Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders

Raymond K. Walters[1], Mark J. Adams[2], Amy E. Adkins[3], Fazil Aliev[4], Silviu-Alin Bacanu[5], Anthony Batzler[6], Sarah Bertelsen[7], Joanna M. Biernacka[8], Tim B. Bigdeli[9], Li-Shiun Chen[10], Toni-Kim Clarke[2], Yi-Ling Chou[10], Franziska Degenhardt[11], Anna R. Docherty[12], Pierre Fontanillas[13], Jerome Foo[14], Louis Fox[10], Josef Frank[14], Ina Giegling[15], Scott Gordon[16], Laura M. Hack[17], Annette M. Hartmann[15], Sarah M. Hartz[10], Stefanie Heilmann-Heimbach[11], Stefan Herms[11,18], Colin Hodgkinson[19] , Per Hoffmann[11,18], Jouke Jan Hottenga[20], Martin A. Kennedy[21], Mervi Alanne-Kinnunen[22], Bettina Konte[15], Jari Lahti[23,24], Marius Lahti-Pulkkinen[24], Lannie Ligthart[20], Anu Loukola[22], Brion S. Maher[25], Hamdi Mbarek[20], Andrew M. McIntosh[26], Matthew B. McQueen[27], Yuri Milaneschi[28], Teemu Palviainen[22], John F. Pearson[29], Roseann E. Peterson[30], Renato Polimanti[31], Samuli Ripatti[22], Euijung Ryu[32], Nancy L. Saccone[33], Jessica E. Salvatore[4,30], Sandra Sanchez Roige[34], Melanie Schwandt[35], Richard Sherva[36], Fabian Streit[14], Jana Strohmaier[14], Nathaniel Thomas[3], Jen-Chyong Wang[7], Bradley T. Webb[5], Robbee Wedow[37], Leah Wetherill[38], Amanda G. Wills[39], 23andMe Research Team[# 13], Jason D. Boardman[37], Danfeng Chen[40], Doo-Sup Choi[41], William E. Copeland[42], Robert C. Culverhouse[43], Norbert Dahmen[44], Louisa Degenhardt[45], Benjamin W. Domingue[46], Sarah L. Elson[13], Mark Frye[47], Wolfgang Gäbel[48], Marcus Ising[49], Emma C. Johnson[10], Margaret Keyes[50], Falk Kiefer[51], John Kramer[52], Samuel Kuperman[52], Susanne Lucae[49], Michael T. Lynskey[53], Wolfgang Maier[54], Karl Mann[51], Satu Männistö[55], Jeanette Nance McClintick[56], Jacquelyn L. Meyers[57], Bertram Müller-Myhsok[58], John I. Nurnberger[38,59], Aarno Palotie[1,22,60] Ulrich Preuss[15,61], Katri Räikkönen[24], Maureen D. Reynolds[62], Monika Ridinger[63], Norbert Scherbaum[64], Marc Schuckit[34], Michael Soyka[65,66], Jens Treutlein[14], Stephanie Witt[14], Norbert Wodarz[67], Peter Zill[66], Daniel E. Adkins[12,68], Joseph M. Boden[21], Dorret Boomsma[20], Laura J. Bierut[10], Sandra A. Brown[34,69], Kathleen K. Bucholz[10], Sven Cichon[18], E. Jane Costello[42], Harriet de Wit[70], Nancy Diazgranados[71] , Danielle M. Dick[3,72], Johan G. Eriksson[73], Lindsay A. Farrer[36,74], Tatiana M. Foroud[38], Nathan A. Gillespie[30], Alison M. Goate[7], David Goldman[19,35] , Richard A. Grucza[10] , Dana B. Hancock[75], Kathleen Mullan Harris[76], Andrew C. Heath[10], Victor Hesselbrock[77], John K. Hewitt[78], Christian Hopfer[79], John Horwood[21], William Iacono[50], Eric O. Johnson[80], Jaakko A. Kaprio[22,81], Victor Karpyak[47], Kenneth S. Kendler[5], Henry R. Kranzler[82], Kenneth Krauter[83], Paul Lichtenstein[84], Penelope A. Lind[16], Matt McGue[50], James MacKillop[85], Pamela A. F. Madden[10], Hermine Maes[86], Patrik Magnusson[84], Nicholas G. Martin[16], Sarah E. Medland[16], Grant W. Montgomery[87], Elliot C. Nelson[10], Markus M. Nöthen[88], Abraham A. Palmer[34,89], Nancy L. Pedersen[84], Brenda Penninx[28], Bernice Porjesz[57], John P. Rice[10], Marcella Rietschel[14], Brien P. Riley[5], Richard Rose[90], Dan Rujescu[15], Pei-Hong Shen[19] , Judy Silberg[30], Michael C. Stallings[78], Ralph E. Tarter[62], Michael M. Vanyukov[62], Scott Vrieze[50], Tamara L. Wall[34], John B. Whitfield[16], Hongyu Zhao[91], Benjamin M. Neale[1], Joel Gelernter[92*], Howard J. Edenberg[38,56*], Arpana Agrawal[10*]

*Corresponding authors: Joel Gelernter (joel.gelernter@yale.edu); Howard Edenberg (edenberg@iu.edu); Arpana Agrawal (Arpana@wustl.edu)

[1]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[2] University of Edinburgh, Division of Psychiatry

[3]Department of Psychology & College Behavioral and Emotional Health Institute, Virginia Commonwealth University

[4]Virginia Commonwealth University, Department of Psychology

[5]Virginia Commonwealth University Alcohol Research Center; Virginia Institute for Psychiatric and Behavioral Genetics; Department of Psychiatry, Virginia Commonwealth University

[6]Mayo Clinic, Psychiatric Genomics and Pharmacogenomics Program

[7]Icahn School of Medicine at Mount Sinai, Department of Neuroscience

[8]Mayo Clinic, Department of Health Sciences Research, and Department of Psychiatry and Psychology

[9]Department of Psychiatry and Behavioral Sciences, State University of New York Downstate Medical Center

[10]Washington University School of Medicine, Department of Psychiatry

[11]Institute of Human Genetics, University of Bonn; and Department of Genomics, Life & Brain Center, University of Bonn

[12]University of Utah, Department of Psychiatry

[13]23andMe, Inc.

[14]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University

[15]Martin-Luther-University Halle-Wittenberg, Department of Psychiatry, Psychotherapy and Psychosomatics

[16]QIMR Berghofer Medical Research Institute

[17]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine

[18]Human Genomics Research Group, Department of Biomedicine, University of Basel Institute of Medical Genetics and Pathology, University Hospital Basel

[19]NIH/NIAAA, Laboratory of Neurogenetics, NIAAA, NIH, USA

[20]Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[21]University of Otago, Christchurch, New Zealand

[22]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland

[23]Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland

[24]Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

[25]Johns Hopkins Bloomberg School of Public Health

[26]University of Edinburgh, Division of Psychiatry; Centre for Cognitive Ageing and Cognitive Epidemiology

[27]Department of Integrative Physiology, University of Colorado Boulder

[28]Department of Psychiatry, Amsterdam Public Health Research Institute, VU University Medical Center/GGz inGeest, Amsterdam, the Netherlands

[29]Biostatistics and Computational Biology Unit, University of Otago, Christchurch, New Zealand

[30]Virginia Commonwealth University, Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry

[31]Department of Psychiatry, Yale School of Medicine and VA CT Healthcare Center, West Haven, CT, USA

[32]Mayo Clinic, Department of Health Sciences Research

[33]Washington University School of Medicine, Department of Genetics

[34]University of California San Diego, Department of Psychiatry

[35]NIH/NIAAA, Office of the Clinical Director

[36]Department of Medicine (Biomedical Genetics), Boston University School of Medicine

[37]Institute of Behavioral Science and Department of Sociology, University of Colorado

[38]Indiana University School of Medicine, Department of Medical and Molecular Genetics

[39]University of Colorado School of Medicine, Department of Pharmacology

[40]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[41]Mayo Clinic, Department of Molecular Pharmacology and Experimental Therapeutics

[42]Duke University Medical Center, Department of Psychiatry and Behavioral Sciences

[43]Washington University School of Medicine, Department of Medicine and Division of Biostatistics

[44]Department of Psychiatry, University of Mainz

[45]National Drug and Alcohol Research Centre, University of New South Wales

[46]Stanford University Graduate School of Education

[47]Mayo Clinic, Department of Psychiatry and Psychology

[48]Department of Psychiatry and Psychotherapy, University of Düsseldorf

[49]Max-Planck-Institute of Psychiatry

[50]University of Minnesota, Department of Psychology

[51]Department of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University

[52]University of Iowa Roy J and Lucille A Carver College of Medicine, Department of Psychiatry

