Sirugo, G., S.M. Williams, and S.A. Tishkoff, *The Missing Diversity in Human Genetic Studies*. *Cell*,
387 2019. **177**(1): p. 26-31.
- 388 18. Li, D., H. Zhao, and J. Gelernter, *Strong association of the alcohol dehydrogenase 1B gene*
389 *(ADH1B) with alcohol dependence and alcohol-induced medical diseases*. *Biol Psychiatry*, 2011.
390 **70**(6): p. 504-12.
- 391 19. Li, D., H. Zhao, and J. Gelernter, *Strong protective effect of the aldehyde dehydrogenase gene*
392 *(ALDH2) 504lys (*2) allele against alcoholism and alcohol-induced medical diseases in Asians*.
393 *Hum Genet*, 2012. **131**(5): p. 725-37.

- 394 20. Pierucci-Lagha, A., et al., *Diagnostic reliability of the Semi-structured Assessment for Drug*
395 *Dependence and Alcoholism (SSADDA)*. Drug Alcohol Depend, 2005. **80**(3): p. 303-12.
- 396 21. Kalayasiri, R., et al., *Clinical features of methamphetamine-induced paranoia and preliminary*
397 *genetic association with DBH-1021C-->T in a Thai treatment cohort*. Addiction, 2014. **109**(6): p.
398 965-76.
- 399 22. Howie, B., et al., *Fast and accurate genotype imputation in genome-wide association studies*
400 *through pre-phasing*. Nat Genet, 2012. **44**(8): p. 955-9.
- 401 23. 1000 Genomes Project Consortium, et al., *A global reference for human genetic variation*.
402 Nature, 2015. **526**(7571): p. 68-74.
- 403 24. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide*
404 *association studies*. Nat Genet, 2006. **38**(8): p. 904-9.
- 405 25. Patterson, N., A.L. Price, and D. Reich, *Population structure and eigenanalysis*. PLoS Genet, 2006.
406 **2**(12): p. e190.
- 407 26. Zhou, X. and M. Stephens, *Efficient multivariate linear mixed model algorithms for genome-wide*
408 *association studies*. Nat Methods, 2014. **11**(4): p. 407-9.
- 409 27. Gaziano, J.M., et al., *Million Veteran Program: A mega-biobank to study genetic influences on*
410 *health and disease*. J Clin Epidemiol, 2016. **70**: p. 214-23.
- 411 28. Hunter-Zinck, H., et al., *Measuring genetic variation in the multi-ethnic Million Veteran Program*
412 *(MVP)*. doi: <https://doi.org/10.1101/2020.01.06.896613>, 2020.
- 413 29. Loh, P.R., et al., *Mixed-model association for biobank-scale datasets*. Nat Genet, 2018. **50**(7): p.
414 906-908.
- 415 30. Willer, C.J., Y. Li, and G.R. Abecasis, *METAL: fast and efficient meta-analysis of genomewide*
416 *association scans*. Bioinformatics, 2010. **26**(17): p. 2190-1.
- 417 31. Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis*. Am J Hum Genet, 2011.
418 **88**(1): p. 76-82.
- 419 32. Banda, Y., et al., *Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the*
420 *Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort*. Genetics, 2015. **200**(4):
421 p. 1285-95.
- 422 33. Hinrichs, A.S., et al., *The UCSC Genome Browser Database: update 2006*. Nucleic Acids Res, 2006.
423 **34**(Database issue): p. D590-8.
- 424 34. Ge, T., et al., *Polygenic prediction via Bayesian regression and continuous shrinkage priors*. Nat
425 Commun, 2019. **10**(1): p. 1776.
- 426 35. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets*.
427 Gigascience, 2015. **4**: p. 7.
- 428 36. Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in*
429 *genome-wide association studies*. Nat Genet, 2015. **47**(3): p. 291-5.
- 430 37. Brown, B.C., et al., *Transethnic Genetic-Correlation Estimates from Summary Statistics*. Am J
431 Hum Genet, 2016. **99**(1): p. 76-88.
- 432 38. Turley, P., et al., *Multi-trait analysis of genome-wide association summary statistics using MTAG*.
433 Nat Genet, 2018. **50**(2): p. 229-237.
- 434 39. Chen, I.C., et al., *CUX2, BRAP and ALDH2 are associated with metabolic traits in people with*
435 *excessive alcohol consumption*. Sci Rep, 2020. **10**(1): p. 18118.
- 436 40. Schumann, G., et al., *Genome-wide association and genetic functional studies identify autism*
437 *susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption*. Proc Natl Acad
438 Sci U S A, 2011. **108**(17): p. 7119-24.
- 439 41. Sanchez-Roige, S., et al., *Genome-wide association study of alcohol use disorder identification*
440 *test (AUDIT) scores in 20 328 research participants of European ancestry*. Addict Biol, 2019. **24**(1):
441 p. 121-131.

- 442 42. Jorgenson, E., et al., *Genetic contributors to variation in alcohol consumption vary by*
443 *race/ethnicity in a large multi-ethnic genome-wide association study*. Mol Psychiatry, 2017.
444 **22**(9): p. 1359-1367.
- 445 43. Schumann, G., et al., *KLB is associated with alcohol drinking, and its gene product beta-Klotho is*
446 *necessary for FGF21 regulation of alcohol preference*. Proc Natl Acad Sci U S A, 2016. **113**(50): p.
447 14372-14377.
- 448 44. Zhang, Y., et al., *Estimation of complex effect-size distributions using summary-level statistics*
449 *from genome-wide association studies across 32 complex traits*. Nat Genet, 2018. **50**(9): p. 1318-
450 1326.
- 451 45. Edenberg, H.J. and J.N. McClintick, *Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and*
452 *Alcohol Use Disorders: A Critical Review*. Alcohol Clin Exp Res, 2018. **42**(12): p. 2281-2297.
- 453 46. Vasiliou, V., A. Pappa, and T. Estey, *Role of human aldehyde dehydrogenases in endobiotic and*
454 *xenobiotic metabolism*. Drug Metab Rev, 2004. **36**(2): p. 279-99.
- 455 47. Kim, J.W., et al., *Associations of BRAP polymorphisms with the risk of alcohol dependence and*
456 *scores on the Alcohol Use Disorders Identification Test*. Neuropsychiatr Dis Treat, 2019. **15**: p. 83-
457 94.
- 458 48. Ozaki, K., et al., *SNPs in BRAP associated with risk of myocardial infarction in Asian populations*.
459 Nat Genet, 2009. **41**(3): p. 329-33.
- 460 49. Kamatani, Y., et al., *Genome-wide association study of hematological and biochemical traits in a*
461 *Japanese population*. Nat Genet, 2010. **42**(3): p. 210-5.
- 462