Supplementary Information

A. Supplementary Methods

1 Cohort descriptions

1.1 Case/control cohorts

Comorbidity and Trauma Study (CATS)

Sample description: This study consisted of opioid dependent individuals aged 18 and older recruited from opioid substitution therapy clinics in the greater Sydney area and genetically unrelated individuals with little or no lifetime opioid misuse from neighborhoods in geographic proximity to these clinics. All subjects were of European-Australian descent. Additional details are available in [1].

Alcohol dependence measure: All participants were assessed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA). Alcohol dependence was defined using DSM-IV criteria. For the purposes of these analyses, controls were defined as those who had a lifetime history of alcohol drinking but did not meet criteria for alcohol abuse or dependence. No other comorbid diagnoses were excluded.

Christchurch Health and Development study (CHDS)

Sample description: The Christchurch Health and Development study (CHDS)[2,3] is a longitudinal study of a birth cohort of 1,265 children collected in mid-1977 from urban Christchurch, New Zealand. Data on social circumstances, health, development and wellbeing of the participants was obtained from the cohort at birth, 4 months, 1 year, annually to age 16 years, and at 18, 21, 25, 30, and 35 years. All study information was collected on the basis of signed consent from study participants and all information is fully confidential. All aspects of the study have been approved by the Canterbury (NZ) Ethics Committee.

Alcohol dependence measure: At ages 18, 21, 25, 30 and 35 years cohort members were questioned about their substance use behaviours and problems associated with substance use since the previous assessment (alcohol, tobacco, cannabis, other illicit drugs), using the relevant sections of the Composite International Diagnostic Interview (CIDI) to assess DSM-IV symptom criteria for substance use disorders. Using this information, lifetime alcohol dependence was classified on the basis of whether the participant met DSM criteria for alcohol dependence at any assessment up to age 35.

Collaborative Study on the Genetics of Alcoholism (COGA case/control)

Sample description: COGA is a multi-site study of alcohol dependent probands and their family members. Alcohol dependent probands were recruited from inpatient and outpatient facilities. Community probands and their family members were also recruited from a variety of sources. A subset of alcohol dependent cases and genetically unrelated controls were genotyped using the Illumina HumanMap 1M BeadChip. The sample used here included 847 alcohol dependent cases and 552 controls of European-American descent. Additional details are available in [4].

Alcohol dependence measure: All participants were assessed using the Semi-Structured Assessment for the Genetics of Alcoholism[5,6]. Cases met criteria for a lifetime history of DSM-IV alcohol dependence. Controls reported a history of alcohol drinking, but did not meet criteria for alcohol dependence, abuse or harmful use, nor did they meet criteria for abuse/dependence on illicit drugs.

Study of Addiction: Genetics and Environment (SAGE), Collaborative Genetic Study of Nicotine Dependence (COGEND) & Family Study of Cocaine Dependence (FSCD)

Sample description: Subjects for the Study of Addiction: Genetics and Environment (SAGE) were selected from three large, complementary studies: COGA[7], Family Study of Cocaine Dependence (FSCD)[8], and the Collaborative Genetic Study of Nicotine Dependence (COGEND)[9]. We analyze these subsets separately and remove overlap between cohorts (Supplementary Methods). COGA participants were assessed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA). FSCD and COGEND participants were assessed using polydiagnostic instruments closely based on the SSAGA. Genotyping was conducted using the Illumina Human1Mv1_C BeadChips. Further details of the SAGE samples are available in [10].

Alcohol dependence measure: Cases reported a lifetime history of DSM-IV alcohol dependence. Genetically unrelated control subjects reported alcohol drinking but had no significant alcohol-dependence symptoms and did not meet criteria for a diagnosis of illicit drug dependence.

German Study on the Genetics of Alcoholism (GESGA)

Sample description: Patients were recruited from consecutive admissions to the psychiatry and addiction medicine departments of several German psychiatric hospitals participating in the German Addiction Research Network (for detailed description see [11,12]). All patients were male and of self-reported German ancestry and fulfilled DSM-IV criteria for AD. Control subjects had been drawn from three population based cohort studies (KORA: https://www.helmholtz-

muenchen.de/kora; popgen: https://www.epidemiologie.uni-kiel.de/biobanking/biobank-popgen; HNR: https://www.uni-due.de/recall-studie) in Germany and a Munich community sample.

Alcohol dependence measure: Alcohol dependence was assessed using DSM-IV criteria. Patients received a consensus diagnosis of two clinical psychiatrists and were assessed using one (dependent on recruiting center) of the following (semi-)structured interviews conducted by trained clinical staff members: Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), Composite International Diagnostic Interview (CIDI) or Structured Clinical Interview for DSM-IV (SCID). Control samples are mainly population based and can thus comprise alcohol dependent individuals.

Gene-Environment-Development Initiative (GEDI) – Duke University (GSMS)

Sample description: The Duke arm of the NIDA-funded Gene-Environment-Development Initiative (GEDI) combined existing phenotypic and environmental data from two large prospective studies, the Great Smoky Mountains Study (GSMS) and the Caring for Children in the Community (CCC) study. For each of the two population-based contributing studies, genome-wide genotyping was conducted using a common platform (Illumina Human660W-Quad v1), generating a total genotyped sample of ~1300 subjects. Further details of the GEDI-Duke sample are available in [13,14].

Alcohol dependence measure: Participants of both studies were assessed via structured interviewing using the Young Adult Psychiatric Assessment and its early life extension (i.e., YAPA and CAPA), yielding diagnoses and symptom scales for a wide range of substance use disorders (SUDs). Alcohol dependence was defined using DSM-IV criteria. For the purposes of these analyses, controls were defined as those who had a lifetime history of alcohol drinking but

did not meet criteria for alcohol abuse or dependence. No other comorbid diagnoses were excluded.

Center on Antisocial Drug Dependence (CADD)

Sample description: The sample of 1,901 unrelated adolescents were aggregated from several studies described elsewhere[15–18]. This cohort was over-selected for adolescent behavioral disinhibition, with half of the participants ascertained specifically from high-risk populations (i.e. recruited through substance abuse treatment, special schools, or involvement with the criminal justice system; see supplement of [19] for additional criteria for clinical probands). CADD GWAS participants had an average age of 16.5 (SD = 1.4, range = 13.0–19.9), 28.9% were female, and 37.3% of participants reported non-Caucasian ancestry.

Alcohol dependence measure: Lifetime Alcohol Dependence was assessed with the CIDI-SAM and defined as meeting alcohol dependence at any wave for this longitudinal study.

Phenomics and Genomics Sample (PAGES)

Sample description: Individuals in this study were recruited as part of a large schizophrenia case control sample from the Munich greater area and consisted of stable schizophrenia inpatients or outpatients and healthy volunteers. All participants were genetically unrelated, schizophrenia patients were of Caucasian, psychiatrically healthy volunteers of German descent. Candidates with a history of head injury or neurological diseases were excluded.

Alcohol dependence measure: Alcohol dependence was assessed using DSM-IV criteria using the semi-structured interview Structured Clinical Interview for DSM-IV (SCID) conducted by trained staff members.

Spit for Science (S4S)

Sample description: Subjects were drawn from longitudinal study of college students attending a public university in the mid-Atlantic United States (http://spit4science.vcu.edu)[20]. The current analytic sample consisted of a total of 3,030 cases and controls, including 252 cases and 1863 controls of European-American ancestry and 74 cases and 841 controls of African ancestry. 35.8% of the subjects were male and all subjects were at least 18 years old.

Alcohol dependence measure: Alcohol dependence diagnoses were assessed using items adapted from the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA)[5]. Queried each year, cases met DSM-IV criteria for a diagnosis of lifetime alcohol dependence in at least one wave of data collection.

NIAAA Intramural (NIAAA)

Sample description: Participants were recruited under two NIH Institutional Review Board-approved screening and assessment protocols and were comprehensively assessed at the National Institutes of Health Clinical Center (Bethesda, Maryland, USA) between 2005 and 2015. All participants provided written informed consent. Genotyping of the participants was conducted at the NIAAA Laboratory of Neurogenetics (Rockville, MD, USA).

Alcohol dependence measure: Lifetime alcohol dependence was assessed using DSM-IV criteria.

Mayo Clinic Center for Individualized Treatment of Addiction (CITA)

Sample description: The alcohol dependent patients in this sample were recruited as part of the Mayo Clinic Center for Individualized Treatment of Addiction (CITA) pharmacogenomics study of acamprosate response. The CITA study was approved by the Institutional Review Board of the Mayo Clinic Rochester and Mayo Clinic Health System. All participants signed informed consent approved by the Institutional Review Board and gave permission for use of their data in future genetic studies of alcohol dependence and related phenotypes. This study recruited men and women between the ages of 18 and 80 with a primary diagnosis of current alcohol dependence based on DSM-IV-TR criteria with the last drink 5 or more days before enrollment. We excluded subjects unable to provide informed consent; those unable to speak English; those with psychotic disorders or unstable psychiatric or medical conditions; women who were pregnant, lactating, or planning to become pregnant; subjects taking disulfiram; and those allergic to acamprosate. Participants were recruited from community-based residential and outpatient treatment programs affiliated with Mayo Clinic in Rochester, Minnesota, and the Mayo Clinic Health System sites in Austin, Minnesota, Albert Lea, Minnesota, and La Crosse, Wisconsin. In addition, self-referred participants residing in communities adjacent to referral sites not enrolled in treatment programs but interested in taking acamprosate, were recruited. Detailed description of the study sample, recruitment sites and enrollment procedures are described in earlier publications[21,22].

Controls were selected from the Mayo Clinic Biobank[23]. The biobank participants were mainly recruited from internal and family medicine department at Mayo Clinic and provided broad consent that allowed use of their biological specimens, health-related questionnaire, and

electronic medical records. Potential controls with ICD9 or ICD10 codes in their electronic medical record indicating alcohol use disorders were excluded.

Alcohol dependence measure: In the case sample, a semi-structured interview known as the Psychiatric Research Interview of Substance and Mood Disorders (PRISM) was conducted by trained and certified interviewers and was used to systematically assess for the presence of lifetime as well as current Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) criteria for alcohol dependence. The control sample was not specifically evaluated for substance use disorders as part of this study beyond the ICD code exclusions.

Alcohol Dependence in African Americans (ADAA)

Sample description: Data from "Alcohol Dependence in African Americans: A Case-Control Genetic Study" (ADAA) was funded by NIH grant R01 AA017444. The data were collected between 2009 and 2013 and consisted of cases recruited from treatment centers in St. Louis Missouri and controls screened for the absence of alcohol use disorder recruited from households selected from neighborhoods in proximity to neighborhoods of residence of case participants.

Alcohol dependence measure: Cases met criteria for DSM-IV alcohol dependence. Controls were alcohol-exposed but did not meet criteria for alcohol abuse (DSM-IV).

1.2 Family-based cohorts

Brisbane Longitudinal Twin Study (BLTS)

Sample description: Beginning in 1992, the Brisbane Longitudinal Twin Study (BLTS) consists of 3,561 individuals: 1,422 twin pairs and 717 additional siblings first enrolled at age 12 years and now aged 30 years and older[24] (see also [25]). The sample is: genetically informative (MZ and

DZ twins, and often parents and siblings; genotyped for 610,000 common single nucleotide polymorphisms - SNPs); (b) large; (c) longitudinal with many participants have been assessed at 12, 14, 16 and 21 years of age; (d) well characterized for behavioral and brain-related outcomes; (e) rich in biological samples; and includes (f) a subgroup [n=969] who have undergone MRI scanning. As part of an ongoing US NIH/NIDA funded project beginning 2009, measures of lifetime cannabis use, abuse and dependence data are collected, along with diagnostic data for nicotine, alcohol, and other illicit substances, as well as pilot epidemiological data for ecstasy and methamphetamine use. The average age at interview is 25.65 years (SD=3.65, range=18-38yrs). The entire BLTS sample and 1,549 of their parents have GWAS data (Illumina 610k chip)[26] imputed on the GRCh37 assembly. The final sample included individuals with both genotypic and lifetime alcohol use data.

Alcohol dependence measure: DSM-IV alcohol dependence was coded as the endorsement of 3 or more dependence criteria. Individuals exposed to alcohol were controls.

Gene-Environment-Development Initiative (GEDI) – Virginia Commonwealth University (VTSABD)

Sample description: The VCU arm of the NIDA-funded Gene-Environment-Development Initiative (GEDI) combined existing phenotypic and environmental data from the Virginia Twin Study of Adolescent Behavioral Development (VTSABD) study, a population-based multi-wave, cohort-sequential twin study of adolescent psychopathology and its risk factors, and two follow-up studies, the Young Adult Follow Up (YAFU) and the Transitions to Substance Abuse (TSA) study. For each of the contributing studies, genome-wide genotyping was conducted using a

common platform (Illumina Human660W-Quad v1), generating a total genotyped sample of ~900 subjects. Further details of the GEDI-VCU sample are available in [13,27].

Alcohol dependence measure: Participants were assessed via structured interviewing using the Child Adult Psychiatric Assessment (CAPA), a Structured Clinical Interview for DSM-IV (SCID)-based assessment of psycho-pathology in young adult twins for YAFU and the Life Experiences Interview (LEI) for TSA, yielding diagnoses and symptom scales for a wide range of substance use disorders (SUDs). Alcohol dependence was defined using DSM-IV criteria. No comorbid diagnoses were excluded.

Minnesota Center for Twin and Family Research (MCTFR)

Sample description: The MCTFR is a community-based longitudinal sample including pedigrees designed to include two rearing parents and two offspring[28]. Assessments across subsets of the study varied but were readily harmonized to DSM-IIIR and DSM-IV diagnoses. As part of the GEDI, genotyping was carried out using the Illumina Human660W-Quad array. The final GWAS sample included 1,631 genotyped spouse pairs and 1,404 families with genotyped parents and offspring (at least 1).[29]

Alcohol dependence measure: Cases met criteria for DSM-IIIR alcohol dependence; see Supplementary Note B.1 for more details on the relationship between DSM-IIIR and DSM-IV. Controls reported lifetime alcohol use.