[53]Addictions Dept, Institute of Psychiatry, Psychology & Neuroscience, King's College London

[54]Department of Psychiatry, University of Bonn

[55]Institute for Health and Welfare, Finland

[56]Indiana University School of Medicine, Biochemistry/Molecular Biology

[57]Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center

[58]Department of Statistical Genetics, Max-Planck-Institute of Psychiatry

[59]Department of Psychiatry, Indiana University School of Medicine

[60]Department of Medicine, Department of Neurology and Department of Psychiatry Massachusetts General Hospital, Boston, MA, USA

[61]Vitos Hospital Herborn, Dept. of Psychiatry and Psychotherapy, Herborn

[62]University of Pittsburgh, School of Pharmacy

[63]Department of Psychiatry and Psychotherapy, University of Regensburg Psychiatric Health Care Aargau

[64]LVR-Hospital Essen, Department of Psychiatry and Psychotherapy, Department of Addictive Behaviour and Addiction Medicine, Medical Faculty, University of Duisburg-Essen

[65]Medical Park Chiemseeblick in Bernau-Felden

[66]Psychiatric Hospital, Ludwig-Maximilians-University

[67]Department of Psychiatry and Psychotherapy, University of Regensburg

[68]University of Utah, Department of Sociology

[69]University of California, San Diego School of Medicine, Department of Psychology

[70]University of Chicago, Department of Psychiatry and Behavioral Neuroscience

[71]NIAAA Intramural Research Program

[72]Department of Human & Molecular Genetics, Virginia Commonwealth University

[73]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland and National Institute for Health and Welfare, Finland

[74]Departments of Neurology, Ophthalmology, Epidemiology, and Biostatistics, Boston University Schools of Medicine and Public Health

[75]RTI International, Behavioral and Urban Health Program, Behavioral Health and Criminal Justice Division

[76]Department of Sociology and Carolina Population Center, University of North Carolina at Chapel Hill

[77]University of Connecticut School of Medicine, Department of Psychiatry

[78]University of Colorado Boulder, Institute for Behavioral Genetics

[79]University of Colorado Denver, School of Medicine

[80]RTI International, Fellows Program

[81]Department of Public Health, University of Helsinki, Helsinki, Finland

[82]University of Pennsylvania Perelman School of Medicine, Center for Studies of Addiction, Department of Psychiatry and VISN 4 MIRECC, Crescenz VAMC

[83]University of Colorado Boulder, Department of Molecular, Cellular, and Developmental Biology

[84]Department of Medical Epidemiology and Biostatistics, Karolinska Instituet, Stockholm, Sweden

[85]Peter Boris Centre for Addictions Research, McMaster University/St. Joseph's Healthcare Hamilton; Michael G. DeGroote Centre for Medicinal Cannabis Research

[86]Virginia Commonwealth University, Virginia Institute for Psychiatric and Behavioral Genetics

[87]The Institute for Molecular Bioscience, University of Queensland

[88]Institute of Human Genetics, University of Bonn School of Medicine & University Hospital Bonn

[89]University of California San Diego, Institute for Genomic Medicine

[90]Department of Psychological & Brain Sciences, Indiana University, Bloomington, IN

[91]Department of Biostatistics, Yale School of Public Health, Yale University

[92]Departments of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine; VA Connecticut Healthcare System

Keywords: Genome-wide association study, alcoholism, psychiatric disorders, alcohol use, pleiotropy

ABSTRACT

Liability to alcohol dependence (AD) is heritable, but little is known about its complex polygenic architecture or its genetic relationship with other disorders. To discover loci associated with AD and characterize the relationship between AD and other psychiatric and behavioral outcomes, we carried out the largest GWAS to date of DSM-IV diagnosed AD. Genome-wide data on 14,904 individuals with AD and 37,944 controls from 28 case/control and family-based studies were meta-analyzed, stratified by genetic ancestry (European, N = 46,568; African; N = 6,280). Independent, genome-wide significant effects of different *ADH1B* variants were identified in European (rs1229984; p = 9.8E-13) and African ancestries (rs2066702; p = 2.2E-9). Significant genetic correlations were observed with schizophrenia, ADHD, depression, and use of cigarettes and cannabis. There was only modest genetic correlation with alcohol consumption and inconsistent associations with problem drinking. The genetic underpinnings of AD only partially overlap with those for alcohol consumption, underscoring the genetic distinction between pathological and non-pathological drinking behaviors.

INTRODUCTION

Excessive alcohol use is a leading contributor to morbidity and mortality. One in 20 deaths worldwide is attributable to alcohol consumption, as is 5.1% of the global burden of disease[1]. Alcohol dependence (AD), as defined by the Fourth Edition of the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)[2], is a serious psychiatric disorder characterized by tolerance, withdrawal, loss of control over drinking and excessive alcohol consumption despite negative health and social consequences. Among alcohol drinkers, 12% meet criteria for DSM-IV AD during their lifetimes[3]. In the United States, only 25% of those with AD ever receive treatment[4,5].

AD is moderately heritable (49% by a recent meta-analysis)[6] and numerous genome-wide association studies (GWAS) have aimed to identify loci contributing to this genetic variance (see[7] for a review). According to one study, common SNPs are responsible for as much as 30% of the variance in AD[8], but few have been identified to date. Variants in the genes responsible for alcohol metabolism[9–19] (*ADH1B* and, to a lesser extent, *ADH1C* and others[20–22], e.g., *ADH4)* have been strongly implicated, initially in East-Asians[9,11,12] and more recently in people of European origin (EU) and in African-Americans (AAs)[13–15]. The association between AD (and related problem drinking phenotypes) and rs1229984, a missense SNP (Arg48His) in *ADH1B* that affects the conversion of alcohol to acetaldehyde, represents one of the largest common-variant effect sizes observed in psychiatry, with the His48 allele accelerating ethanol metabolism and affording approximately 3-fold reduction in likelihood of AD across numerous studies (e.g.,[14,23,24]). Another functional polymorphism, rs671 in *ALDH2* (Glu504Lys), strongly affects alcohol metabolism by blocking conversion of acetaldehyde to acetate, but is rare except in some Asian populations[9–11,17,18]. *ADH1B*

and *ALDH2* polymorphisms, however, only explain a small proportion of the heritable variation in AD in populations of European ancestry.

In this study, we compiled the largest numbers of carefully diagnosed alcohol dependent individuals and alcohol-exposed controls to date, from both case-control and family studies. These included substantial numbers of both European ancestry (EU, N = 46,568, including 38,686 unrelated individuals) and African-American ancestry (AA, N = 6,280, including 5,799 unrelated individuals) subjects. Each study was subjected to stringent quality control (QC) before conducting GWAS within each population of each study, followed by a genome-wide meta-analysis. We estimated the heritability (SNP-h^2) of AD and examine the extent to which aggregate genetic variation in AD is related to traits from 42 other GWAS, including continuous measures of alcohol consumption.

METHODS

Samples: We collected individual genotypic data from 14 case/control studies and 9 family-based studies and summary statistics from GWAS of AD from 5 additional cohorts (**Table 1**; see **Supplementary Information** for cohort descriptions). AD was defined as meeting criteria for a DSM-IV (or DSM-IIIR in one instance) diagnosis of AD. Excepting three cohorts with population-based controls (N=7,015), all controls were screened for AD. Individuals with no history of drinking alcohol and those meeting criteria for DSM-IV alcohol abuse were additionally excluded as controls where applicable (**Supplementary Information**).

Quality control and imputation: Data for the genotyped cohorts that shared raw data were deposited to a secure server for uniform quality control (QC). QC and imputation of the 14 case/control studies was performed using the ricopili pipeline (https://github.com/Nealelab/ricopili). For 9 family-based cohorts, an equivalent pipeline,

263 picopili (https://github.com/Nealelab/picopili), was developed for QC, imputation, and
264 analysis appropriate for diverse family structures, including twins, sibships and
265 extended pedigrees (**Supplementary Information**).

266

267 After initial sample and variant QC, principal components analysis (PCA) was used to
268 identify population outliers for exclusion and to stratify samples in each study by
269 continental ancestry. Identified EU and AA ancestry populations were confirmed by PCA
270 with the 1000 Genomes reference panel[25]. Final sample and variant QC, including filters
271 for call rate, heterozygosity, and departure from Hardy-Weinberg equilibrium (HWE),
272 was then performed within each ancestry group in each cohort (see **Supplementary**
273 **Information**). Samples were also filtered for cryptic relatedness within and between
274 cohorts and for departures from reported pedigree structures.

275

276 Each cohort was imputed using SHAPEIT[26] and IMPUTE2[27], using the cosmopolitan (all
277 ancestries) 1000 Genomes reference panel. Consistency of minor allele frequencies
278 (MAF) with the reference panel was verified prior to imputation, with SNPs in EU
279 cohorts compared to MAF in European population samples and AA cohorts compared
280 to MAF in African population samples. Imputed SNPs were then filtered for INFO score
281 > 0.8 and allele frequency > 0.01 prior to analysis.

282

283 *Association Analysis:* A GWAS for AD status was performed within each ancestry
284 stratum of each sample using an association model appropriate for the study design
285 (**Table 1**). For case/control studies, GWAS was performed using logistic regression with
286 imputed dosages. For family-based studies of small, simple pedigrees (e.g., sibships),
287 association with imputed genotypes was tested using generalized estimating equations
288 (GEE). For more complex pedigrees, imputed genotypes were tested using logistic
289 mixed models. Sex was included as a covariate, along with principal components to
290 control for population structure (**Supplementary Information**). Details of the analytic

model, software used, effective N, number of SNPs and principal components are presented for each sample in Supplementary **Table S1**.

In addition to this primary analysis, subsets of genetically unrelated individuals were selected from each family-based cohort (i.e. taking one individual per family) and used to perform a conventional case/control GWAS using logistic regression. This was used in place of the family-based GWAS for estimation of effect sizes and inclusion in estimation of SNP-h^2 and genetic correlations (r_g) using LD score regression analyses.

Meta-analysis: The primary discovery meta-analysis of all ancestry-stratified GWAS (N_{case} = 14,904; $N_{control}$ = 37,994) was conducted in METAL[28]. As the different study designs (family vs. case-control) produced effect sizes that were not comparable, results were combined using weighting by effective sample size (see **Supplementary Information**). Separate ancestry-specific discovery meta-analyses of EU (N = 46,568) and AA (N = 6,280) cohorts, respectively, were also performed. Heterogeneity was evaluated across all cohorts and between study design subsets (**Supplementary Information**). Power analysis was performed using CaTS[29] with the estimated effective sample size.

In addition to the discovery meta-analyses, we conducted meta-analyses for two design subsets. First, we performed sample size weighted meta-analysis of the subset of genetically unrelated individuals in EU (N = 38,686) and AA (N = 5,799) cohorts for use in LD score regression (LDSR) analysis. Second, we performed inverse-variance weighted meta-analysis of genetically unrelated individuals in genotyped cohorts to estimate within-ancestry effect sizes for EU (N = 28,757) and AA (N = 5,799). These effect sizes were then used to compare trans-ancestral fine mapping results using inverse-variance weighted fixed effects, random effects[30], and Bayesian[31] models

(**Supplementary Information**). Supplementary **Table S2** provides an overview of the various meta-analytic models that were fitted to data.

Heritability and Genetic Correlation Analysis: LDSR analysis[32] was performed to estimate the heritability explained by common SNPs in meta-analyses of unrelated EU and AA samples, respectively. LDSR was performed using HapMap3 SNPs and LD scores computed from 1000 Genomes reference samples corresponding to each population (Supplementary Information). Conversion of h^2_g estimates from observed to liability scale was performed assuming population prevalences of 0.159 and 0.111 for AD in alcohol-exposed EU and AA individuals, respectively[3].

Genetic correlation between AD and 42 traits from LD Hub[33] and other published studies[34–44] was examined with the same unrelated EU meta-analysis (10,206 cases and 28,480 controls) and precomputed European LD scores using LDSR. To avoid increasing the multiple testing burden, redundant or highly-correlated phenotypes were reduced by manually selecting the version of the phenotype with the greatest predicted relevance to AD, largest sample size, or highest heritability (**Supplementary Information**).

Replication: As described below, a locus on chromosome 3 was genome-wide significant (GWS) in the trans-ancestral discovery meta-analysis. The minor allele, associated with lower AD risk in our analysis, had low frequency in all EU samples except the Finnish cohorts; it was also higher in AAs. To seek replication, we examined the association between this locus and DSM-IV AD in two independent AA samples (Yale-Penn 2, n = 911 cases and 599 controls, and COGA AAfGWAS, n = 880 cases and 1,814 controls; Supplementary Information) using GEE (Yale-Penn) and Genome-Wide Association/Interaction Analysis and Rare Variant Analysis with Family Data (GWAF; in COGA) respectively. Association with AD status, broadly defined using

hospital and death records, was also examined in the FINRISK cohort (1,232 cases and 22,614 controls) using Firth logistic regression[45].

RESULTS

GWAS meta-analyses*:* In both the EU and AA analyses, GWS loci ($p < 5E-8$) were identified in the ADH gene cluster on chromosome 4 (**Figure 1** for Manhattan plot; **Table 2** for top loci; Supplementary **Figure S1** for QQ plot for discovery GWAS showing polygenic signal). Examining individual populations, rs1229984 in *ADH1B* was the strongest associated signal from the analysis in EU ($p = 9.8E-13$), while rs2066702, also in *ADH1B*, was the most significant variant in AA ($p = 2.2E-9$; **Figure 2** shows the regional association plots for the *ADH1B* locus for the discovery, EU, AA and trans-ancestral meta-analysis). Clumping for linkage disequilibrium (LD; $r^2 < .1$ within 500kb) suggested multiple independent signals within this locus in both populations (**Table 2**), with differing leading alleles reflecting different LD structures and allele frequencies in each population (Supplementary **Figure S2A** and **S2B** show LD patterns in the ADH locus, including *ADH1B*, in AA and EU respectively). Conditional analysis controlling for rs2066702 (Supplementary **Figure S3** for results in AA) and rs1229984 (Supplementary **Figure S4** for results in EU) was inconclusive due to limited power, but was tentatively consistent with the existence of additional independent effects in the region (Supplementary **Table S3** shows marginal and conditional effect sizes for genome-wide significant SNPs in the *ADH1B* locus). The most promising support for an independent signal arises from rs894368 (marginal odds ratio = 0.887, $p = 6.9E-7$; conditional odds ratio = 0.890, $p = 6.8E-6$; Supplementary Information). Results from the trans-ancestral meta-analysis reinforced the robust effects of rs1229984 and other *ADH1B* SNPs on liability to AD (regional association plot for rs1229984 in Supplementary **Figure S5A** (inverse-variance weighted), **S5B** (modified random-effects) and **S5C** (Bayesian)) across various analytic models.

We also verified whether variants affecting *ADH1B* expression (eQTLs) were associated with AD. Considering GTEx data V7 (available at https://www.gtexportal.org/), 263 variants were reported to affect *ADH1B* expression in different human tissues (FDR q<0.05). After LD-informed clumping and the exclusion of variants in LD with the GWS coding alleles (i.e., rs1229984 and rs2066702), three variants (i.e., rs11939328, rs10516440, rs7664780) were considered with respect to their association with AD. SNP rs10516440 showed a genome-wide significant association with AD with contribution from both AA and EU analyses (trans-ancestry p = 4.72E-8; EU p = 3.97E-6; AA p = 1.97E-3). In line with the effect of the coding variants where the protective allele is associated with increased *ADH1B* enzymatic activity, the rs10516440*A allele was associated with reduced AD risk and increased *ADH1B* expression, which was consistent across multiple tissues (multi-tissue p = 1.42E-76).

A novel locus on chromosome 3, rs7644567, also achieved GWS in the meta-analysis (p = 3.03E-8; Supplementary **Figure S6** for regional association plot), primarily attributable to contributions from the AA samples (p = 6.64E-6) with the major, A, allele being associated with AD risk liability. The G allele has an MAF = 0.29 in AA, but MAF<0.01 in most EU samples, except in FinnTwin (MAF = .032) and NAG-Fin (MAF = 0.054). In AA, rs7644567 does not appear to be in high LD with other variants (Supplementary **Figure S6**) and did not replicate in two independent AA samples. In the independent FINRISK cohort (MAF = .045), there was modest evidence for association (p = 0.019), but with risk associated with the minor G allele (**Supplementary Table S4** for results in each replication sample).

Overall, there was limited evidence for heterogeneity across all cohorts, within ancestry, between ancestries, or between study designs within ancestry (**Supplementary Information**; Supplementary **Figure S7-S13**). Gene-level association testing with

MAGMA[46] did not identify any additional genes in EU or AA (Supplementary **Table S5** for top 20 genes in EU and AA).