Center for Education and Drug Abuse Research (CEDAR) – Substance Abuse and the Dopamine System Study (SADS)

Sample description: Participants were recruited from the Pittsburgh, Pennsylvania, metropolitan area through newspaper advertisements, social service agencies, substance abuse treatment

programs and various other media. For this project, the sample is drawn from two combined studies with distinct but related ascertainment schemes, from the same Greater Pittsburgh population, joined in the Substance Use Disorder Liability: Candidate System Genes study (R01 DA019157)[30]. CEDAR (P50 DA005605) is a longitudinal family/high-risk study of substance use disorder (SUD)[31]. Parents from a sample of nuclear families, ascertained in CEDAR through the father who did or did not have a DSM-III-R SUD (DSM-IV was introduced after this study started) related to illicit drugs (an illegal substance or nonmedical use of a prescribed psychoactive drug), provided a source for male and female cases and controls. All diagnoses have been revised using DSM-IV criteria, and the SADS participants were diagnosed accordingly. Control subjects had no substance (including alcohol) use disorder, or Axis I or II psychiatric disorder. Participants from the SADS study (R01 DA011922) were males 14-18 years of age having a DSM-IV diagnosis of substance dependence related to use of illicit drugs. In both CEDAR and SADS subsamples, probands having a psychiatric disorder other than SUD qualified for the study unless they had a lifetime history of psychosis or any other condition where valid reporting was uncertain. The vocabulary subscale of WISC-III (subjects below age 16) or WAIS-III (age 16 and older) was administered prior to implementation of the protocol and was required to be in the normal range (>70). Since psychiatric comorbidity is common among substance abusers, cases were not excluded for any Axis I or Axis II disorders. The CEDAR and SADS subjects were self-identified European-Americans from the same Greater Pittsburgh geographic area, and the genomic inflation factor based on all genotyped SNPs, evaluating the excess false-positive rate, was satisfactory at .9812. For this analysis, CEDAR-SADS contributed a sample of 468 European-Americans (169 females and 299 males), average age 25.8 (SD=3.73; range 16.0-34.0) genotyped on Illumina Human660W-Quad BeadChips.

Alcohol dependence measure: Lifetime alcohol use disorder was diagnosed using an expanded version of the Structured Clinical Interview for DSM-III-R-outpatient version (SCID-OP).

Swedish Twin Registry (STR)

Sample description: From the population-based Swedish Twin Registry, participants of the SALT study[32] were invited to the TwinGene study in which blood samples were taken and a simple health checkup performed between 2004 to 2008[33]. Samples from 9,900 unique genomes were genotyped using the Illumina OmniExpress 700K chip.

Alcohol dependence measure: Cases were defined as individuals endorsing 3 or more DSM-IV alcohol dependence criteria. Controls reported alcohol use but did not meet criteria for abuse.

Yale-Penn

Sample description: Yale-Penn subjects were recruited in the eastern US, predominantly in Connecticut and Pennsylvania. They were administered the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)[34] to derive DSM-IV diagnoses of lifetime alcohol dependence (and other major psychiatric traits). The study received IRB approval from all participating institutions and written informed consent was obtained from all study participants. Additional information is available in the relevant GWAS publications (e.g. [35–38]).

Alcohol dependence measure: DSM-IV diagnoses from the SSADDA.

COGA (fam)

Sample description: COGA is a multi-site study of alcohol dependent probands and their extended families (details available in [7]). Initially, a sample of unrelated alcohol dependent cases

(n=847) and alcohol exposed controls aged 25 years or older (n=552) was constructed (COGA-cc)[4]. In a follow-up genotyping effort, a subset of the most genetically informative families was selected for a family-based GWAS (COGA-fGWAS)[39,40]. This sample consisted of 118 European-American families with 2,232 individuals with genotyping data.

Alcohol dependence measure: All participants were assessed using the Semi-Structured Assessment for the Genetics of Alcoholism[5]. Cases met criteria for a lifetime history of DSM-IV alcohol dependence. Controls reported a history of alcohol drinking, but did not meet criteria for alcohol dependence, abuse or harmful use.

Australian Alcohol and Nicotine Studies (OZ-ALC-NAG)

Sample description: Participants were recruited from twins and their relatives who had participated in questionnaire- and interview-based studies on alcohol and nicotine use and alcohol-related events or symptoms (as described in [41]). They were living in Australia and of predominantly European ancestry.

Alcohol dependence measure: Assessed using DSM-IV criteria. Most alcohol-dependent cases were mild, with 70% of those meeting alcohol dependence criteria reporting only three or four dependence symptoms and fewer than 5% reporting seven dependence symptoms.

Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD)

Sample description: Participants in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD)[42] were recruited in Ireland and Northern Ireland between 1998 and 2002. Briefly, probands were ascertained in community alcoholism treatment facilities and public and private hospitals. Probands were eligible for inclusion if they met DSM-IV criteria for lifetime AD and

if all four grandparents had been born in Ireland, Northern Ireland, Scotland, Wales, or England. Probands, siblings, and parents were interviewed by clinically trained research interviewers, most of whom had extensive clinical experience with alcoholism. We assessed lifetime history of AD using a modified version of the Semi-Structured Assessment of the Genetics of Alcoholism (SSAGA) interview, version II[5], demographic characteristics, other comorbid conditions, alcohol-related traits, personality features, and clinical records. All participants provided informed consent. We included 815 probands and siblings in genotyping. Controls were genotyped from 2,048 DNA samples from healthy, unpaid volunteers donating blood at the Irish Blood Transfusion Service and obtained from the Trinity College Biobank https://www.tcd.ie/ttmi/facilities/trinity-biobank/ at Trinity College Dublin. Biobank controls were eligible if they denied any problems with alcohol or history of mental illness and if all four grandparents had been born in Ireland, Northern Ireland, Scotland, Wales, or England. Because of the sample source, controls were not formally screened for AD. Information about age and sex was available for these subjects.

Alcohol dependence measure: DSM-IV criteria for lifetime AD. Because of the sample source, controls were not formally screened for AD.

1.3 Summary statistics cohorts

Netherlands Study of Depression and Anxiety / Netherlands Twin Register (NESDA/NTR)

Sample description: Unrelated participants of European ancestry from the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Registry (NTR) were included in the analyses (N=2,023). NESDA is a longitudinal study focusing on the course and consequences of depression and anxiety disorders. Subjects for NESDA were recruited from three sources,

namely the general population, mental health organizations and general practices[43]. NTR participants are ascertained because of the presence of twins or triplets in the family and consist of multiples, their parents, siblings and spouses. Twins are born in all strata of society and NTR represents a general sample from the Dutch population[44,45].

Alcohol dependence measure: In NESDA, lifetime AD diagnoses according to DSM-IV were ascertained using the Composite Interview Diagnostic Instrument. From NESDA, healthy controls were selected including participants without lifetime AD and alcohol abuse diagnoses, and those never exposed to alcohol. In NTR, controls were added if they score low on CAGE (=0) and low heavy drinking over time (low on alcohol consumption: frequency: less than 1 time per week, and quantity: less than 1 glass per week).

FinnTwin Nicotine Addiction Genetics (NAG-Fin)

Sample description: The NAG-Fin participants originate from the Older Finnish Twin Cohort[46] consisting of adult twins born in 1938-1957. Based on earlier questionnaires, twin pairs concordant for ever-smoking were recruited along with their family members (mainly siblings) for the Nicotine Addiction Genetics (NAG) study[47]. A total of 747 families including 2,193 subjects were assessed by DNA sample collection, structured psychiatric interview based on the SSAGA (Semi-Structured Assessment for the Genetics of Alcoholism), and additional questionnaires. The interview and questionnaires yielded detailed phenotypic information on lifetime smoking behavior and alcohol use, including DSM-IV diagnoses for nicotine and alcohol dependence. Genotype data was generated with the Illumina Human670-QuadCustom BeadChip (at the Wellcome Trust Sanger Institute) and the Illumina HumanCoreExome-12v1-0 BeadChip (at the Broad Institute of MIT and Harvard). Both co-twins from dizygotic twin pairs

were included when available, whereas only one co-twin from each monozygotic pair was included. All available siblings were also included. Altogether 1,576 adults (average age 54.1 (SD 5.9, range 30-79), 59.7% males) were included in the analyses.

Alcohol dependence measure: Lifetime Alcohol Dependence was assessed using the SSAGA.

FinnTwin12 (FT12)

Sample description: The FinnTwin12 participants originate from the Younger Finnish Twin Cohort. FinnTwin12 is a population-based longitudinal study of five consecutive birth cohorts (1983-1987) designed to examine genetic and environmental determinants of health-related behaviors, with a particular focus on use and abuse of alcohol[46]. A total of 1,852 twin individuals from 367 monozygotic and 575 dizygotic twin pairs were assessed by DNA sample collection and structured psychiatric interview based on the SSAGA (Semi-Structured Assessment for the Genetics of Alcoholism) at age 14, and 1,347 of them were interviewed again using the SSAGA as young adults. The age 22 interview yielded detailed phenotypic information on lifetime smoking behavior and alcohol use, including DSM-IV diagnoses for nicotine and alcohol dependence. Genotype data was generated with the Illumina Human670-QuadCustom BeadChip (at the Wellcome Trust Sanger Institute) and the Illumina HumanCoreExome-12v1-0 BeadChip (at the Broad Institute of MIT and Harvard). Altogether 962 subjects (average age 22.4, SD 0.7, range 20-27; 46.8% males) were included in the analyses.

Alcohol dependence measure: Lifetime Alcohol Dependence was assessed using the SSAGA.

National Longitudinal Study of Adolescent to Adult Health (Add Health)

Sample description: The National Longitudinal Study of Adolescent to Adult Health (Add Health) is an ongoing, nationally-representative longitudinal cohort study of 20,000+ adolescents followed into adulthood for 20+ years across five interview waves from 1994-2018. Extensive longitudinal social, behavioral, environmental, and biological data are available, and the design included an embedded genetic subsample of MZ and DZ twins, full sibs, half sibs, and unrelated adolescents in the same household. Genome-wide data are available on 9,975 individuals using two Illumina platforms (Human Omni1-Quad BeadChip, Human Omni-2.5 Quad BeadChip) consisting of 631,990 SNPs. Add Health is a multiracial and multiethnic sample with substantial numbers of individuals with Hispanic and Asian ancestry. For more information about the design of Add Health see [48,49].

Alcohol dependence measure: Lifetime DSM-IV alcohol dependence was assessed using questionnaire modeled on the Composite-International Diagnostic Interview, Substance Abuse Module (CIDI-SAM).

Helsinki Birth Cohort Study (HBCS)

Sample description: The Helsinki Birth Cohort Study (HBCS) is composed of 8,760 individuals born between the years 1934-44 in one of the two main maternity hospitals in Helsinki, Finland. Between 2001 and 2003, a randomly selected sample of 928 males and 1,075 females participated in a clinical follow-up study with a focus on cardiovascular, metabolic, mental, and reproductive health and cognitive function. There were 1,620 women and men (43.4% men) with valid genotype and phenotype data. The mean age of the participants was 61.5 years (SD=2.9). DNA was extracted from blood samples and genotyping was performed with the modified

Illumina 610k chip by the Wellcome Trust Sanger Institute, Cambridge, UK according to standard protocols. Detailed information on the selection of the HBCS participants and on the study design can be found elsewhere[50]. Research plan of the HBCS was approved by the Institutional Review Board of the National Public Health Institute and all participants have signed an informed consent.

Alcohol dependence measure: Alcohol dependence diagnoses were extracted from the Hospital Discharge Register (HDR), which contained data on all hospitalizations in psychiatric and general hospitals in Finland between 1969 and 2008. The HDR also includes personal and hospital ID numbers, dates of hospital admission and discharge, and primary as well as up to three subsidiary diagnoses at discharge. We also identified alcohol dependence diagnoses as causes of death from the National Causes of Death-Register (CDR), which contains records of primary and subsidiary causes of death from all deaths in Finland. Diagnoses were entered into the HDR and CDR according to the International Classification of Diseases, Eighth Revision (ICD-8) until 1986, according to the ICD-9 using the Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition (DSM-III-R) criteria until 1995, and according to the ICD-10 since 1996. In the current study, the primary diagnoses and subsidiary diagnoses of alcohol dependence (ICD-8/9: 303.9 and ICD-10 F10.2) from either register served to index the alcohol dependence. In our sample, we identified 36 cases with alcohol dependence based on the HDR and CDR (2.2% of the total sample).

1.4 Replication cohorts

FINRISK

FINRISK is a population-based cohort study designed to assess risk factors for cardiovascular disease and other chronic diseases. The study design has been extensively described elsewhere[51]. Briefly, independent random and representative population cohorts have been surveyed and interviewed at five-year intervals since 1972. Participants are also linked to population health registries. Genotyping was performed in batches over the study waves using standard genotyping arrays. For the current study, lifetime alcohol dependence status was inferred from ICD codes for hospitalization and cause of death in the linked registry data.

Yale-Penn 2

Participants of Yale-Penn 2 were recruited and ascertained following the same protocol as Yale-Penn 1, described above, with a larger proportion of samples coming from unrelated individuals rather than families. DSM-IV diagnoses of lifetime alcohol dependence were derived from the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)[34]. Genotyping was performed using Illumina HumanCoreExome. Participants were grouped separately from Yale-Penn 1 based on the epoch of recruitment and the platform used for genotyping. Written informed consent was obtained from subjects as approved at each site by the respective institutional review boards, and certificates of confidentiality were obtained from NIDA and NIAAA.