Heritability and genetic correlations: LD score based liability-scale SNP-heritability of AD was estimated at h^2_g = 0.090 (SE = 0.019, p = 8.02E-7) in the meta-analysis of unrelated EU samples. Exclusion of the *ADH1B* locus did not substantially modify this estimate (h^2_g = 0.089, SE = 0.0185). Nominally significant heritability from common variants was also estimated for the meta-analysis of unrelated AA individuals based on LDSR with scores computed from 1000 Genomes African populations (p = .017), but the quantitative estimate of h^2_g was unstable depending on the choice of reference panel, reflecting the challenge of correctly specifying LDSR and robustly modelling LD for the admixed AA population (**Supplementary Information**).

Significant genetic correlation with AD in EU was observed for 16 traits (significant genetic correlations in **Figure 3;** all genetic correlations in Supplementary **Table S6**), after correction for multiple testing (p = 1.19E-3 for 42 traits). The largest positive correlations were with ever smoking tobacco (rg = .708, p = 1.3E-7) and lifetime cannabis use (rg = .793, p = 2.5E-4), and with other psychiatric disorders and traits, especially schizophrenia (rg = 0.357, p = 3.2E-11), ADHD (rg = .444, p = 4.2E-6), and depressive symptoms (rg = .603, p = 2.6E-7). Educational attainment (rg = -0.424, p = 6.8E-9) and age at first birth (higher values indicate that subjects were older when they had their first child, rg = -0.63, p = 2.0E-9) showed significant inverse genetic correlation with AD suggesting that liability to AD risk was genetically related to lower educational attainment and lower age at which one had their first child.

Unexpected patterns of genetic correlation were observed when comparisons were made to other alcohol-related measures. AD was genetically correlated with alcohol consumption in a meta-analysis of the Alcohol Genome-wide Association (AlcGen) and

Cohorts for Aging and Research in Genomic Epidemiology Plus (CHARGE+) consortia[36] (rg = .695, p = 6.9E-6) but only modestly with alcohol consumption from the recent large UK Biobank analysis[37] (rg = 0.371, p = 5.2E-5). Liability to AD was not correlated with genome-wide SNPs from a recent GWAS of the Alcohol Use Disorders Identification Test (AUDIT) in 23andMe[38] (rg = 0.076, p = 0.65), perhaps due to the low levels of drinking observed in this population[38]. Additional analysis indicates AD is genetically correlated with GWAS of delay discounting in the 23andMe sample[42] (rg = 0.478, p = 6.0E-3), suggesting behavioral phenotypes in the cohort are still informative to AD.

Associations with other GWS loci: We examined results for the eight independent variants associated at GWS levels with alcohol consumption in the UK Biobank[37] (**Supplementary Table S7**). Among the UK Biobank findings, three of the four reported variants in the ADH region of chromosome 4 (rs145452708 – a proxy for rs1229984, rs29001570 and rs35081954) were associated in the present study with AD (p ranging from 3.5E-5 – 2.3E-10) with sign concordant effects; the remaining variant was excluded from our analysis due to MAF <0.01. The UK Biobank lead variant in *KLB*, rs11940694, was nominally associated with AD (p = .0097), though this does not surpass multiple testing correction for the eight GWS alcohol consumption loci. We see little evidence (p >0.2) for association of AD with the reported loci at *GCKR* and *CADM2*, which may be due to differences in power for the given effect size or because these genes exert an influence on liability to consume alcohol but not later problems. The locus on chromosome 18 showed limited regional association with AD, but the index variant was not present in our analysis because it no longer appears in the 1000 Genomes Phase 3 reference panel[25].

Power analysis: Only 3 additional loci reach p < 1E-6 (**Table 2**). Power analyses indicated that the current meta-analysis is expected to have at least 63% power to detect very common variants (MAF ≥ 0.25) with odds ratios ≥ 1.10 at p < 1E-6 (41% for

459 $p < 5E-8$; **Supplementary Figure S14** for power analysis curves). Power is lower for
460 less common variants (MAF ≥ .05) even with odds ratios ≥ 1.20 at $p < 1E-6$ (60% power)
461 and $p < 5E-8$ (38% power).

462

463 **DISCUSSION**

464

465 To our knowledge, this is the largest GWAS of rigorously-defined AD. We identified loci
466 in *ADH1B* that differed between EU and AA, as well as novel genetic correlations
467 between AD and psychiatric disorders (e.g., schizophrenia), tobacco and cannabis use,
468 and behavioral outcomes (e.g., educational attainment). Analyses also revealed a
469 genetic distinction between GWAS results for alcohol consumption and AD. Although
470 larger sample sizes can be amassed by focusing on quantitative measures of
471 consumption, only the upper tail is relevant to AD (as a medical diagnosis) and even
472 that does not capture other aspects of disordered drinking (e.g., loss of control,
473 withdrawal) directly. Conversely, cases derived from electronic medical records (e.g.,
474 ICD codes) may result in a high rate of false negatives, while self-screening instruments
475 (e.g. AUDIT scores) is best suited to analyses of disordered drinking when a sufficiently
476 high threshold or score cut-off is applied to pinpoint severity. Our study has the
477 advantage of greater diagnostic precision via use of semi-structured interviews to
478 diagnose AD systematically in a majority of the constituent studies.

479

480 The genome-wide significant SNPs reaffirm the importance of functional variants
481 affecting alcohol metabolism to the risk of AD. The top association in *ADH1B*,
482 rs1229984, is a missense variant that is amongst the most widely studied in relation to
483 alcohol use, misuse and dependence. The resulting amino acid substitution (Arg48His)
484 increases the rate at which ADH1B oxidizes ethanol to acetaldehyde[10,11]. Early studies
485 on Asian populations in which the derived allele is common demonstrated strong
486 protection against the development of AD[9–11]. In EUs and AAs, the protective allele is

present at much lower frequencies (EU MAF = 3-4%, AA MAF < 1%), but recent large-scale studies have shown an association between this locus and alcohol consumption and problems at GWS levels in EU with similar effect size[14,15]. The lead variant in AA cohorts, rs2066702 (Arg370Cys), is another functional missense variant in *ADH1B,* and it, similarly, encodes an enzyme with an increased rate of ethanol oxidation[10,11]. The allele encoding Cys370 is common among AAs, but rare in other populations[10]. Our results clearly show that these two different functional SNPs in *ADH1B* both affect risk for alcoholism, with their relative importance dependent upon allele frequency in the population studied. Larger future studies will be needed to evaluate the evidence for additional independent effects in the chromosome 4 locus.

The only other locus to reach significance was rs7644567 on chromosome 3, primarily driven by AA cohorts due to the variant's very low MAF in EU. This locus did not replicate in independent African or Finnish ancestry samples. We note that the conventional genome-wide significance threshold is derived for European ancestry samples, and thus is likely to be too lenient in GWAS of African-ancestry cohorts due to higher genetic diversity and corresponding increase in the effective number of independent tests in the GWAS[47,48]. As an illustration, in 4 samples from the current study that included both EU and AA participants, the number of independent SNPs identified upon LD pruning was 1.7- to 2.3-fold greater in AA than EU subjects. Much larger studies in AA and other non-EU populations will clearly be important to elucidate additional loci.

Despite limited SNP-level findings, there is significant evidence for polygenic effects of common variants in both EU and AA cohorts. The estimated h^2_g = .09 for AD in EU is only modestly lower than those recently reported for alcohol consumption (h^2_g = .13)[37] and AUDIT scores (h^2_g = .12)[38], and comparable to estimates derived for cigarettes-per-day[33]. Our h^2_g estimate is lower than a prior report[8], likely reflecting a combination of differences in estimation method and greater heterogeneity in ascertainment strategy

across samples in the current study. The latter is especially relevant given that we incorporated population-based cohorts with a wide range of ages at ascertainment and cultural environments, as well as cohorts enriched for other substance use disorders.

Comparing our GWAS to recent GWAS of alcohol consumption measures suggests that the liability underlying normative patterns of alcohol intake and AD are only partially overlapping. Genome-wide, we observe only modest genetic correlation (significantly < 1) with log-scaled alcohol consumption by participants in AlcGen and CHARGE+ Consortia cohorts[36] (rg = .695) and in the UK Biobank[37] (rg = .371), and no significant correlation with GWAS of log-scaled AUDIT scores in 23andMe participants[38] (rg = .076). We also observe only partial replication of the 8 loci significantly associated with consumption in the UK Biobank. One key factor in interpreting the differences between these traits and AD is that the distribution of consumption levels and AUDIT scores can be highly skewed in population samples, with most individuals at the low (nonpathological) end of the spectrum. This effect may be especially pronounced among the older, healthy volunteers of the UK Biobank cohort[49] and the 23andMe cohort, which is more educated and has higher socioeconomic status than the general US population [38]. We hypothesize that the variants that affect consumption at lower levels may differ substantively from those that affect very high levels of consumption in alcohol dependent individuals, who are also characterized by loss of control over intake[50]. This appears to be the case in one prior study that used specific cut-offs to harmonize AUDIT scores with AD data and noted significant concordance in SNP-h^2 estimates[51] – according to that study, the optimal cutoffs for their sample were ≥6 and ≥9 for women and men respectively. However, there is a need for a further detailed characterization of how AUDIT cut-offs may be applied to maximize concordance with genetic liability to AD diagnosis risk. The strongly negative genetic correlation between educational attainment and AD, in contrast to positive genetic correlations of education with consumption and AUDIT scores, further underscore this distinction between

544 normative/habitual levels of alcohol intake and diagnosed AD, at least in the respective
545 populations studied.