COGA African-American Family GWAS (COGA AA fGWAS)

This cohort from COGA consists of AD probands ascertained through treatment facilities as described for the other COGA cohorts above. Individuals from families that self-identified as Black/African-American were genotyped on the Illumina 2.5M array. Ancestry was further compared across all available EU and AA data and a final set of 2,382 individuals from 482 families comprised the AA family GWAS sample. The AA family cohort has been further described in more detail elsewhere.[52] Cases and controls for the replication analysis were defined in an identical manner to the primary phenotype for this analysis.

1.5 Polygenic risk score cohorts

Avon Longitudinal Study of Parents and Children (ALSPAC)

The Avon Longitudinal Study of Parents and Children (ALSPAC) recruited 14,541 pregnant women residing in Avon, UK, with expected dates of delivery April 1, 1991, to December 31, 1992; 14,541 is the initial number of pregnancies for which the mothers enrolled in the ALSPAC study and had either returned at least 1 questionnaire or attended a "Children in Focus" clinic by July 19, 1999. Of these initial pregnancies, there was a total of 14,062 live births and 13,988 children who were alive at 1 year of age. Subsequent phases of enrollment increased the sample size over time 14,775 live births and 14,701 children who were alive at 1 year of age. The phases of enrollment are described in more detail elsewhere[53,54]. The study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). Ethical approval for the

study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

ALSPAC participants were administered questionnaires online or via post. The wave of data collection included in the current study was sent to participants at approximately age 23 and included questions about past year alcohol use. Study data were collected and managed using REDCap (Research Electronic Data Capture)[55] electronic data capture tools hosted at the University of Bristol. Individuals who responded that they had never had a full drink of alcohol, or that they had not used alcohol within the past year, were coded as missing. Questionnaire items allowed for the calculation of DSM-5[56] alcohol use disorder (AUD) criteria counts, from which a binary diagnosis was then derived based on a cutoff of two or more criteria endorsed. Individuals missing data on half or more of the items were coded as missing; those responding to at least half of the items, but fewer than all of the items, were assigned a prorated score. Phenotypic and genetic data were available for N=2,723 individuals, of whom 337 met criteria for AUD diagnosis.

Generation Scotland (GS)

The Scottish Family Health Study (GS) is a family-based cohort recruited from the general population of Scotland from 2006-2011 (N=24,084)[57]. Genotyping was performed using the Illumina OmniExpress BeadChip and after quality control which removed SNPs with a call rate <98%, a minor allele frequency (MAF) of <1% or those showing deviation from HWE ($p < 5 \times 10\text{-}6$), 561,125 autosomal SNPs and 19,904 individuals were available for analysis. This study obtained informed consent from all participants and was conducted under generic approval from

the National Health Service National Research Ethics Service (approval letter dated 17 June 2011, Ref 11/NW/0382). All components of GS have received ethical approval from the National Health Service Tayside Committee on Medical Research Ethics (REC Reference Number: 05/S1401/89) and written consent for the use of data was obtained from all participants.

Alcohol consumption was assessed using a preclinical questionnaire and participants identified as current drinkers, former drinkers or never drinkers. Alcohol intake was self-reported as units consumed in the previous week. The CAGE questionnaire was administered during a re-contact of GS in 2015[58] and consists of 4 questions designed as a screening tool for alcohol problems[59]. After removing former and never drinkers there were 6,906 individuals available for analysis with both CAGE data and genotype data.

2 Quality control

2.1 Case/control cohorts

Quality control (QC) was performed separately for each case/control cohort using ricopili (https://github.com/Nealelab/ricopili).

Following the standardized ricopili pipeline, variants in each cohort were first filtered for call rate (<5% missingness), followed by individual-level filters for call rate (<2% missingness) and heterozygosity ($|F_{het}| > .20$). If chromosome X variants were available for the cohort the sex checks were also performed to ensure concordance with reported sex. Variants were then filtered for call rate (<2% missingness), differential missingness between cases and controls (absolute difference < 2%), invariant markers, and departure from Hardy-Weinberg equilibrium in cases (P

> 1e-10) or controls (P > 1e-6). In cohorts involving multiple genotyping batches, variants were also filtered for association with batch controlling for the phenotype.

QC was performed prior to estimation of relatedness and principal components (described below). In cases of cryptic relatedness or ancestry outliers, QC was repeated after outlier removal to ensure no additional variants or individuals failed QC after removal of the affected individuals.

2.2 Family-based cohorts

QC for family-based cohorts was performed using picopili (https://github.com/Nealelab/picopili). This QC, imputation, and analysis pipeline was developed for the current analysis with the aim of paralleling the functionality of ricopili (https://github.com/Nealelab/ricopili) with appropriate modifications for the analysis of family-based GWAS cohorts.

QC of the family-based cohorts applied the same basic filters as the case/control QC pipeline (i.e. call rates, heterozygosity, discordant sex checks, differential missingness, and departure from Hardy-Weinberg equilibrium). Where applicable, tests were based on allele frequencies computed from founders in the family-based cohort using PLINK 1.9[60]. In addition, family-based cohorts were QCed to remove individuals or variants with excessive Mendelian error rates. After QC, remaining Mendelian errors were set to missing.

As in the case/control cohorts, QC was repeated after stratification by ancestry and removal of ancestry outliers and instances of cryptic relatedness.

2.3 Summary statistics

For cohorts contributing summary statistics, pre-imputation QC was performed by the respective studies according to their chosen analysis protocols. For HBCS, pre-imputation QC included filtering genotypes for SNP clustering probability for each genotype > 95%, filtering individuals and markers for call rate > 95% (99% for markers with MAF < 5%), and filtering markers for MAF > 1% and HWE p > 1E-6. Heterozygosity and gender checks were performed for all individuals and any discrepancies were removed. In NAG-Fin and FT12, a minimum minor allele frequency of >0.01 and missingness <0.05 were set for analyses. Genotyping and pre-GWAS quality control for NESDA/NTR has been previously described[61]. Briefly, sample QC included filtering for genotyping rate < 90%, absolute heterozygosity statistic (F_{het}) greater than .075, and excessive Mendelian errors or departure from expected gender. Variants were filtered for low minor allele frequency (< .005), departure from Hardy-Weinberg equilibrium ($p < 10^{-12}$), and low call rate (< 95%), along with platform-specific QC. For Add Health, mismatches on heterozygosity and sex were removed but no additional sample filtering was conducted prior to imputation.

3 Principal components analysis and relatedness estimation

3.1 Case/control cohorts

Principal components analysis (PCA) and relatedness estimation were performed within each cohort using a more stringently QCed set of variants. Specifically, variants were filtered for

allele frequency (minor allele frequency > 5%) and Hardy-Weinberg equilibrium (P > 1e-3), and strand ambiguous SNPs and variants in regions of high LD (i.e. the MHC and the chromosome 8 inversion region) were removed. Remaining variants were then pruned for linkage disequilibrium (LD; pairwise $r^2 < 0.2$). Using this strictly QCed set of SNPs, relatedness was then estimated in each cohort using PLINK[60]. Cryptically related pairs of individuals ($\hat{\pi} > 0.2$) were filtered to remove one individual from each related pair, preferentially keeping cases with alcohol dependence and dropping individuals related to multiple other individuals in the cohort.

PCA was then performed using EIGENSOFT[62,63] to infer ancestry. Where appropriate, individuals in a study were stratified by ancestry into European and African ancestry cohorts for analysis (Supplementary Table 1). Additional PCA including 1000 Genomes reference samples were performed to verify the identity of ancestry clusters.

After stratification by ancestry, the full ricopili pipeline of QC, relatedness estimation, and PCA was repeated within each ancestry stratum of each cohort. Remaining PCA outliers within each ancestry group were removed as necessary.

3.2 Family-based cohorts

PCA and relatedness estimation for the family-based cohorts was performed using picopili (https://github.com/Nealelab/picopili). Following the same strategy as the case/control cohorts, variants were first QCed for missingness (< 2%), minor allele frequency (> 5%), and Hardy-Weinberg equilibrium (P > 1e-4). The reported pedigree was used to define founders for computing these filters in PLINK[60]. Indels, strand ambiguous SNPs, and variants in previously

reported regions of high LD[64] were also removed, and remaining variants were pruned for pairwise LD.

For PCA in each cohort, a subset of unrelated individuals ($\hat{\pi} < .09375$; midpoint between 3rd and 4th degree relatives) was identified using PRIMUS[65]. To the extent estimates of relatedness may be upwardly biased in diverse cohorts, membership in the "unrelated" set will be conservative. PCA was then performed in this unrelated set, and the resulting SNP weights were used to project PCs for the remaining related samples using EIGENSOFT[62,63]. This procedure assures that the PCA is performed with unrelated individuals, in order to prevent family structure from biasing the PCA solution[66], while providing results for all individuals. As in case/control cohorts, these PCA results were then used to identify and remove ancestry outliers and to stratify cohorts by continental ancestry group. PCA including 1000 Genomes reference samples was used to confirm the ancestry of each PCA cluster. QC and PCA were repeated within each ancestry group after stratification.

Relatedness estimation was then performed to confirm that genetic relatedness was consistent with the reported pedigree structure of each cohort and to remove instances of cryptic relatedness. In cohorts with a homogeneous population structure after stratification and outlier removal (assessed by visual inspection of PCA results), relatedness estimates were computed using PLINK[60]. For cohorts with remaining structure (e.g. AA cohorts, and Finnish admixture in STR) relatedness was instead estimated using REAP[67]. For estimation in REAP, admixture solutions were estimated using the previously defined "unrelated" set and projected to remaining samples using ADMIXTURE[68].

Relatedness estimates were compared to the reported pedigree structure to identify possible errors or cryptic relatedness. In particular, this filtering aimed to identify: (a) reported parent/offspring pairs with estimated identity-by-descent (IBD) proportions not matching expectations; (b) apparent parent/offspring pairs from IBD values that were not reported in the pedigree; (c) cryptic relatedness ($\hat{\pi} > .09375$) between individuals reported to be in different families; and (d) individuals who were genetically unrelated ($\hat{\pi} < .09375$) to all other individuals of their reported pedigree. Where possible these issues were resolve by confirmation of pedigree data from the original cohort. Unresolved relatedness problems, most commonly instances of cryptic relatedness between families, were then resolved by filtering individuals. As in the case/control cohorts, this filtering prioritized post-QC sample size, preferentially keeping individuals with alcohol dependence, individuals without a missing phenotype, and individuals in larger pedigrees.

3.3 Summary statistics samples

In HBCS, relatedness checks were performed to remove any discrepancies from the expected relatedness structure. Multidimensional scaling (MDS) components were then computed in PLINK. Mismatches between observed and expected relatedness were similarly filtered in NESDA/NTR, and principal components were calculated in Eigenstrat with 1000 Genomes Phase 3 as a reference set[63]. In both NAG-Fin and FT12, both co-twins from dizygotic twin pairs were included when available, whereas only one co-twin from each monozygotic pair was included. NAG-Fin also retained all available siblings. No ancestral outliers were observed in either cohort, and a Genetic Relatedness Matrix (GRM) was used to account for any additional

potential admixture. Similarly, in Add Health, a GRM was computed in GCTA to account for admixture within specified ancestral groups.

4 Imputation

4.1 Case/control cohorts

Imputation of case/control cohorts was performed using ricopili (https://github.com/Nealelab/ricopili). Prior to imputation, each cohort was aligned to 1000 Genomes Project Phase 3 reference data[69,70]. LiftOver[71] to human genome build hg19 was performed if needed, and matching of chromosome, position, and alleles to the reference data was verified. To assist with match strand flips and strand ambiguous SNPs, allele frequencies were also checked against 1000 Genomes reference data. For European ancestry cohorts, SNPs were excluded if their allele frequency difference by more than 0.15 from 1000 Genomes European ancestry individuals; for African ancestry individuals, SNPs were filtered for allele frequency differences greater than 0.25 compared to 1000 Genomes African ancestry individuals. The looser threshold was specified in African ancestry cohorts to account for varying degrees of admixture, and generally yielded higher quality imputation results (data not shown).

After alignment to the 1000 Genomes Project Phase 3 reference[70], each cohort was phased using SHAPEIT[72] and imputed using IMPUTE2[73,74]. Imputation dosages and best-guess genotypes were saved for analysis, as described below. PCA was performed within each cohort using best-guess genotypes to compute principal components (PCs) for use as covariates in GWAS following the same procedure described above. For this post-imputation PCA, best-guess

genotypes were strictly filtered for quality (call rate > 99% for genotype calls with posterior probabilities > 0.8, MAF > 5%) and more stringently pruned for LD (pairwise $r^2 < 0.1$, and removal of additional previously-identified regions of high LD[64]).

4.2 Family-based cohorts

Family-based cohorts were imputed using picopili (https://github.com/Nealelab/picopili) paralleling the same procedure described above for case/control cohorts. Each cohort was matched to the 1000 Genomes Project Phase 3 imputation reference data following the same set of heuristics as are implemented in ricopili. Pre-phasing and imputation were then performed with SHAPEIT[72] and IMPUTE2[73,74] with two primary changes to accommodate the family data. First, phasing was performed for each chromosome rather than in 3 MB genomic chunks in order to assist in identifying any long regions of haplotype sharing between family members. Second, the duoHMM algorithm in SHAPEIT[75] was enabled to allow use of pedigree information in refining haplotype calls.

After imputation, best-guess genotypes were called (minimum posterior probability > 0.8) and QCed for call rate (missingness < 2%), INFO score > 0.6, and allele frequency > 0.005. (Additional filtering was applied prior to meta-analysis, see below.) Any apparent mendelian errors in the imputed pedigrees were set as missing. After QC, post-imputation PCA was then performed to compute PCs for use as covariates in the GWAS using the same protocol as the PCA performed in the family-based cohorts prior to imputation (see above).

4.3 Summary statistics

All summary statistics cohorts were imputed by their respective studies. For HBCS, NAG-Fin and FT12, imputation was performed using the 1000 Genomes Phase I integrated variant set (v3 / April 2012; NCBI build 37 / hg19) as the reference sample with IMPUTE2[73]. NESDA/NTR data were phased and imputed to 1000 Genomes Phase 3 V5 reference panel has been previously described[61]. After imputation all SNPs were converted to best guess genotypes using Plink 1.90[60]. Add Health data were imputed using the Haplotype Reference Consortium on the Michigan Imputation Server[76].

5 Cross-cohort relatedness and ancestry confirmation

After imputation, QCed best-guess genotypes from each cohort were merged to allow filtering for cryptic relatedness between cohorts. Imputed genotypes were filtered for allele frequency and imputation quality (i.e. INFO score, call rate at posterior probability > .80) within each cohort, and then merged and filtered to variants passing QC across cohorts. As in the within-cohort relatedness checks, the passing variants were then pruned for LD and used to estimate genetic relatedness between all pairs of individuals. Relatedness among EU cohorts was estimated using PLINK[60], while relatedness with AA cohorts was estimated using REAP[67] to account for varying admixture.

In cases of observed cross-cohort cryptic relatedness ($\pi > 0.1$), individuals were removed from each related pair as in the within-cohort relatedness filtering. In order to maximize effective sample size, priority was given to keeping individuals with an alcohol dependence diagnosis, individuals in cohorts with small sample sizes, and individuals who were part of a pedigree in a

family-based study. Individuals with cryptic relatedness to a large number of other samples were prioritized for removal. Instances of known overlap between the cohorts (e.g. among the cohorts in SAGE) were also verified and filtered accordingly.

Unrelated individuals were also used to verify ancestry assignment of the EU and AA cohorts, respectively, by merging the cohorts of each ancestry with 1000 Genomes Project reference samples and performing PCA. PCA results from the merged genotyped samples confirm that the AA (Supplementary Figure S12A) and EU cohorts (Supplementary Figure S12B) cluster with the European and African ancestry reference samples, respectively, with the AA cohorts showing evidence of admixture between European and African ancestry that is consistent with other published studies of African-American individuals[77,78].

Table 1 reports final sample sizes for analysis after filtering for cross-cohort relatedness. GWAS were performed separately in each cohort (and for EA and AA within a cohort) using the set of individuals who passed this relatedness check.

6 Genome-wide association

6.1 Case/control cohorts

Genome-wide association studies (GWAS) were performed in each case/control cohort using PLINK[60]. Logistic regression was performed to test association between alcohol dependence and the imputed additive dosage of each variant, controlling for sex and principal components (PCs). Sex was excluded as a covariate in GESGA due to a lack of female cases; instead variants were

filtered to remove any variants with substantial allele frequency differences between male and female controls.

The number of PCs included as covariates to control for confounding from population structure varied by ancestry and sample size. In EU cohorts, the number of PC covariates was determined by cohort sample size in order to reflect differential power of PCA to detect true population structure[79]. Specifically, in EU cohorts with fewer than 2000 samples or fewer than 500 cases, the first 5 PCs were included as covariates; larger cohorts included the first 10 PCs. The number of cases was included as a criterion to prevent over-fitting to PCs in large cohorts with strongly skewed case/control ratios (e.g. S4S).

In AA cohorts, we included as covariates the top PCs associated with genome-wide population structure, as opposed to local ancestry tracts, up to a maximum of 5 or 10 PCs based on the same sample size thresholds as in EU cohorts (see Supplementary Note B.2). In practice, this resulted in the use of between 1 and 5 PCs in each cohort (Supplementary Table S1).

6.2 Family-based cohorts

GWAS was performed in each family-based cohort using imputed genotypes for each variant. The association model used to test association for each variant was selected based on the complexity of the pedigree structure in each cohort's family-based design. Cohorts with a simple pedigree structure were tested using generalized estimating equations (GEE). Cohorts with more complex pedigrees that performed poorly in the GEE model were tested using generalized linear

mixed models (GLMM). Both models are described below. Sex and PC covariates were included following the same protocol as described above for case/control cohorts.

Generalized Estimating Equations (GEE)

GWAS of family-based cohorts with simple pedigrees (Supplementary Table 1) were performed using the GEE model[80]. For family *i* with individual *j* the logistic GEE model specifies the mean and variances of phenotype *Y*

$$E\left(Y_{ij}\middle|\boldsymbol{x}_{ij}\right) = \pi_{ij} = \frac{e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_{ij}'\boldsymbol{\beta}}}$$

$$Var\left(Y_{ij}\middle|\boldsymbol{x}_{ij}\right) = \pi_{ij}(1 - \pi_{ij})$$

with correlation structure

$$Corr\left(Y_{ij}, Y_{kl}\middle|\boldsymbol{x}_{ij}, \boldsymbol{x}_{kl}\right) = \begin{cases} \rho, & i = k \\ 0, & \text{otherwise} \end{cases}$$

where **x** includes an intercept term, the SNP to be tested, and any desired covariates. In other words, the covariance matrix for the observed phenotypes *Y* is block diagonal with the blocks defined by individuals in the same family. This exchangeable correlation structure within family is likely to be correctly specified when all individuals within a family have the same degree of relatedness and that structure is the same across families (e.g. a sib-pair design). For more complex family structures this simple covariance structure is unlikely to hold, which motivates the use of a more flexible generalized mixed model (see below).

GEE models were fit in R using *geepack*[81]. Imputed variants were fit in the model using QCed best-guess genotypes. Robust sandwich standard errors were used to account for possible

misspecification of the block diagonal correlation matrix. GWAS results for a given SNP were evaluated based on the Wald test of the corresponding regression coefficient β.

Generalized Linear Mixed Model (GLMM)

For more complex pedigrees, GWAS was performed using a generalized linear mixed model with logistic link function (i.e. a logistic mixed model[82]). Unlike the GEE, the GLMM is implemented with an arbitrary covariance matrix between individuals, allowing for more complex and varied correlation structures from relatedness within families.

The logistic mixed model is specified similar to a conventional logistic regression, with an added random effects term similar to a linear mixed model. In the generalized form,

$$\eta_i = g(\mu_i) = G_i\beta + X_i\alpha + b_i$$

where *G* are observed genotypes, *X* are other observed covariates, and *g()* is the standard logistic link function.

$$g(\mu_i) = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$$

The random effects term b_i is assumed to follow

$$b \sim N(0, \tau \boldsymbol{K})$$

where **K** is the genetic relatedness matrix (GRM). Arbitrary specification of this GRM **K** is a key feature of the GLMM model.

We fit the GLMM using best-guess genotypes with the package GMMAT in R[83]. The GRM **K** is estimated in PLINK[60] using the same strictly QCed set of SNPs used for post-imputation PCA (see above). As is recommended for mixed models, GRMs are generated following a leave-one-

chromosome-out (LOCO) approach that omits the chromosome containing the SNP to be tested from the calculation of the GRM to prevent confounding[84]. Each SNP in the GWAS is evaluated using a score test; this is necessary to maintain computational feasibility for the GWAS but forgoes calculation of effect sizes and standard errors for each variant[83].

Comparing the GLMM and GEE models, it may be noted that the GLMM implies structured covariance on the latent scale rather than on the observed scale as in the GEE. Both models include logistic regression as a special case, but the GLMM and GEE models are not nested with one another. As might be anticipated by the model differences, simulations show mixed results for which model is preferable depending on the choice of simulation setting[83,85]. Empirically, we do observe less inflation of genome-wide test statistics in cohorts with complex pedigrees when using the GLMM model compared to the GEE model (data not shown). For the current study we rely on both models to maintain compatibility with the conventional logistic regression model for GWAS and choose the most appropriate model for each cohort based on pedigree structure with attention to practical benefits (e.g. interpretable effects sizes, computational tractability) and appropriateness of the accompanying model assumptions (i.e. exchangeable correlations within family).

GWAS of Unrelated Individuals

In addition to the primary family-based analyses, a subset of unrelated individuals was selected from each family-based cohort to perform a conventional case/control GWAS. Unrelated individuals were chosen to maximize the effective sample size for case/control analysis within each cohort. GWAS was then performed using logistic regression with the imputed genotypes in

PLINK[60]. Sex and PC covariates were included following the same protocol as the case/control GWAS, as described above. In EU cohorts, the subset of unrelated individuals was also used to perform sex-specific GWAS, subject to the same sample size requirements as the case/control cohorts.

6.3 Summary statistics

GWAS was performed separately within each summary statistic cohort following standard protocols for the respective studies. In HBCS, association analyses were conducted using the score test from SNPTEST2[86], with covariates including sex, age and the first 3 MDS components. NAG-Fin and FT12 were analyzed using Genome-wide Efficient Mixed Model Association (GEMMA v0.94)[87] with sex and age as covariates and with the computed GRM used to control for any remaining ancestry structure and relatedness. The NESDA/NTR GWAS was conducted using a mixed linear model as implemented in GCTA[84,88] (--mlma-loco) with adjustment for sex, age, age squared, and four principal components of genetic ancestry as covariates. Add Health data were similarly analyzed using a mixed linear model association framework within GCTA[84,88] with sex as a covariate.

7 Genome-wide meta-analysis

We performed three batches of primary meta-analyses. First, we perform meta-analysis of all samples (including related individuals and summary statistic cohorts). Second, we perform meta-analysis of unrelated individuals (i.e. using the GWAS of unrelated individuals rather than GEE or GLMM results for family-based cohorts). Third, we perform meta-analysis of unrelated genotyped samples only (i.e. excluding summary statistic cohorts). Within each of these batches

we stratify by ancestry. The full set of meta-analysis designs is described in Supplementary Table S2.

7.1 Meta-analysis with related samples

The primary discovery meta-analysis is performed using all available samples, including related individuals and summary statistic cohorts (14,904 cases, 37,944 controls). In addition to this primary meta-analysis across ancestries, meta-analysis is also performed within AA cohorts (3,335 cases, 2,945 controls) and EU cohorts (11,569 cases, 34,999 controls) separately.

These meta-analyses were performed using p-values with weights defined by the effective sample size of each cohort. These weights were defined to account for the differences in case/control balance and degree of relatedness within each cohort, while allowing meta-analysis without comparable effect size estimates from the GLMM or summary statistic cohorts (see Supplementary Note B.3 for more detail).

For meta-analysis, results from each cohort were filtered for imputation INFO score (> 0.8), minor allele frequency (> 1%), and expected minor allele count (MAC) in cases and controls (> 5). GWAS results from summary statistics cohorts were filtered according to the same criteria after being aligned to match the same genomic reference as the genotyped cohorts (e.g. matching rsids, positions, and alleles). Cohorts with an extreme case/control ratio (i.e. STR and HBCS) were more strictly filtered to require MAF in controls corresponding to MAC >5 alleles in cases and MAC >10 alleles in cases and controls. This stricter filtering addressed observed instability in the results for these cohorts at low allele counts. Results from each meta-analysis were further

filtered to only report results for variants with an effective sample size > 1000 and > 15% of the maximum effective sample size for the meta-analysis, as well as requiring expected minor allele counts of at least 20 across the included cohorts. These filtering criteria were also applied to subsequent meta-analyses.

7.2 Meta-analysis with unrelated samples

To support planned secondary analyses, we also performed genome-wide meta-analyses restricting to primarily unrelated samples. In particular, for family-based cohorts the case/control GWAS of unrelated individuals was included in the meta-analysis rather than the family-based GEE or GLMM analysis. This analysis was performed separately for each ancestry group. Final sample sizes for these meta-analyses of unrelated individuals were 2,991 cases and 2,808 controls for AA, and 10,206 cases and 28,480 controls for EU.

These meta-analyses were designed to allow secondary analysis with methods that depend on the relationship between sample size and p-values as an indicator of effect size but do not directly require effect size estimates. For the current paper, this principally includes LD score regression[89] and gene-based analysis with MAGMA[90]. As appropriate for these analyses, we also focus on ancestry-specific meta-analyses to allow modelling of the different LD structure within each ancestry.

We note that the inclusion of the summary statistic cohorts in this meta-analysis means the included individuals are not fully unrelated. Most of the summary statistic cohorts included some number of related individuals, most frequently within a mixed model framework. Throughout

this paper, however, we treat the summary statistic cohorts as analyses of unrelated individuals since information on the degree of relatedness within the cohort is unavailable, i.e. to compute effective sample sizes. Therefore we included them in this analysis of unrelated individuals as well. We note however that comparison of LD score regression results to analysis excluding the summary statistic cohorts suggests that the impact of this inclusion is minimal (see below), with a primary benefit of increasing sample size and thus improving precision.

7.3 Meta-analysis with unrelated genotyped samples

Meta-analysis of unrelated genotyped samples was performed using conventional inverse-variance weighted fixed effects meta-analysis in METAL[91]. This analysis excluded the summary statistic cohorts and restricted the family-based cohorts to unrelated individuals only. Meta-analysis was performed for both European (EU) and African (AA) ancestry cohorts. Total sample sizes for this meta-analysis were 8,485 cases and 20,272 controls in EU cohorts, and 2,991 cases and 2,808 controls in AA cohorts.

This analysis was primarily intended to provide estimates of variant effect sizes, and also served as the baseline for conditional analysis of independent effects in the chromosome 4 locus. This restricted set of samples is necessary for estimation of effect sizes because many of the summary statistic cohorts relied on GWAS with a linear rather than logistic link function and thus do not have comparable effect sizes to the genotyped cohorts, and because effects sizes are unavailable for the family-based cohorts with complex pedigrees analyzed using the GLMM score test.