546

547 The current analysis also identified robust genetic correlation of AD with a broad variety
548 of psychiatric outcomes. This correlation is strongest for aspects of negative mood,
549 including neuroticism and major depressive disorder, as also seen in twin studies[52,53]
550 and through recent specific molecular evidence for pleiotropy[54,55]. Taken together with
551 evidence from other recent genomic studies[54], and null correlations for other GWAS of
552 alcohol consumption, these findings suggest that major depression may only share
553 genetic liability with alcohol use at pathological levels.

554

555 AD was also negatively genetically correlated with AFB, which is an indicator of
556 reproductive tempo and correlated with age at first consensual sexual intercourse[56].
557 This is consistent with evidence of common genetic liability to early, risky behaviors
558 underlying AD and AFB[57]. Nominally significant genetic correlation with delay
559 discounting (i.e. favoring immediate rewards) and the strong genetic correlation of AD
560 with ADHD, cigarette smoking and cannabis use may similarly reflect a shared genetic
561 factor for risk-taking and reduced impulse control.

562

563 Lower genetic correlations were observed for most biomedical and anthropometric
564 outcomes. Liver enzymes GGT and ALT, once proposed as possible biomarkers for
565 alcohol abuse[58], showed, as expected, nominal evidence for genetic correlation with AD
566 but neither survived multiple testing correction. Notably, we did not find any association
567 between AD and body-mass index (BMI). Negative genetic correlations with BMI were
568 previously reported for both alcohol consumption[37] and AUDIT scores[38], but there is
569 prior evidence that BMI has differing underlying genetic architecture in the context of AD
570 and outside of that context[59]. The negative genetic correlations observed in those
571 studies are consistent with studies of light to moderate drinking, which is also

associated with healthier lifestyle behaviors, while heavy and problematic drinking is typically associated with weight gain[60].

This study benefits from precision in diagnostic assessment of AD, known alcohol exposure in a majority of the controls, and careful quality control that excluded overlaps of individuals between studies while combining case-control and twin/family-based study designs. Despite these strengths our sample size was insufficient to identify additional GWS loci robustly. Power analyses indicate that additional SNPs associated with AD are likely to have small effect sizes, consistent with other psychiatric disorders (e.g. depression[61]). This supports the pressing need for collection of large numbers of well characterized cases and controls. The differences between our results and the study of AUDIT scores[38], however, highlight that ascertainment and trait definition must also be taken into account. Careful study of how screening tools, such as the AUDIT, correlate to genetic liability to AD (as defined by DSM-IV or similar) could substantially boost sample sizes for future AD GWAS. There is also a continued need to characterize the genetic architecture of AD in non-EU populations.

We show a novel genetic distinction between drinking in the pathological range (AD) and habitual drinking that does not cross the threshold into pathology or dependence. Larger future samples will allow us to uncover additional pleiotropy between pathological and non-pathological alcohol use as well as between AD and other neuropsychiatric disorders.

Table 1: Descriptive statistics for cohorts in the meta-analysis of AD.

				European (EU)				African - American (AA)			
		Male	Ages	N Total		N Unrelated		N Total		N Unrelated	
Dataset	**PMID**	**(%)**	**(years)**	**Case**	**Control**	**Case**	**Control**	**Case**	**Control**	**Case**	**Control**
Case-control: Logistic Regression											
Comorbidity and Trauma Study (CATS)	23303482	56%	18-67	572	817	572	817	--	--	--	--
Christchurch Health and Development Study (CHDS)	23255320	48%	16-30	112	500	112	500	--	--	--	--
Collaborative Study of the Genetics of Alcoholism - case-control cohort (COGA-cc)	20201924	54%	18-79	583	363	583	363	--	--	--	--
Family Study of Cocaine Dependence (FSCD)	18243582	51%	18-60	266	174	266	174	255	241	255	241
German Study of the Genetics of Alcoholism (GESGA)	19581569	65%	18-84	1314	2142	1314	2142	--	--	--	--
Gene-Environment Development Initiative - Great Smoky Mountains Study (GEDI-GSMS)	8956679	57%	9-26	42	565	42	565	--	--	--	--
Center on Antisocial Drug Dependence (CADD)	25637581	70%	13-20	400	577	400	577	51	51	51	51
Phenomics and Genomics Sample (PAGES)	28371232	57%	18-74	37	523	37	523	--	--	--	--
Collaborative Study on the Genetics of Nicotine Dependence (COGEND Nico)	17158188	34%	25-82	135	272	135	272	46	232	46	232
COGEND - Study of Addiction: Genetics and Environment (COGEND SAGE)	20202923	37%	18-77	311	225	311	225	104	103	104	103
Spit For Science	24639683	36%	>18	252	1863	252	1863	74	841	74	841
National Institute on Alcohol Abuse and Alcoholism Intramural (NIAAA)	n/a	67%	>18	442	206	442	206	404	110	404	110
Mayo Clinic Center for the Individual Treatment of Alcohol Dependence (CITA)	25290263	55%	≥18	378	646	378	646	--	--	--	--
Alcohol Dependence in African Americans (ADAA)	n/a	57%	18-69	--	--	--	--	794	297	794	297
Family-based, twins and sibs: GEE											
Brisbane Longitudinal Twin Study (BLTS)	23187020	43%	18-30	60	938	51	546	--	--	--	--
GEDI - Virginia Twin Study on Adolescent Behavioral Development (GEDI-VTSABD)	9294370	38%	8-32	209	503	188	318	--	--	--	--
Minnesota Center for Twin and Family Research (MCTFR)	23942779	41%	16-21	609	2100	553	1274	--	--	--	--
Center for Education and Drug Abuse Research (CEDAR)	21514569	63%	16-34	59	200	54	152	--	--	--	--
Swedish Twin Registry (STR)	23137839	47%	40-83	76	8311	76	6112	--	--	--	--
Yale-Penn	24166409	58%	16-79	1094	301	1004	252	--	--	--	--
Family-based, large/complex pedigrees: Logistic Mixed Model											
Collaborative Study of the Genetics of Alcoholism - family cohort (COGA-fam)	23089632	45%	12-88	605	682	168	138	--	--	--	--
Australian Alcohol and Nicotine Studies (OZ-ALC-NAG)	21529783	45%	18-82	1571	3069	1111	805	--	--	--	--
Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD)	15770118	50%	17-84	721	1814	436	1802	--	--	--	--
Yale-Penn	24166409	51%	16-79	--	--	--	--	1607	1070	1263	933
Summary statistics											
Netherlands Study of Depression and Anxiety / Netherlands Twin Registry (NESDA/NTR)	18197199	31%	>18	390	1633	390	1633	--	--	--	--
Finnish Nicotine Addiction Genetics Project (NAG-Fin)	17436240	52%	30-92	439	1137	439	1137	--	--	--	--
FinnTwin12 (FT12)	17254406	47%	20-27	88	874	88	874	--	--	--	--
National Longitudinal Study of Adolescent to Adult Health (Add Health)	25378290	47%	24-34	768	2981	768	2981	--	--	--	--
Helsinki Birth Cohort Study (HBCS)	16251536	43%	56-70	36	1583	36	1583	--	--	--	--
Total				**11569**	**34999**	**10206**	**28480**	**3335**	**2945**	**2991**	**2808**