7.4 Trans-ancestral modelling

To fully evaluate the pattern of genetic effects between the EU and AA ancestry cohorts we considered multiple models for trans-ancestral meta-analysis. These methods have been developed to identify genetic effects that may not be well represented by conventional fixed effects meta-analysis. In particular, this includes any instances of ancestry-specific effects or ancestry-specific differences in the magnitude of an effect which could be related to differences in allele frequency, LD structure, or other factors.

Specifically, we evaluated the modified random effects model proposed by Han & Eskin[92] and MANTRA, a Bayesian method proposed by Morris[93]. We apply both methods since they have both been evaluated to perform well for trans-ancestral modelling[94] but are based on distinct models.

The Han & Eskin random effects model[92] combines the test of mean effects for a variant in a random effects model with the test for heterogeneity at the variant. This combined test evaluates a null hypothesis that the variant has no association with the phenotype in all cohorts, with an alternative hypothesis that there may be either a non-zero average association across cohorts or variation between cohorts (which implies the association must be non-zero in at least one cohort). This contrasts with a traditional random effects model, which treats cross-cohort variability with zero mean effect as a null result. We perform the trans-ancestral meta-analysis using the Han & Eskin model[92] as implemented in Metasoft (http://genetics.cs.ucla.edu/meta/index.html) and evaluate significance with the conventional 5E-8 p-value threshold.

MANTRA[93] is a Bayesian model that considers potential clustering of effects by ancestry. True marginal effects within each cluster are assumed to be normally distributed, with a prior geometric distribution on the number of clusters. Clustering of populations is informed by the pairwise genetic distance (F_{ST}) between the populations. The fitted model is compared to the null hypothesis that the variant's marginal effect is zero in all ancestries and evaluated using Bayes' Factor (BF) for this model comparison. We defined genome-wide significance for this test as log(BF) > 6.1 based on previous work suggesting that this BF threshold provides a similar false positive rate and statistical evidence against the null hypothesis as the p < 5E-8 threshold for GWAS[94]. For the current study MANTRA was implemented using software provided by the method's author[93]. Default priors were used for the probability of heterogeneity across ancestry (0.5), and the mean (uniform) and variance (exponential with expected value of 1) parameters for the distribution of effect sizes within each ancestry cluster. As previously noted[93], these priors are intentionally weak and used for computational efficiency. Estimation is performed using a Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm with default burn-in and convergence criteria.

Because both the MANTRA and Han & Eskin methods involve modelling differences in the estimated effect size between ancestries they can only be evaluated among the meta-analysis cohorts with effect size estimates, namely the analyses of unrelated individuals from genotyped cohorts (11,476 cases, 23,080 controls; Supplementary Table S2). Inverse-variance weighted meta-analysis was performed within each ancestry group (i.e. EU and AA) before trans-ancestral meta-analysis of those two sets of results. For comparison, we also perform fixed effects meta-

analysis of the EU and AA results using conventional inverse-variance weights as a baseline for evaluating the impact of additional modelling for trans-ancestral effects.

For these analyses we apply the same per-cohort GWAS QC (e.g. INFO score, MAF) as in the meta-analysis of unrelated individuals. Variants are excluded from the trans-ancestral meta-analyses if they aren't present in both EU and AA ancestries, if the effective sample size is less than 15% of the total effective sample size, or if the expected minor allele count in cases is < 20.

All three trans-ancestral meta-analyses (fixed effects, Han & Eskin random effects, and MANTRA) yielded genome-wide significance for the chromosome 4 *ADH1B* locus, as indexed by rs1229984 (Supplementary Figure S1), and rs9571413 on chromosome 13 (not shown). The results for the three meta-analysis methods are highly similar both genome-wide and for specific associations in the *ADH1B* locus. There is no evidence of associations identified by the trans-ancestral models that are sufficiently heterogeneous across ancestries that they go undetected by the fixed effects model. This provides reassuring evidence to support the use of the fixed effects model for the primary meta-analysis of EU and AA cohorts in the full data where use of the random effects or MANTRA methods is prevented by the lack of effect size estimates.

The chromosome 13 SNP, rs9571413, is an uncommon intergenic SNP. The results suggest that the minor allele is a risk variant (fixed effects OR=1.326) just surpassing genome-wide significance in each analysis (fixed effects p=3.90E-8; random effects p=4.83E-8; MANTRA log(BF)=6.11). In contrast, this variant only nominally approached significance in the primary discovery meta-analysis (p=1.54E-5), reflecting much stronger evidence of association among

EU case/control cohorts (p=5.19E-6) than in EU family cohorts (p=.0743), summary statistic cohorts (p=0.979), or AA cohorts (p=.118). These differences are insufficient to demonstrate significant heterogeneity between all cohorts (I^2=20.5, p=.196) or between EU case/control and family cohorts (I^2=69.7, p=.069) but do cast doubt on the nominally significant association observed among unrelated genotyped individuals. Considering the weaker result for rs9571413 in the primary discovery meta-analysis, along with the nominal significance of the SNP (which would not survive correction for the multiple meta-analysis versions in the current paper nor adjustment for genomic control), we do not reject the null hypothesis for association of rs9571413 with alcohol dependence in the current paper.

8 Cross-cohort heterogeneity

While the trans-ancestral meta-analysis methods aim to use trans-ancestral differences for fine-mapping and to improve sensitivity to loci with varying effect sizes across cohorts, is it also important to evaluate potential systematic differences between cohorts related to other study design factors. Such study is particularly important to identify areas where the fixed effects meta-analysis may be misleading.

For that reason, we evaluated heterogeneity using Cochran's Q test[95] for both the omnibus test of heterogeneity between all cohorts and targeted comparisons with fewer degrees of freedom between sets of cohorts defined by differences in study design. In particular, we evaluated:

- The omnibus test of heterogeneity between all cohorts in the discovery meta-analysis (Supplementary Figure S6)
- The omnibus test of heterogeneity among AA cohorts (Supplementary Figure S7A)

- The omnibus test of heterogeneity among all EU cohorts (Supplementary Figure S7B)
- The 1 degree of freedom test of heterogeneity between EU and AA cohorts (Supplementary Figure S7C)
- The 1 degree of freedom test of heterogeneity between the EU family-based cohorts with simple (GEE model) versus complex (GLMM model) pedigrees (Supplementary Figure S8A)
- The 1 degree of freedom test of heterogeneity between family-based and case/control EU cohorts (Supplementary Figure S8B)
- The 1 degree of freedom test of heterogeneity between genotyped and summary statistic EU cohorts (Supplementary Figure S8C)

All tests of heterogeneity were done based on the meta-analysis of P values under a fixed effects model with weights defined by effective sample size.

One variant, rs4673609, reached genome-wide significance for heterogeneity among the African ancestry cohorts (p=8.78e-10). Heterogeneity for this variant primarily reflects opposing trends for association of the A allele in FSCD (OR=1.62, p=1.42E-3) and CADD (OR=3.93, p=7.32E-4) compared to NIAAA (OR=0.53, p=5.92E-4) and COGEND Nico (OR=.37, p=2.04E-4). The European ancestry meta-analysis does not show any trend towards association (p=.545) or heterogeneity (p=.760) for this variant, nor is there a trend towards association in any individual cohort ($p > 0.1$ in all cohorts). Heterogeneity at this variant in AA cohorts may reflect differences in background haplotypes, statistical artifacts from the small cohorts with the observed trends, or other study-specific factors, but nevertheless the observed heterogeneity appears restricted to this single variant.

Apart from this single variant, none of these comparisons identified significant inflation of genome-wide heterogeneity statistics between the different study designs in the discovery GWAS (Supplementary Figures S6-S8). This broad consistency across study designs supports the use of the fixed effects meta-analysis as the primary discovery GWAS results for the current study of alcohol dependence.

9 Assessment of genome-wide significant loci

9.1 Conditional analysis of the *ADH1B* region

Clumping GWAS results from the primary discovery meta-analysis suggested that the chromosome 4 locus may contain multiple independent effects, both within and between ancestry. Supplementary Figures S2A and S2B illustrate the pattern of LD in the chromosome 4 locus in European and African ancestry reference data, respectively. Estimated D' values suggest a primary central haplotype structure covers most of the locus, but with partially independent clusters of additional variants on both sides of that core signal.

To further evaluate the possibility of independent effects we performed conditional analysis for all variants in the locus controlling for the lead variant. Specifically, we performed GWAS in European ancestry cohorts controlling for rs1229984 as a covariate; for African ancestry cohorts conditional analyses controlled for rs2066702 and also included rs1229984 as a covariate in cohorts where it passed imputation quality filters. Other covariates were kept the same as the primary GWAS in each cohort. Analysis was performed using unrelated genotyped samples to enable comparison of effect sizes between the conditional and marginal GWAS results and

because individual-level data were not available for conditional analysis in the summary statistic cohorts.

We compare results in each ancestry and trans-ancestral analysis under the fixed effects model with inverse variance weights (Supplementary Figure S3). Detailed results for the putative independent effects, based on LD clumping in each population of the discovery GWAS, are reported in Supplementary Table S3. We also note that two of the SNPs identified by LD clumping, are in strong LD with known missense variants rs698 and rs1693482 in *ADH1C*; results for conditional analysis of these two coding SNPs are also reported in Supplementary Table S3.

No variants reached genome-wide significance in the conditional analysis (Supplementary Figure S3), including the variants identified as potentially independent signals in the full discovery GWAS (i.e. LD $r^2 < 0.1$ with the *ADH1B* index SNP in the relevant population in 1000 Genomes reference data). However, the suggested independent variants also are not significant in the marginal analysis of the unrelated genotyped samples available for the conditional meta-analysis (Supplementary Table S3), reflecting the reduced sample size of meta-analysis in the unrelated genotypes samples compared to the full discovery GWAS used to identify the suggested independent variants. When directly comparing the marginal and conditional GWAS in the unrelated genotypes samples, most of the suggested independent variants have a modestly attenuated effect size and an increased standard error in the conditional analysis, leading to less significant results (i.e. higher P values). The effect sizes weren't fully attenuated to the null in the conditional analysis, however, and variants rs3811802 and rs894368 did not show attenuation

in the conditional trans-ancestral analysis. In addition, the conditional trans-ancestral analysis suggests rs112346244 as an alternative lead variant, as part of a highly correlated haplotype, versus those identified by LD clumping of the discovery GWAS. While none of these results are conclusive evidence of an independent signal at any specific variant in the region, it is suggestive and worthy of attention in future analyses.

9.2 Association with eQTLs for expression of ADH1B

Given the association of AD with *ADH1B* coding variants, we also tested whether variants affecting *ADH1B* expression (eQTLs) were associated with AD. Considering data from the Genotype-Tissue Expression (GTEx) project[96] (V7; available at https://www.gtexportal.org/), 262 variants were reported to affect *ADH1B* expression in different human tissues (FDR $q<0.05$), although not significant in liver. After LD-informed clumping and the exclusion of variants in LD with the genome-wide significant 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 EA analyses (trans-ancestry p = 4.72E-8; EA p = 3.97E-6; AA p = 1.97E-3). The other two LD-clumped eQTLs, rs11939328 and rs7664780, did not show any association with AD ($p > 0.05$). Located in the intergenic region between *ADH1C* and *ADH7*, rs10516440 is a LD proxy ($r^2 > 0.9$) of rs6827898 (Table 2) in populations of European and African descent. The rs10516440*A allele was associated with reduced AD risk and increased *ADH1B* expression, in line with the effect of the coding variants where the protective allele is associated with increased

ADH1B enzymatic activity. The association of rs10516440*A with expression of *ADH1B* was consistent across multiple tissues (multi-tissue p = 1.42E-76). The same rs10516440 variant is also an eQTL for *ADH1A* (multi-tissue p = 6.72E-33), and *ADH1C* (multi-tissue p = 1.9E-39).

9.3 Evaluation of novel association on chromosome 3 (rs7644567)

Because the observed association (p < 5E-8) of rs7644567 with AD in the discovery meta-analysis is novel but has concerning statistical properties (e.g. lack of other associated SNPs in the locus, few cohorts contributing to the association) we carefully scrutinized rs7644567. As described below, we find that rs7644567 clearly passes QC thresholds for inclusion in the current meta-analysis but has limited supporting evidence for association from LD proxies and fails to replicate in three external cohorts.

QC metrics for rs7644567

Rs7644567 is rare in non-Finnish European populations (non-Finnish European ancestry MAF=.007, Finnish ancestry MAF=.066) and relatively common in those of African ancestry (MAF=.329; based upon 1000 Genomes populations[69]). We observed allele frequencies consistent with the reported population allele frequencies within each of these ancestries. For the discovery meta-analysis, results for rs7644567 reflect contributions from six genotyped AA cohorts and two summary statistics cohorts from Finland. In the two remaining AA cohorts rs7644567 failed to reach the imputation quality threshold (INFO .74-.77). Nearly all of the European ancestry cohorts either failed to impute rs7644567 with sufficient quality (INFO < 0.8; 12 cohorts) or the variant was too rare (MAF < .01 or expected MAC < 5; 12 cohorts). In the one remaining European ancestry cohort logistic regression failed to converge for the variant. The

large number of cohorts where rs7644567 failed to meet MAF and MAC thresholds for inclusion in the meta-analysis is consistent with the limited MAF, along with the limited number of variants in LD with rs7644567, as described below. By comparison, rs7644567 was well imputed in the 6 passing AA cohorts (INFO 0.96-1.05) and two summary statistics cohorts from Finland (INFO 0.96-0.98). The 2 AA cohorts with lower imputation quality are distinguished by having fewer and less informative (in terms of LD with rs7644567) genotyped SNPs in the region.

In sum, the availability of results passing QC for rs7644567 are fully consistent with the expected population allele frequencies and available variants for imputation, and the variant easily passes QC criteria in the included cohorts.

LD proxies of rs7644567

The regional Manhattan plot of rs7644567 (Supplementary Figure S4A) highlights the lack of other genome-wide significant variants in the locus. Focusing on the results from AA cohorts, we observed a handful of variants with meaningful LD to rs7644567 that nevertheless have limited association with AD (Supplementary Figure S4B). The strongest observed LD proxy was rs13098461 (r^2 = .877 in African ancestry populations from 1000 Genomes), with a limited number of other proxies in the region (4 variants with r^2 > .60; proxies identified with LDlink[97]), consistent with the variants present in the AA meta-analysis.

The lead proxy variant, rs13098461, had an estimated effect size similar to rs7644567 in the unrelated AA individuals (rs13098461*T OR=1.238, p=2.2E-3; compared to OR=1.229, p=4.3E-5 for rs7644567*A) but remained much less significant due to a lower available sample size

from cohorts where the variant passes QC. On the other hand, rs13098461 is more common in European ancestry populations (MAF=.061) and is the best available proxy for rs7644567 in European populations with MAF > .01 (identified with LDlink[97]). Specifically, the rs13098461*C allele remains predominantly on the same haplotype as rs7644567*G, though with a weaker correlation due to the difference in allele frequency (r^2=.298, D'=1.0). The association of rs13098461 with AD showed a much weaker effect size in unrelated genotyped EU individuals (OR=1.073, p=.275) and no evidence of association in the full EU discovery meta-analysis (p=.786). As a result, the full discovery meta-analysis reports weaker evidence of association with AD for rs13098461 than rs7644567 despite a much larger available sample size in the current meta-analysis (rs7644567 p = 1.36E-8, effective N = 6,204; rs13098461 p = .2358, effective N = 22,246). This result doesn't rule out association of rs7644567 with AD, but it does suggest any true association of rs7644567 is through effects poorly tagged by rs13098461 in European ancestry individuals.

Replication analysis

We tested whether the observed association of rs7644567 with AD replicated in three external cohorts: FINRISK, Yale-Penn 2, and COGA AA fGWAS. Full descriptions of each cohort are in Section 1.4 above. These cohorts allow evaluation of the association in both African and Finnish ancestries, the two ancestries contributing to the significant result in the discovery GWAS.

Results for the test of association of the rs7644567*G allele with AD in each of the three replication cohorts are reported in Supplementary Table S4. In all three replication cohorts the rs7644567*G allele trends towards association with a risk-increasing effect, as indicated by the

sign of the Z scores, as opposed to the risk-decreasing effect reported for that allele in the discovery meta-analysis. In the FINRISK cohort this reversed effect is nominally significant (p=.019), though would not survive multiple testing correction for three replication cohorts. Taken together, these sign discordant effects and nominally significant trend in the opposed direction clearly do not replicate the association of rs7644567 with AD reported by the discovery meta-analysis. As a result, we conclude that the current study has insufficient evidence to firmly reject the null hypothesis for association of rs7644567 with AD.

Biological annotation of rs7644567

Although the evidence for association of rs7644567 with AD is inconclusive in the current meta-analysis, we briefly review existing evidence for the biological effects of this variant.

rs7644567 is an intergenic SNP located 120 kb upstream of *RBMS3*. It has not been significantly associated with any other phenotypes in GWAS Catalog[98]. Given that rs7644567 is rare in European populations however, it is unlikely to be well-covered in many genome-wide studies.

Among the brain regions available in GTEx[96] V7, eQTL results suggest the rs7644567*A allele is nominally associated with increased expression of *RMBS3* in cerebellar hemisphere (p = 9.9E-3). Analysis of Hi-C data with HUGIn[99–101] suggests the region containing rs7644567 has Bonferroni significant chromatin contact with *RBMS3* and its promoter region in the liver and a neural progenitor cell line, and may also have contact with *RBMS3* and its promoter region in hippocampus (FDR < .05; Supplementary Figure S5). *RBMS3* encodes an RNA-binding protein that is upregulated in liver fibrosis[102] and has been suggested as a tumor suppressor gene for

multiple cancers[103–105]. The HUGIn analysis also indicates chromatin contacts of the rs7644567 region with points near and within *GADL1* in liver. *GADL1* encodes a glutamate decarboxylase-like protein involved in taurine synthesis rather than glutamate decarboxylation[106], and variants in *GADL1* are strongly associated with blood metabolite levels[107].

10 Power analysis

10.1 Power to detect loci at $p < 5e\text{-}8$ and $p < 1e\text{-}6$

Power calculations for the current meta-analysis were performed using CaTS[108], which is freely available for download (http://csg.sph.umich.edu/abecasis/cats/download.html). CaTS estimates the power of GWAS of a dichotomous phenotype to detect a risk variant with a given allele frequency and effect size (i.e. relative risk [RR]) at a specified significance threshold given the number of cases and controls and the population prevalence of the phenotype.

For the current study we evaluated the power for common variants (MAF > .01) with odds ratios (ORs) between 1.05 and 1.3. ORs in this range are consistent with the effects of top loci identified for other complex traits (though it is likely that many additional variants have effect sizes below this range). We convert ORs to RRs following the approximation derived by [109]:

$$RR = \frac{OR}{(1 - K) + (K * OR)}$$

where K is the population prevalence of the phenotype.

We consider power for the full discovery meta-analysis, as well as the ancestry-specific discovery meta-analyses for EU and AA. For power analysis in EU and AA we assume the population prevalence of AD in alcohol-exposed individuals is .159 and .111, respectively[110]. For

power analysis of the full discovery meta-analysis we take a weighted average of these prevalences, proportional to sample size in the current study, yielding K=.151.

Sample sizes for the current meta-analysis were specified using the effective sample size calculations used for weighting the meta-analysis (see Section B.3). Consistent with those derived sample sizes we assume the effective sample size is consistent with a GWAS of equal numbers of cases and controls. For example, we compute power for the full discovery sample (N_{eff}=31,844) assuming it is equivalent to a GWAS of 15,922 cases and 15,922 controls.

The results for these power calculations are shown in Supplementary Figure S11. As reported in Supplementary Figure S11, we estimated power to reach genome-wide significance ($p < 5E-8$) in the current meta-analysis of AD, as well as power to reach $p < 1E-6$. The latter threshold is of interest due to the observation that relatively few loci in the current study reach this threshold compared to GWAS of other complex traits. Thus although power to identify genome-wide significant effects in the current GWAS may be somewhat limited, there is better power to identify suggestive evidence for loci at $p < 1E-6$. Therefore, the limited number of loci reaching the more liberal threshold provides stronger evidence that remaining variants associated with AD that are not detected in the current analysis (but whose existence are implied by the significant polygenicity and SNP-heritability estimates in EU and AA from LDSR) are expected to have smaller ORs and/or lower MAF.

10.2 Comparison to GWAS effect sizes of other disorders

To provide additional perspective on the power in the current GWAS of AD, we compared the current power analysis to reported effect sizes from large GWAS of other disorders. This analysis aims to provide better intuition about the meaning of the range of effect sizes that the power analysis indicates the current GWAS of AD would be well powered to detect at either the level of genome-wide significance or $p < 1E\text{-}6$.

For this comparison, we evaluated the number of loci the current GWAS of AD would be expected to identify if the top loci for AD have effect sizes and allele frequencies comparable to each of three disorders: schizophrenia[111], class I obesity[112], and major depression[113]. We first tabulated the reported independent genome-wide significant loci from the largest available GWAS of each disorder[111–113], including the reported OR and MAF of the index variant in each locus. Where possible (schizophrenia and obesity), we focused on the reported OR from the replication portion of the study to avoid biasing of effect sizes by winner's curse. For schizophrenia we also restricted to loci that are genome-wide significant in the initial discovery GWAS to further protect against winner's curse. From the available results this yielded ORs and MAFs for 105 reported loci for schizophrenia, 25 loci for class I obesity, and 44 loci for major depression.

For the reported loci, we then computed power to detect effects with the given OR and MAF in the current discovery GWAS of AD, as well as power in the EU discovery GWAS and the GWAS of unrelated EU samples. Assuming the loci are independent, this power per locus can then be summarized by using the binomial probabilities to compute the expected number (and

95% confidence interval) of loci from the observed distribution of OR and MAF from each of the disorders that would be detected at a given alpha level in the current GWAS of AD.

Supplementary Table S8 reports the number of loci the current GWAS of AD would be expected to detect (and corresponding 95% confidence intervals) at $p < 1E\text{-}6$ and $p < 5E\text{-}8$ if the true effect sizes of top loci in AD matched the effect sizes and odds ratios reported for top loci in schizophrenia[111], class I obesity[112], or major depression[113]. These can be interpreted as a lower bound on the number of loci the current AD GWAS would be expected to detect if the full distribution of effect sizes matched these disorders, since the reported numbers do not account for additional loci with true effects that have not been identified as genome-wide significant in the respective disorders. The expected number of loci also does not include the probability of loci reaching the given alpha under the null (e.g. given one million independent loci that all have null effects we would expect on average one locus to reach $p < 1e\text{-}6$ in a given GWAS).

The results shown in Supplementary Table S8 suggest that the current discovery GWAS of AD observes significantly fewer loci at the genome-wide significance threshold or at $p < 1E\text{-}6$ than would be expected if the top loci for AD had effect sizes similar to schizophrenia or class I obesity. The same pattern is observed when focusing on the GWAS of unrelated EU samples, suggesting that mild heterogeneity in the trans-ancestral or family-based analysis would not be sufficient to explain few observed loci in the current GWAS if true effect sizes resembled schizophrenia or obesity. Instead, the GWAS results for AD could be more consistent with the top true effect sizes for AD being similar to, or perhaps slightly larger than, the effect sizes reported for top loci in major depression. We also note that AD also has a prevalence and SNP-

heritability more similar to major depression than schizophrenia. In sum, the comparative power analysis suggests that sample sizes somewhere between those for class I obesity (55,229 cases and 104,894 controls to identify 25 significant loci) and major depression (130,664 cases, 330,470 controls to identify 44 loci) may be required to identify a larger number of top common variant loci associated with AD similar to these other disorders.

11 LD score regression intercept and SNP heritability

The proportion of variance explained by all common SNPs – i.e. the SNP-heritability h^2_g was estimated using LD score regression (LDSR)[89] with the python package ldsc (https://github.com/bulik/ldsc). All SNP heritability estimates are reported on the liability scale assuming a population prevalence of alcohol dependence of 15.9% in alcohol-exposed individuals of European ancestry[110]. The heritability estimates were converted to the liability scale using the standard correction factor

$$\frac{K(1-K)}{z^2}\frac{K(1-K)}{P(1-P)}$$

Where K is the population prevalence, P is the in-sample prevalence, and z is the density of the normal distribution at the Kth quantile[114].

For EU cohorts, LDSR was performed using pre-computed LD scores based on 1000 Genomes Project reference data[69] on individuals of European ancestry (available for download at https://data.broadinstitute.org/alkesgroup/LDSCORE/). Evaluation of the intercept in the meta-analysis of unrelated EU individuals (10,206 cases, 28,480 controls) suggests modest inflation (intercept = 1.018, one-sided p=2.25e-3) though polygenic signal remains the primary source of deviation from the null hypothesis genome-wide (LDSR confounding ratio = 0.298). Partitioning

heritability using functional categories and selection-related metrics[115,116] does not meaningfully decrease the intercept (intercept = 1.015, one-sided p=.033, ratio = .256), suggesting that the intercept is not primarily inflated due to LDSR model misspecification. The inflation also does not appear to be due to cryptic relatedness in the summary statistic cohorts, since LDSR of the results from the meta-analysis of unrelated genotypes samples shows nominally higher inflation (intercept = 1.023, p=7.74e-5). The estimate of SNP heritability is similar robust, with generally consistent estimates from univariate LDSR of the meta-analysis for unrelated EU individuals (h_g^2=.090, 8.02e-7), partitioned LDSR for those results (h_g^2=.119, p=7.69e-5), or univariate LDSR of the results for unrelated genotyped EU samples (h_g^2=.085, p=1.94e-4).

We also performed LDSR using the AA meta-analysis results. Identifying an appropriate reference sample for computing LD scores is complicated by the admixture in this population. The pattern of LD blocks genome-wide may vary widely depending on the mosaic of local ancestry tracts, and those ancestry patterns are likely to vary between individuals and between cohorts (see Supplementary Note B.2 regarding similar complications in PCA). Therefore we evaluated LDSR in the AA results with multiple reference panels built from 1000 Genomes Project reference data: European ancestry individuals, African ancestry individuals, and African ancestry individuals in the American Southwest (ASW).

LDSR suggests nominally significant genetic signal from polygenic effects, rather than other sources of confounding, in regression with LD scores from African (h_g^2=.286, p=.0168) or European (h_g^2=.116, p=.0402) ancestry individuals. Regression with ASW samples showed a similar trend but was non-significant (h_g^2=.153, p=.0597). All SNP heritability estimates are

given assuming a population prevalence of alcohol dependence of 11.1% among alcohol-exposed African American individuals[110]. Given the instability of these estimates depending on the choice of LD reference panel, as well as the clear methodological concerns of performing this analysis in an admixed ancestry cohort, we specifically do not endorse any of these point estimates of heritability. This instability also prevents further analysis of cross-ancestry genetic correlation using a method such as popcorn[117]. We do however note the general trend of significance for these estimates, suggesting a correlation between the genome-wide meta-analysis results that is consistent with the presence of true polygenic effects in the AA cohorts. The LD score regression intercept in AA also indicates that there is no evidence of inflated results from population stratification or other confounding regardless of the choice of LD reference panel (African ancestry intercept=0.9911, se=0.0057; European ancestry intercept=0.9952, se=0.0062; ASW intercept=0.9966, se=0.005). Hopefully future analyses will clarify appropriate methods for estimating SNP heritability in cohorts of admixed ancestry and increasing samples sizes for AA cohorts will allow improved precision in estimating the contribution of polygenic effects to alcohol dependence in this population to accompany the corresponding EU ancestry estimates.