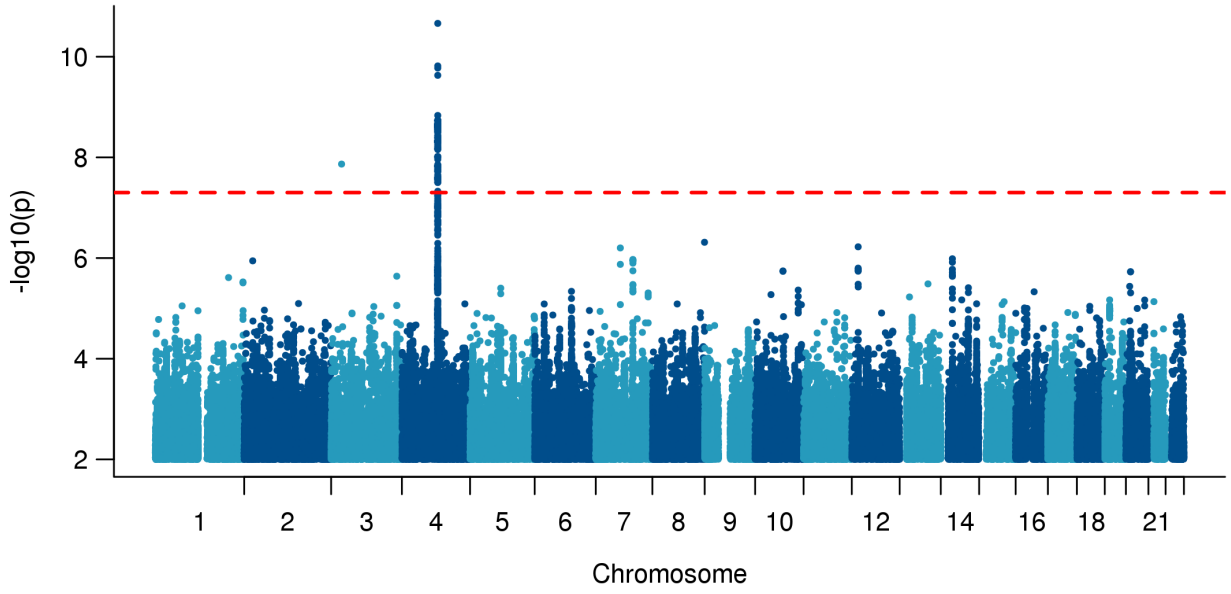
Overview of numbers of alcohol dependent cases and controls from each cohort in the current analysis, including the number of genetically unrelated individuals. Cohorts are listed by study design. Sample sizes are listed after QC exclusions and stratified by ancestry group. PubMed identifiers (PMID) are listed for previous publications describing each cohort, along with the percentage of male samples and the age range in the cohort.

Table 2: Top 10 loci from the discovery meta-analysis of alcohol dependence by ancestry

						A1 Allele Freq.		INFO score		Effect size		Discovery meta-analysis p-value		
SNP	**CHR**	**BP**	**A1**	**A2**	**Gene**	**EU**	**AA**	**EU**	**AA**	**EU OR**	**AA OR**	**EU**	**AA**	**Trans**
Trans-ancestral meta-analysis (14,904 cases, 37,944 controls)														
rs1229984	4	100239319	T	C	*ADH1B*	0.040	0.014	0.904	0.910	0.486	0.912	**9.79E-13**	3.48E-01	**2.18E-11**
rs1789912	4	100263942	T	C	*ADH1C*	0.418	0.132	1.000	1.020	1.106	1.211	1.98E-07	1.32E-03	**1.47E-09**
rs2066702	4	100229017	A	G	*ADH1B*	--	0.215	--	0.989	--	0.731	--	**2.21E-09**	**2.21E-09**
rs6827898	4	100295863	A	G	--	0.123	0.112	0.963	0.942	1.145	1.270	5.21E-07	9.31E-04	**2.97E-09**
rs7644567	3	29201672	A	G	*RBMS3*	--	0.705	--	0.997	--	1.229	--	6.64E-06	**1.36E-08**
rs894368	4	100309313	A	C	--	0.309	0.386	0.994	0.962	0.887	0.981	**1.93E-08**	9.73E-01	3.30E-07
rs116338421	8	145761256	C	G	*ARHGAP39*	--	0.172	--	0.974	--	0.755	--	4.86E-07	4.86E-07
rs79171978	12	17798824	C	G	--	0.099	0.027	0.989	0.986	1.201	1.016	5.47E-08	8.18E-01	5.98E-07
rs2461618	7	68667233	A	G	--	--	0.088	--	0.984	--	0.669	--	6.30E-07	6.30E-07
rs8017647	14	32456358	T	C	--	0.792	0.565	0.998	0.991	0.901	0.923	8.05E-06	4.71E-02	1.03E-06
African ancestry meta-analysis (3,335 cases, 2,945 controls)														
rs2066702	4	100229017	A	G	*ADH1B*	--	0.215	--	0.989	--	0.731	--	**2.21E-09**	**2.21E-09**
rs5781337	1	223883425	CA	C	--	0.263	0.212	0.982	0.927	1.007	0.664	8.85E-01	1.62E-07	6.59E-02
rs116338421	8	145761256	C	G	*ARHGAP39*	--	0.172	--	0.974	--	0.755	--	4.86E-07	4.86E-07
rs3857224	4	100129685	T	C	*ADH6*	0.315	0.585	0.994	1.000	0.970	0.814	2.40E-01	5.86E-07	2.36E-03
rs2461618	7	68667233	A	G	--	--	0.088	--	0.984	--	0.669	--	6.30E-07	6.30E-07
rs10784244	12	62035165	G	A	--	0.153	0.484	1.000	0.998	1.041	1.226	6.26E-02	1.04E-06	2.49E-04
rs17199739	16	25444288	G	A	--	0.176	0.096	0.993	0.955	0.994	0.693	4.25E-01	1.11E-06	8.66E-03
rs79016499	11	93010988	T	C	--	--	0.066	--	0.928	--	1.729	--	1.36E-06	--
rs740793	17	3846353	G	A	*ATP2A3*	0.453	0.350	0.973	0.970	0.996	1.370	4.66E-01	1.48E-06	3.44E-01
rs143258048	3	75982870	A	AC	*ROBO2*	--	0.028	--	0.879	--	0.490	--	1.86E-06	--
European ancestry meta-analysis (11,569 cases, 34,999 controls)														
rs1229984	4	100239319	T	C	*ADH1B*	0.040	0.014	0.904	0.910	0.486	0.912	**9.79E-13**	3.48E-01	**2.18E-11**
rs894368	4	100309313	A	C	--	0.309	0.386	0.994	0.962	0.887	0.981	**1.93E-08**	9.73E-01	3.30E-07
rs3811802	4	100244221	G	A	*ADH1B*	0.454	0.529	0.958	0.956	1.162	0.914	**2.40E-08**	2.19E-02	1.22E-04
rs79171978	12	17798824	C	G	--	0.099	0.027	0.989	0.986	1.201	1.016	5.47E-08	8.18E-01	5.98E-07
rs1154445	4	100288521	G	T	--	0.425	0.134	0.970	0.986	1.137	1.211	1.80E-07	2.63E-02	**1.48E-08**
rs6827898	4	100295863	A	G	--	0.123	0.112	0.963	0.942	1.145	1.270	5.21E-07	9.31E-04	**2.97E-09**
rs4388946	12	17935154	C	A	--	0.240	0.297	0.988	0.976	1.137	0.950	7.14E-07	1.87E-01	7.05E-05
rs1229863	4	100252386	A	T	*ADH1B*	0.174	0.038	0.990	0.989	1.145	1.254	7.80E-07	4.26E-02	9.28E-08
rs34929220	15	69769635	T	C	*DRAIC*	0.690	0.937	0.898	0.943	0.893	1.028	1.02E-06	8.38E-01	7.38E-06
rs113659074	4	100252308	T	G	*ADH1B*	0.068	0.093	0.980	0.947	0.800	1.166	1.54E-06	6.63E-02	2.99E-04