12 Gene-level association testing

Gene-level association tests were performed with MAGMA[90] using FUMA[118]. Analysis was performed with default settings for 19,436 protein-coding genes with 1000 Genome Phase 3 reference data[69]. Because these tests depend on the LD structure around each gene they were performed separately in European and African ancestry cohorts. Results from GWAS of unrelated individuals only (i.e. sub-sampling within family cohorts) were used as input to ensure valid inference from the input sample sizes and P values for MAGMA. The top results from

these gene-based tests are reported in Supplementary Table S5. No genes reach Bonferroni-corrected significance (P < 2.57E-6 = 0.05 / 19,346 genes) in either ancestry.

13 Genetic correlation

Genetic correlation of common variant risk for AD with common genetic effects on other traits was estimated using LDSR[119]. Given that the vast majority of available GWAS results of other traits of interest is for European-ancestry samples, as well as the issues with LD in admixed AA samples as noted above, analysis was restricted to genetic correlation with the GWAS of unrelated EU individuals (N_{case}=10,206, $N_{control}$=28,480).

Where possible, genetic correlation with publicly available GWAS results was computed using LD Hub. Traits from LD Hub (http://ldsc.broadinstitute.org/)[120] were selected for inclusion in this analysis based on relevance to AD and the expected power for LDSR analysis of that trait (e.g. based on the reported z-score of the SNP-heritability estimate for the trait). In the interest of maximizing power for analysis of correlation with traits of interest to AD, traits were additionally filtered to avoid redundancy between traits (e.g. excluding earlier GWAS of educational attainment in favor of the most recent published results) to limit the multiple testing burden of the overall genetic correlation analysis. Ultimately, LD Hub was used to estimate genetic correlation for 24 traits: smoking initiation (i.e. ever vs. never smoked), cessation (i.e. former vs. current smoker) and cigarettes per day[121]; depressive symptoms, neuroticism, and subjective well-being[122]; cross-disorder analysis of 5 disorders from the Psychiatric Genomics Consortium[123]; schizophrenia[111]; bipolar disorder[124]; Alzheimer's disease[125]; age of first birth and number of children[126]; parents age at death[127]; coronary artery disease[128]; Type 2 Diabetes[129];

heart rate[130]; HDL cholesterol[131]; leptin[132]; serum creatinine[133]; HbA1C levels[134]; adult height[135]; body mass index[136]; and chronotype and sleep duration[137].

Genetic correlation for 21 additional traits was computed using the python package ldsc (https://github.com/bulik/ldsc). This selection of traits was evaluated using the same criteria as the list of traits selected from LD Hub. Analysis was performed with ldsc rather than LD Hub in cases where the latest available GWAS data for the trait had not been publicly released and/or included in LD Hub's repository of results at the time of this analysis. GWAS results in this category include: cannabis use initiation[138]; nicotine dependence[139]; two analyses of alcohol consumption[140,141]; AUDIT scores[142]; attention deficit/hyperactivity disorder (ADHD)[143]; major depressive disorder[113]; anorexia nervosa[144]; autism spectrum disorder[145]; obsessive-compulsive disorder[146]; educational attainment[147]; delay discounting[148]; risk-taking behavior[149]; Townsend deprivation score; liver enzymes GGT, ALT, AST, and ALP[150]; and intracranial, caudate, and putamen brain volumes[151]. The GWAS of Townsend deprivation score is from an initial phenome-wide analysis of 337,199 genotyped individuals from UK Biobank[152] (https://github.com/Nealelab/UK_Biobank_GWAS), and assesses the socioeconomic status of an individual's neighborhood at the time of participation.

For analyses using ldsc, genetic correlation was estimated using GWAS results for common HapMap3 SNPs and previously-computed LD scores from 1000 Genomes Project reference data on individuals of European ancestry (i.e. conventional "./eur_w_ld_chr/" scores). These LD scores are freely available for download from

https://data.broadinstitute.org/alkesgroup/LDSCORE/. Both the heritability and genetic correlation intercept terms in the regression were left unconstrained for all analyses.

Genetic correlation results for all 45 traits are reported in Supplementary Table S6. All correlations were tested for difference from r_g=0 and evaluated for nominal ($p < 0.05$) and Bonferroni-adjusted ($p < 1.11E-3$ for 45 traits) significance. Genetic correlation between AD and the two alcohol consumption GWAS were additionally tested for r_g<1.

We note that there is substantial known sample overlap between our meta-analysis of AD and the cohorts including in the GWAS for many of the traits in this genetic correlation analysis. As previously described[89], the intercept term of LDSR analysis of genetic correlation can be interpreted as an index of sample overlap or other correlated confounding between the two studies. Evaluation of the intercept term in the current analysis shows noteworthy covariance intercepts for many traits, especially for other analyses of psychiatric disorders in the Psychiatric Genomics Consortium (Supplementary Table S6). These results are generally consistent with the known sample overlap with those studies and highlight the importance of leaving the intercept term unconstrained in these LDSR analyses.

14 Polygenic Risk Score Prediction

In order to evaluate how well polygenic signal in the current meta-analysis generalize beyond our cohorts we analyzed polygenic risk scores (PRS) in three external cohorts: *ALSPAC*, *GS*, and *COGA AA fGWAS*. Each of these cohorts are described in Section 1.4 and 1.5 above. In addition to assessing whether PRS created from the current meta-analysis are predictive of AD-related

outcomes, we consider predictive accuracy across ancestry and controlling for the *ADH1B* functional variants or variants associated with alcohol consumption.

14.1 PRS analysis methods

ALSPAC

SNPs from the meta-analysis of AD in unrelated genotyped EU individuals (N_{case}=8,485, $N_{control}$=20,272) were clumped to identify independent loci. Summary statistics (p-values and log-transformed odds ratios) were then used to derive PRS for ALSPAC participants at 7 p-value thresholds: p=0.001, 0.01, 0.10, 0.20, 0.30, 0.40, and 0.50. These scores were tested for association with past year alcohol use disorder (AUD) symptom count and diagnosis, with 10 principal components and sex included as covariates. We report Nagelkerke's R2 differences between a model including the PRS and a model including only principal components and sex as predictors of the outcome (AUD diagnosis or symptom count).

GS

In the *GS* data, PRS were generated using PRSice[153] using the weights from the meta-analysis of AD in unrelated genotyped EU individuals (N_{case}=8,485, $N_{control}$=20,272). An independent PRS for alcohol consumption was created using the summary statistics from a previous GWAS of alcohol consumption in the UK Biobank[141]. Clumping of SNPs in GS was performed using an r^2 threshold of 0.1 and a 250 kb window. Thirteen PRS were created using p-value thresholds of 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1. PRS were scaled to have a mean of 0 and a standard deviation of 1 prior to statistical analyses, such that the β for association reported are scaled. Association analyses between PRS and CAGE score was performed in AS-

REML-R and an inverse relationship matrix created from the kinship information in GS used to control for relatedness in the sample. Age, sex and 4 ancestral MDS components were fit as covariates. The r^2 value within the mixed model was calculated as described by [154] by multiplying the PRS by its regression coefficient and dividing this by the variance of CAGE to give a coefficient of determination between 0 and 1. For the joint analysis of AD and alcohol consumption the most predictive PRS created from each of the two GWAS were entered simultaneously into the regression model to estimate their independent effects on CAGE scores.

COGA AA fGWAS

In the *COGA AA fGWAS* cohort (N=2,828), an independent sample of African ancestry, PRS were created using the summary statistics from the GWAS of unrelated genotyped AA individuals (N_{case}=2,991, $N_{control}$=2,808). PRSice-2[153] was used to generate PRS using SNPs with minor allele frequency > 0.01, missing genotype rate < 0.1 and HWE p-values > p=1E-06. Clumping was done with respect to the linkage disequilibrium (LD) pattern in the 1000 Genome Phase 3 African ancestry sample using an r^2 threshold of 0.1 and a 250 kb window. A series of scores was calculated in COGA that included SNPs meeting increasing p-value thresholds from the discovery GWAS sample ($p<0.001$, $p<0.01$, $p<0.05$, $p<0.10$, $p<0.20$, $p<0.30$, $p<0.40$, $p<0.50$). PRS were also constructed following the same protocol with the results of the GWAS of unrelated genotyped EU samples for comparison. Association of each PRS with AD in the COGA AA fGWAS was tested using logistic mixed effect models controlling for the first three principal components, birth cohort, and sex, and including family id as a random effect to account for familial clustering. PRS were scaled to have a mean of 0 and a standard deviation of 1 prior to statistical analyses. To account for the strongest genome-wide finding in AA,

rs2066702 in *ADH1B*, analyses were rerun for the most predictive threshold in the target sample (threshold of p < 0.5, associated with AD at p = 1.92e-7) while including rs2066702 genotype (coded 0/1/2) as an additional covariate.

14.2 PRS results

PRS based on the meta-analysis of AD were significantly predictive of AD outcomes in all three tested cohorts. In ALSPAC, the AD PRS nominally predicted up to 0.55% of the variance in AUD symptom count (P_T<0.10, p=0.0195) and up to 0.51% of the variance in AUD diagnosis (P_T<0.10, p=0.0169; Supplementary Figure S10A). The AD PRS was also significantly predictive of CAGE in GS (maximum R^2=0.3%, P_T<0.2, p=7.9E-6; Supplementary Figure S10B). PRS derived from the AA GWAS of AD predicted up to nearly 1.7% of the variance in DSM-IV alcohol dependence in the independent COGA AAfGWAS sample (P_T<0.5, R^2=1.65 %, p=1.92E-7; Supplementary Figure S10C). In all three cohorts the best prediction is observed when using a moderate p-value threshold, suggesting that polygenic effects on AD risk are present in variants beyond the top observed hits in *ADH1B*.

Notably, the AD PRS still yielded significant variance explained after controlling for other genetic factors. In *GS*, when both the AD and consumption PRS were entered into the same regression model the AD PRS continued to significantly predict variance in CAGE (dependence PRS: R^2=0.0029, p=1.0E-5; consumption PRS: R^2=0.0028, p=1.3E-5) with R^2 similar to the model without the alcohol consumption PRS, suggesting independent effects in the current AD PRS. In the COGA AAfGWAS, even after the inclusion of rs2066702 genotype in the model the alcohol dependence PRS remained significantly associated with alcohol dependence (p = 2.5E-

07) and accounted for 1.61% of the variance, again suggesting the current AD GWAS does capture polygenic effects beyond the rs2066702 *ADH1B* coding variant.

Lastly, the PRS results demonstrate the better predictive accuracy of ancestry-matched PRS. PRS generated from the unrelated EU discovery GWAS only predicted modest variance in alcohol dependence in the COGA AAfGWAS sample (maximum Nagelkerke R^2 of 0.37%, p=0.01; Supplementary Figure S10D), unlike the nearly 1.7% of variance explained by the PRS computed from AA GWAS results. Importantly, this better prediction is observed despite the substantially smaller discovery sample size for the AA PRS (N_{case}=2,991, $N_{control}$=2,808) compared to the EU PRS (N_{case}=8,485, $N_{control}$=20,272). This result is consistent with previous work showing poor performance of PRS when applied to ancestries differing from the GWAS discovery sample[155].

B. Supplementary Note

1 Relationship of DSM-IV alcohol dependence to other potential case definitions

Save one study (MCTFR) that defined cases using DSM-IIIR criteria, all other cases were defined using DSM-IV criteria. There are subtle distinctions between the two diagnostic classification versions. For instance, DSM-IIIR requires the endorsement of ≥3 of 9 criteria, (DSM-IV requires ≥3 of 7 in the same 12 months) of which 8 criteria are identical to the DSM-IV criteria (with withdrawal being further divided into experience of symptoms and use of substance to alleviate the symptoms) while 1 criterion (intoxication or withdrawal hindering role obligations and in hazardous situations) broadly reflects DSM-IV abuse criteria. Thus, there is a measure of criterion redundancy in DSM-IIIR but across-classification concordance has been shown to be high in adults[156,157] (e.g., kappa 0.92) and adolescents[158] (e.g., kappa 0.96). The recent DSM-5 classification system eliminates the diagnosis of dependence in favor of a severity continuum (i.e., alcohol use disorder) based on the sum of the 7 DSM-IV dependence criteria, 3 DSM-IV abuse criteria (without legal problems) and craving (0-1: unaffected; 2-3 mild; 4-5 moderate; 6-11 severe). Application of DSM-5 criteria modestly increased the prevalence of AUD potentially due to the inclusion of individuals with 2 or more criteria[159]; studies also report that those diagnosed under the DSM-5 but not meeting criteria for DSM-IV dependence may also be less severely affected[159,160]. Thus, given that a majority of our studies preceded the development of DSM-5 criteria, we opted to use DSM-IV dependence as our case definition.

2 Principal components analysis in recently admixed samples

Principal components from PCA are commonly used as covariates in GWAS to control for population structure within a sample in order to protect against inflated results due to population stratification[63]. As part of this PCA, it is important to prune markers for LD in order to avoid regions of high LD having disproportionate influence on the PCA solution[64]. This pruning also improves the correlation of PCs with geographic population structure[161].

One common diagnostic for proper LD pruning in PCA is to test the correlation of genome-wide SNPs with each computed PC (e.g. in ricopili). Strong genome-wide signal is generally consistent with genetic drift along the PC's axis of variation, while strong association with individual loci is likely to reflect either artifacts from LD among SNPs in that region or selection[62]. For instance, SNPs in the *LCT* (lactase) region of chromosome 2 will often strongly correlate with PCs reflecting northern vs. southern European ancestry.