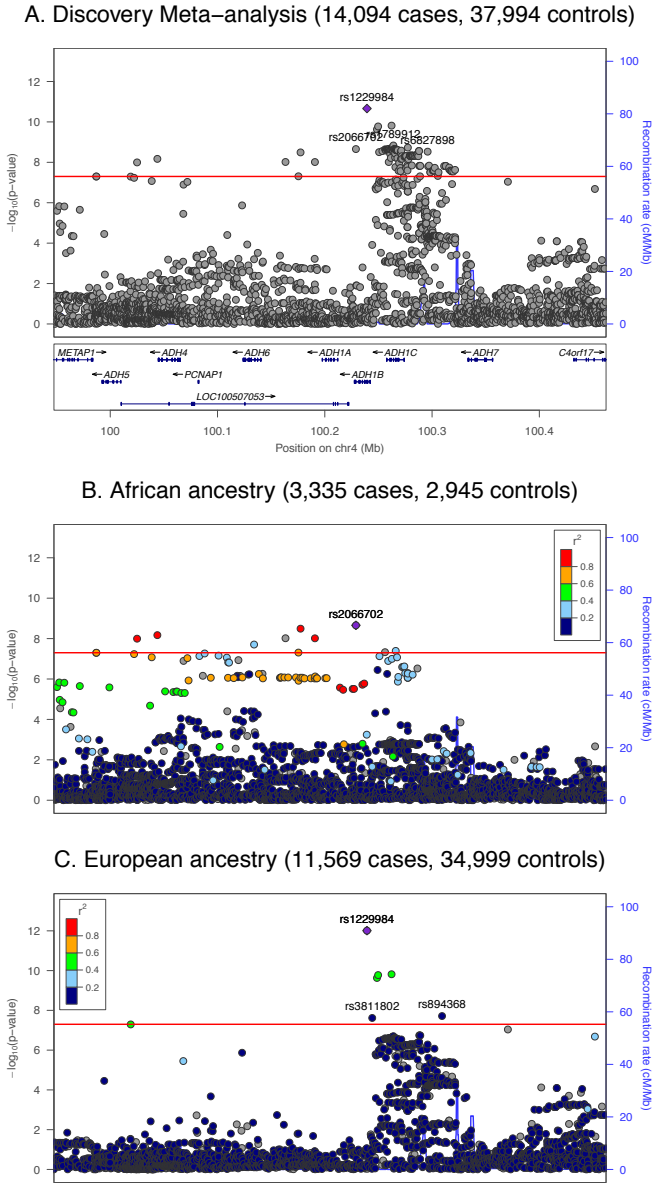
Top 10 nominally independent variants from the discovery trans-ancestral (Trans.) meta-analysis and the discovery meta-analyses in African (AA) and European (EU) ancestry cohorts, respectively. Independent variants are identified based on clumping for LD (pairwise $r^2 < 0.1$) in 1000 Genomes Project Phase 3 data[25]. EU results are clumped using European (EUR) ancestry reference samples, AA results are clumped using African ancestry reference samples from the American Southwest (ASW), and trans-ancestral results are clumped using merged EUR and African ancestry (AFR) reference samples. Meta-analysis p-values and allele frequencies (Freq.) are reported from full discovery meta-analyses. Bold indicates genome-wide significant p-values ($p < 5e\text{-}8$). Odds ratios (OR) and INFO scores are reported from the meta-analyses of the subset of unrelated individuals within each ancestry. Chromosome (CHR) and base pair (BP) position are reported for genome build hg19, with genes annotated by Ensembl VEP[62]. Allele frequency and OR are given with respect to allele 1 (A1).

Figure 1: Manhattan plot of discovery trans-ancestral meta-analysis showing strong evidence for rs1229984 in *ADH1B*.



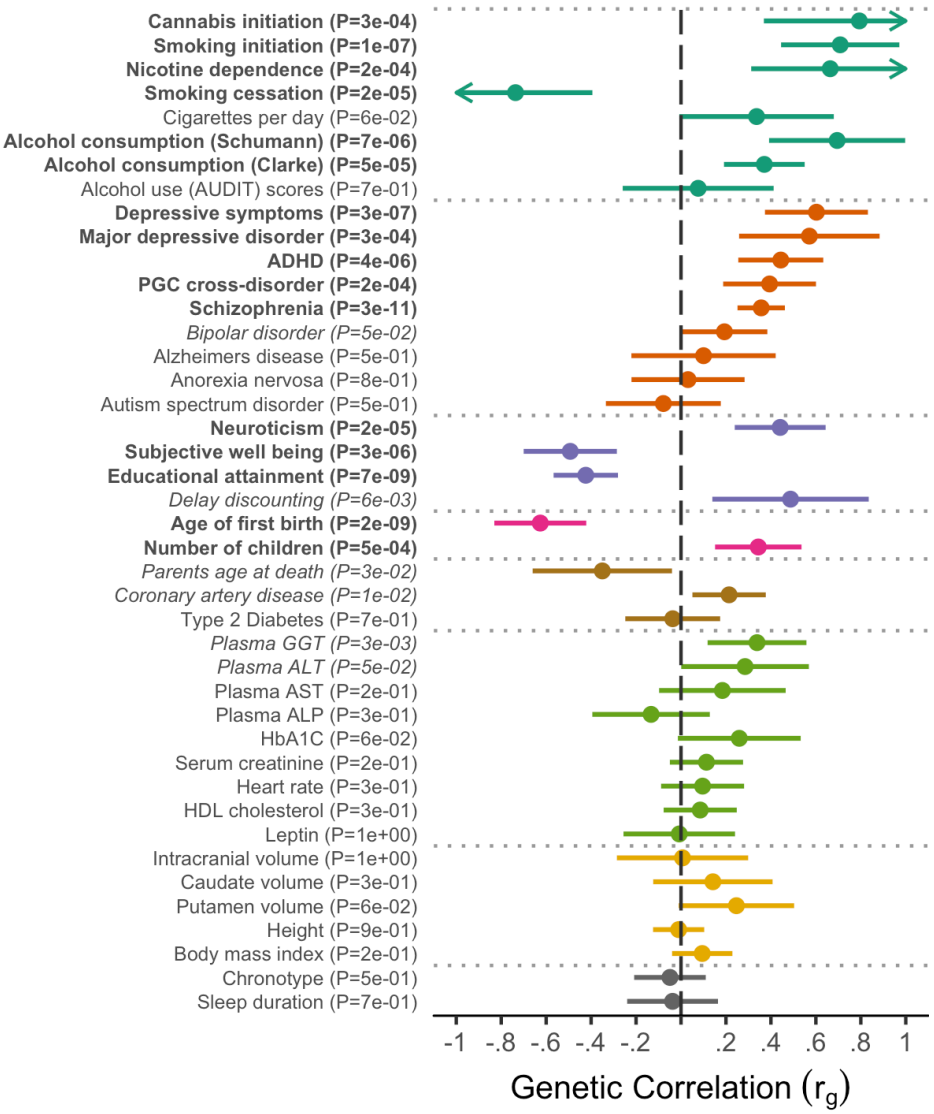
Dashed red reference line indicates genome-wide significance ($p < 5E\text{-}8$). Results are from the discovery meta-analysis of all cohorts (14,904 cases, 37,994 controls) under a fixed effects model weighted by effective sample size.

Figure 2: Regional plots for the ADH1B locus, rs1229984, in the European (EU), African-American (AA) and trans-ancestral discovery meta-analysis.



Results of meta-analysis with effective sample size weights for the *ADH1B* locus in (A) all cohorts, (B) AA cohorts, and (C) EU cohorts. Red reference line indicates the genome-wide significance threshold ($P < 5e\text{-}8$). Within ancestry, colored points reflect the degree of LD (pairwise r^2) to the index variant (indicated by a purple diamond) in 1000 Genomes Project reference data[25] for individuals of (B) African or (C) European ancestry, respectively. No reference LD panel exists for the trans-ancestral sample (A).

Figure 3: Genetic correlations between 42 traits and alcohol dependence in Europeans.



Genetic correlation results from LD score regression with the meta-analysis of AD in unrelated EU individuals (10,206 cases, 28,480 controls). After Bonferroni correction, significant correlations are observed with 16 traits and disorders ($p < 1.2E-3$); bold), with nominally significant results for 6 additional traits and disorders ($p < .05$; italics). Error bars indicate 95% confidence intervals, with arrows indicating intervals extending above 1 or below -1. Phenotypes are organized by research domain.

Ethics statement:

This study was approved by the institutional review board (IRB) of Washington University in St. Louis (Human Research Protection Office; number 201512068). Each contributing cohort obtained informed consent from their participants and received ethics approvals of their study protocols from their respective review boards in accordance with applicable regulations.

Contributors:

The 23andMe research team includes Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson

Data availability:

Summary statistics from the genome-wide meta-analyses will be made available on the Psychiatric Genomics Consortium's downloads page (http://www.med.unc.edu/pgc/results-and-downloads). Individual-level data from the genotyped cohorts and cohort-level summary statistics will be made available to researchers following an approved analysis proposal through the PGC Substance Use Disorder group with agreement of the cohort PIs; contact the corresponding authors for details. Cohort data is also available from dbGaP except where prohibited by IRB or European Union data restrictions (accession numbers to be available before publication).

Code availability:

Code for GWAS of case/control cohorts with ricopili is available at https://github.com/Nealelab/ricopili. Code for GWAS of family-based cohorts with picopili is available at https://github.com/Nealelab/picopili. Code for LD score regression analyses are available at https://github.com/bulik/ldsc. Effective sample size calculations were implemented using PLINK (https://www.cog-genomics.org/plink2), and GMMAT (https://content.sph.harvard.edu/xlin/software.html#gmmat) and geepack (https://cran.r-project.org/web/packages/geepack/index.html) in R (https://cran.r-project.org/); example code is available from the first author by request.

References

1. World Health Organization. *Global Status Report on Alcohol and Health, 2014*. World Health Organization; 2014.
2. American Psychiatric Association. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. *Am Psychiatr Assoc*. 2000. doi:10.1176/appi.books.9780890423349
3. Hasin D, Stinson F, Ogburn E, Grant B. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: Results from the national epidemiologic survey on alcohol and related conditions. *Arch Gen Psychiatry*. 2007;64(7):830-842.
4. Dawson DA, Grant BF, Stinson FS, Chou PS. Estimating the effect of help-seeking on achieving recovery from alcohol dependence. *Addiction*. 2006;101(6):824-834.
5. Dawson DA, Grant BF, Stinson FS, Chou PS, Huang B, Ruan W. Recovery from DSM-IV alcohol dependence: United States, 2001–2002. *Addiction*. 2005;100(3):281-292.
6. Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol Med*. 2015;45(5):1061-1072.
7. 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(8):1312-1327.
8. Palmer RHC, McGeary JE, Heath AC, Keller MC, Brick LA, Knopik VS. Shared additive genetic influences on DSM-IV criteria for alcohol dependence in subjects of European ancestry. *Addiction*. 2015;110(12):1922-1931.
9. Thomasson HR, Edenberg HJ, Crabb DW, et al. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *Am J Hum Genet*. 1991;48(4):677.
10. Edenberg HJ. The Genetics of Alcohol Metabolism: Role of Alcohol Dehydrogenase and Aldehyde Dehydrogenase Variants. *Alcohol Res Heal*. 2007;30(1):5-13.
11. Hurley TD, Edenberg HJ. Genes Encoding Enzymes Involved in Ethanol Metabolism. *Alcohol Res*. 2012;34(3):339-344.
12. Luczak SE, Glatt SJ, Wall TJ. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians. 2006.
13. 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(6):504-512.
14. Bierut LJ, Goate AM, Breslau N, et al. ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African

ancestry. *Mol Psychiatry*. 2012;17(4):445.

15. Frank J, Cichon S, Treutlein J, et al. Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict Biol*. 2012;17(1):171-180.
16. Hart AB, Lynch KG, Farrer L, Gelernter J, Kranzler HR. Which alcohol use disorder criteria contribute to the association of ADH1B with alcohol dependence? *Addict Biol*. 2016;21(4):924-938.
17. 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(5):725-737.
18. Crabb DW, Edenberg HJ, Bosron WF, Li T-K. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2 (2) allele is dominant. *J Clin Invest*. 1989;83(1):314.
19. Meyers JL, Shmulewitz D, Aharonovich E, et al. Alcohol-Metabolizing Genes and Alcohol Phenotypes in an Israeli Household Sample. *Alcohol Clin Exp Res*. 2013;37(11):1872-1881.
20. Macgregor S, Lind PA, Bucholz KK, et al. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum Mol Genet*. 2008;18(3):580-593.
21. Luo X, Kranzler HR, Zuo L, Lappalainen J, Yang B, Gelernter J. ADH4 gene variation is associated with alcohol dependence and drug dependence in European Americans: results from HWD tests and case–control association studies. *Neuropsychopharmacology*. 2006;31(5):1085-1095.
22. Edenberg HJ, Xuei X, Chen H-J, et al. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Hum Mol Genet*. 2006;15(9):1539-1549.
23. Gelernter J, Kranzler HR, Sherva R, et al. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry*. 2014;19(1):41-49. doi:10.1038/mp.2013.145
24. Polimanti R, Gelernter J. ADH1B: From alcoholism, natural selection, and cancer to the human phenome. *Am J Med Genet Part B Neuropsychiatr Genet*. 2017.
25. Consortium T 1000 GP. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
26. O'Connell J, Gurdasani D, Delaneau O, et al. A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genet*. 2014;10(4):e1004234.
27. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.

28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.

29. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006;38(2):209-213.

30. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88(5):586-598.

31. Morris AP. Transethnic Meta-Analysis of Genomewide Association Studies. *Genet Epidemiol*. 2011;35(8):809-822.

32. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.

33. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33(2):272-279. doi:10.1093/bioinformatics/btw613

34. Stringer S, Minică CC, Verweij KJH, et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. *Transl Psychiatry*. 2016;6:e769. doi:10.1038/tp.2016.36

35. Hancock DB, Guo Y, Reginsson GW, et al. Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol Psychiatry*. 2017. doi:10.1038/mp.2017.193

36. Schumann G, Liu C, O'Reilly P, et al. KLB is associated with alcohol drinking, and its gene product β-Klotho is necessary for FGF21 regulation of alcohol preference. *Proc Natl Acad Sci USA*. 2016;113(50):14372-14377.

37. Clarke T-K, Adams MJ, Davies G, 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*. July 2017. doi:10.1038/mp.2017.153

38. Sanchez-Roige S, Fontanillas P, Elson SL, et al. Genome-wide association study of Alcohol Use Disorder Identification Test (AUDIT) scores in 20,328 research participants of European ancestry. *Addict Biol*. 2017. doi:10.1111/adb.12574

39. Demontis D, Walters RK, Martin J, et al. Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv*. 2017:145581. doi:10.1101/145581

40. Duncan L, Yilmaz Z, Gaspar H, et al. Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am J Psychiatry*. 2017;174(9):850-858. doi:10.1176/appi.ajp.2017.16121402

41. Anney RJL, Ripke S, Anttila V, et al. Meta-analysis of GWAS of over 16,000

individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*. 2017;8(1). doi:10.1186/s13229-017-0137-9

42. Sanchez-Roige S, Fontanillas P, Elson SL, et al. Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. *Nat Neurosci*. 2018;21:16-18.

43. Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. *Nature*. 2015;520(7546):224-229. doi:10.1038/nature14101

44. Chambers JC, Zhang W, Sehmi J, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet*. 2011;43(11):1131-1138. doi:10.1038/ng.970

45. Firth D. Bias Reduction of Maximum-Likelihood-Estimates. *Biometrika*. 1993;80(1):27-38. doi:10.1093/biomet/80.1.27

46. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol*. 2015;11(4). doi:10.1371/journal.pcbi.1004219

47. Dudbridge F, Gusnanto A. Estimation of Significance Thresholds for Genomewide Association Scans. *Genet Epidemiol*. 2008;32:227-234.

48. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the Multiple Testing Burden for Genomewide Association Studies of Nearly All Common Variants. *Genet Epidemiol*. 2008;32(4):381-385.

49. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017.

50. Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. *The Lancet Psychiatry*. 2016;3(8):760-773. doi:10.1016/S2215-0366(16)00104-8

51. Mbarek H, Milaneschi Y, Fedko IO, et al. The genetics of alcohol dependence: Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. *Am J Med Genet Part B Neuropsychiatr Genet*. 2015;168(8):739-748.

52. Prescott CA, Aggen S, Kendler KS. Sex-specific genetic influences on the comorbidity of alcoholism and major depression in a population-based sample of us twins. *Arch Gen Psychiatry*. 2000;57(8):803-811.

53. Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry*. 2003;60(9):929-937.

54. Andersen AM, Pietrzak RH, Kranzler HR, et al. Polygenic scores for major depressive disorder and risk of alcohol dependence. *JAMA Psychiatry*. 2017;74(11):1153-1160.

55. Zhou HP, Yang B, Wang Q, et al. Genetic Risk Variants Associated With Comorbid Alcohol Dependence and Major Depression. *JAMA Psychiatry*. 2017. doi:10.1001/jamapsychiatry.2017.3275

56. Barban N, Jansen R, de Vlaming R, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet*. 2016;48(12):1462-1472.

57. Polimanti R, Wang Q, Meda SA, et al. The Interplay Between Risky Sexual Behaviors and Alcohol Dependence: Genome-Wide Association and Neuroimaging Support for LHPP as a Risk Gene. *Neuropsychopharmacology*. 2017;42(3):598-605.

58. Pratt DS, Kaplan MM. Evaluation of Abnormal Liver-Enzyme Results in Asymptomatic Patients. *N Eng J Med*. 2000;342:1266-1271.

59. Polimanti R, Zhang H, Smith AH, et al. Genome-wide association study of body mass index in subjects with alcohol dependence. *Addict Biol*. 2017;22(2):535-549. doi:10.1111/adb.12317

60. Traversy G, Chaput J-P. Alcohol Consumption and Obesity: An Update. *Curr Obes Rep*. 2015;4(1):122-130. doi:10.1007/s13679-014-0129-4

61. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depressive disorder. *bioRxiv*. 2017.

62. Aken BL, Achuthan P, Akanni W, et al. Ensembl 2017. *Nucleic Acids Res*. 2017;45(D1):D635-D642. doi:10.1093/nar/gkw1104