In applying this diagnostic in PCA of the AA cohorts we observed a pattern of strong association between PCs and SNPs in broad loci (e.g. Supplementary Figure S13A in ADAA cohort). Such loci were consistently observed for association with PCs within each cohort after the first few dimensions. The locations of these loci across the genome were not consistent across cohorts, as might be expected if these loci represented signatures of selection in the African-American population. More importantly, the pattern of loci associated with each PC was not consistent when performing PCA with different random subsets of SNPs or when removing individual chromosomes from the PCA computation. More stringent LD pruning prior to PCA also did not

remove the pattern of association with PCs. Together, this suggests that the strong association of these loci with a given PC does not indicate selection along some dimension of genetic ancestry.

Instead we hypothesize that the loci reflect regions where local genetic ancestry deviates from the individual's genome-wide average ancestry. In other words, if the first 1-2 PCs in an AA cohort capture the relative overall admixture between European and African ancestry, then the next set of PCs may capture deviations from that grand mean admixture proportion at the scale of regional ancestry tracts. The inconsistent genomic location of SNPs correlated with these PCs would thus be hypothesized to reflect the relatively stochastic variation in which regions of the genome show similar enough patterns of deviation from average admixture across individuals such that they form a primary axis of genetic variation in a given cohort. This hypothesis is also consistent with the top SNPs in these PC-associated loci being strongly ancestry informative (e.g. rs1991442, lead SNP on chromosome 3 for association with PC 8 in ADAA; ancestral A allele frequency 97% in African ancestry and 44% European ancestry).

To evaluate this hypothesis, we performed local ancestry calling in the ADAA cohort following the protocol of Martin et al.[155] (available at https://github.com/armartin/ancestry_pipeline). Briefly, QCed pre-imputation genotype data for ADAA was merged with genotype data for individuals of European or African ancestry from 1000 Genomes Phase 3 (excluding Americans of African Ancestry in the Southwestern United States [ASW])[69]. The merged data was then phased using HAPI-UR[162] and local ancestry tracts were called using RFMix[163] to identify African and European ancestry haplotypes for each individual. These local ancestry tracts were then processed to estimate each individual's global (genome-wide) proportion of African and

European ancestry as well as the proportions in each chromosome. Chromosome 6 was excluded due to the computational complexity of calling local ancestry across the HLA region in the full ADAA cohort.

Supplementary Figure S13B shows the test of association between an individual's proportion of African ancestry on each chromosome and the calculated value of the 8th principal component, controlling for the individual's global (genome-wide) proportion of African ancestry is included as a covariate. Comparison to the plot of SNP associations with this PC in Supplementary Figure S13A suggests that when loci on a given chromosome are strongly associated with the PC there is a strong relationship of that PC with local African vs. European ancestry proportions on that chromosome. Similar patterns were observed for other PCs in the ADAA cohort. These findings are highly consistent with our hypothesis that these PCs directly reflect these variations in local ancestry.

If these PCs are indeed measures of deviation of local ancestry proportion from the individual's global ancestry proportion, should they still be included as covariates in the GWAS for each cohort? We note that the purpose of PC covariates is to protect against population stratification. Thus for these local ancestry PCs their value as covariates depends on whether local ancestry, beyond the global ancestry proportions, are correlated with non-genetic factors related to alcohol dependence (AD) risk or study ascertainment. This could occur for example if local ancestry patterns differentiate between AA sub-populations or if ancestry-informative markers in those regions are strongly associated with AD.

If, however, control for local ancestry is needed in the AA GWAS, that would imply a need for more than the handful of standard PC covariates. Indeed the 8th PC shown in Supplementary Figure S13A would not be included as a covariate under our current protocol. Instead, if control for local ancestry is desired it may be preferable to call local ancestry directly within each cohort for use as a covariate rather than using these later PCs as a proxy for that structure.

To evaluate whether such control is necessary in our data, we considered the impact of including or omitting these local ancestry PCs as covariates in GWAS for each AA cohort, up to the normal number of PCs for each cohort under our current analysis protocol (i.e. based on sample size). To that end, we performed GWAS of each AA cohort with either the full set of PCs or a reduced set of PCs that omits PCs after the first that showed the characteristic pattern of strong association between the PC and particular loci (i.e. see Supplementary Table S1). Each set of GWAS was then meta-analyzed following the same procedure as the primary meta-analysis, and inflation of genome-wide test statistics was compared.

We observed effectively identical inflation of GWAS results with all PCs (λ_{GC} = 0.9930) vs. GWAS excluding PCs that appear to reflect local ancestry (λ_{GC} = 0.9935). QQ plots of the GWAS results also did not show any substantive differences in the tail of the distribution (Supplementary Figure S14). Notably the GWAS with limited PCs appears well controlled for population structure in an absolute sense as well as in comparison to the full PC analysis.

On this basis, we adopted the analysis with only top PCs (i.e. PCs without localized SNP associations likely to reflect local ancestry) as the primary analysis method for the AA cohorts.

Using this approach, there was negligible evidence of inflation of GWAS results in the AA cohorts (lambda=1.007). Further work may want to evaluate whether this approach is beneficial in other AA cohorts and whether controlling directly for local ancestry provides additional benefits. For the present analysis, however, we anticipate that this procedure controlling for only top PCs is sufficient to control for population stratification in the AA cohorts.

3 Effective sample size in family-based association models

3.1 Motivation

For the genome-wide meta-analysis of AD it is necessary to define weights for the contribution of each cohort to the meta-analysis. In the general case, fixed effects meta-analysis for SNP *j* with *Z* statistics resulting from studies k=1,…,K can be given by

$$Z_j = \frac{\sum_k Z_{jk} w_{jk}}{\sum_k w_{jk}}$$

Ideally, i.e. to maximize power, the weights w_{jk} should be proportional to the inverse of the sampling variance of z_{jk}. When comparable effects sizes (e.g. odds ratios) are available for all studies then the inverse standard error of the effects size can be used ($w_{jk} = 1/SE^2_{jk}$). Alternatively, weights defined using sample size ($w_{jk} = N_{jk}$) may asymptotically equivalent to inverse variance weights when the study design and trait distribution is identical across studies[164,165].

For the current meta-analysis, however, comparable effect sizes are not universally available, and the study design is not consistent across all cohorts. As a result, weighting by simple sample size would not be optimal to maximize power in the meta-analysis. This is true even in the

absence of the family-based cohorts, since simple sample size weighting would not account for differences in the case/control ratio across cohorts[91]. Instead we define weights based on estimates of the effective sample size of each cohort, accounting for the differences in study design and case/control balance.

3.2 Defining weights

Unrelated case/control cohorts

For GWAS of unrelated cases and controls, the effective sample size is given by

$$N_{cc} = \frac{4}{\frac{1}{N_{ca}} + \frac{1}{N_{co}}}$$

Where N_{ca} and N_{co} are the number of cases and controls in the study, respectively[91,166]. The resulting value can be interpreted as the expected sample size that would be required to have the same statistical power as the observed study if equal numbers of cases and controls were included instead of the observed case/control ratio.

Intuitively, this calculation is consistent with the test of association for a given SNP being a test of the difference in allele frequency between cases and controls. In particular, it reflects the reduced power to distinguish allele frequencies when N_{ca} is small compared to N_{co} due to the large uncertainty about frequency in cases regardless of the certainty about the control allele frequency (or equivalently when N_{co} is small).

For meta-analysis, we define weights $w_{jk} = N_{cc,k}$ based on this effective sample size for GWAS of unrelated case/control cohorts. This effective sample size also serves as our baseline for defining weights for the other study designs.

Simple family-based cohorts (GEE)

To define effective sample size weights for the family-based cohorts analyzed with the GEE model we want to account for (a) the impact of relatedness on power for the GEE model and (b) the case/control balance. The goal is to define an effective sample size that is roughly comparable to the $N_{cc,k}$ defined for case/control cohorts.

It can be shown[167,168] that asymptotically the GEE model has power proportional to

$$b\sqrt{\frac{N_{fam}}{v_R}}$$

given the true effect size b, the number of family clusters N_{fam}, and the robust sampling variance v_R. Following the previous derivations[167,168] we can evaluate v_R under a simplified model with a binary exposure that occurs with probability π, with the binary outcome occurring with probability P_0 when the exposure is absent and P_1 when the exposure is present. This corresponds to the simple case where a given SNP is rare enough to have no observed homozygotes with the minor allele (making the SNP exposure binary) but is easily generalizable. Assuming that the working correlation structure **R** has been correctly specified with compound symmetry (exchangeable) correlations within family, then we can fill in for v_R

$$v_R = \frac{1}{\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}}\left[\frac{1}{\pi P_0(1-P_0)} + \frac{1}{(1-\pi)P_1(1-P_1)}\right]$$

$$= \frac{N_{fam}}{\sum_i \frac{n_i}{1+(n_i-1)\rho}} \left[\frac{1}{\pi P_0 (1-P_0)} + \frac{1}{(1-\pi) P_1 (1-P_1)} \right]$$

where **1** is a *N*x1 vector of ones, n_i is the number of individuals within each family *i*, and ρ is the within-family phenotypic correlation[167,168]. Returning to the expression for power in the GEE model, this yields

$$b \sqrt{\frac{N_{fam}}{\frac{N_{fam}}{\sum_i \frac{n_i}{1+(n_i-1)\rho}} \left[\frac{1}{\pi P_0 (1-P_0)} + \frac{1}{(1-\pi) P_1 (1-P_1)} \right]}}$$

Assuming the effect of a single SNP in small $P_0 \approx P_1$, allowing us to simplify

$$b \sqrt{\frac{N_{fam}}{\frac{N_{fam}}{\sum_i \frac{n_i}{1+(n_i-1)\rho}} \left[\frac{1}{\pi P (1-P)} + \frac{1}{(1-\pi) P (1-P)} \right]}}$$

$$= b \sqrt{\frac{1}{\frac{1}{\sum_i \frac{n_i}{1+(n_i-1)\rho}} \left[\frac{1}{\pi (1-\pi) P (1-P)} \right]}}$$

$$= b \sqrt{\pi (1-\pi)} \sqrt{P (1-P)} \sqrt{\sum_i \frac{n_i}{1+(n_i-1)\rho}}$$

We note that π is a function of the minor allele frequency for the SNP, and can be thought of as standardizing the effect size *b*. The remaining terms reflect the impact of case/control balance (i.e. *P*[1-*P*]) and the effective sample size for the related individuals. In particular, the last term simplifies to $\sqrt{N}$ if each family only contains 1 individual (i.e. $n_i = 1$) or if there is no correlation between family members, making the observations functionally independent ($\rho = 0$). At the other

extreme, if all family members are perfectly correlated (ρ = 1) then the last term reduces to the number of families. The denominator of this term is sometimes known as the design effect.

To align this with the effective sample size for unrelated case/control cohorts defined above, we note that the effective sample size function for unrelated cohorts can be rewritten as a function of *P(1-P)*.

$$N_{cc} = \frac{4}{\frac{1}{N_{ca}} + \frac{1}{N_{co}}} = \frac{4}{\frac{1}{PN} + \frac{1}{(1-P)N}} = \frac{4}{\frac{1}{P(1-P)N}} = 4P(1-P)N$$

The scaling by 4 ensures that the effective sample size equals the sample size in a balanced case/control design (i.e. *P*=0.5). Combining this scaling with the above derivation for the power for association testing in the GEE model implies a corresponding effective sample size of

$$N_{gee} = 4P(1-P)\sum_i \frac{n_i}{1+(n_i-1)\hat{\rho}}$$

where the estimated within-family $\hat{\rho}$, computed under the null model with no SNP effects, is substituted for the true ρ to enable estimation of N_{gee}. As desired, this expression for N_{gee} clearly reflects the impact of case/control balance and family structure on the effective sample size of the GEE model. We thus use $w_{jk} = N_{gee,k}$ for as meta-analysis weights for the family-based cohorts analyzed with the GEE model.

Complex family-based cohorts (Logistic mixed model)

To define an effective sample size for the logistic mixed model consistent with the above values for the GEE and case/control models, we first note that specifying v_R as a function of $\boldsymbol{1'R^{-1}1}$ in the above derivation for the GEE model allows generalization to other correlation structures **R**. Derivations by Dang et al.[169] show that the power of generalized linear mixed models (GLMMs)

indeed depends on $\boldsymbol{1'R^{-1}1}$, where $\boldsymbol{R}$ is the marginal phenotypic correlation matrix (i.e. not conditional on random effects), with remaining scaling parameters matching the existing derivation for the GEE[167].

For the mixed model fit in the current study (described in Supplementary Methods Section 5.2), it is evident that under the null hypothesis of no effect for the target SNP

$$cor(Y^*) \propto \tau \boldsymbol{K} + \frac{\pi^2}{3}\boldsymbol{I}$$

Substituting observed values and assuming that the kinship matrix $\boldsymbol{K}$ is standardized to have diag($\boldsymbol{K}$)=$\boldsymbol{1}$, we can then compute

$$\widehat{\boldsymbol{R}} = \frac{1}{\hat{\tau} + \left(\frac{\pi^2}{3}\right)}\left[\hat{\tau}\widehat{\boldsymbol{K}} + \frac{\pi^2}{3}\boldsymbol{I}\right]$$

and following the same derivation as the GEE model approximate the effective sample size for the logistic mixed model as:

$$N_{glmm} = 4P(1-P)\,\mathbf{1}'\widehat{\boldsymbol{R}}^{-1}\mathbf{1}$$

For this calculation we use GRMs computed from genome-wide data to estimate $\boldsymbol{K}$, with the observed GRM standardized to a correlation matrix. For numerical stability in inverting $\boldsymbol{R}$, estimated relationships between families and pairwise relatedness values < 0.05 in $\boldsymbol{K}$ (after standardization) were set to zero. The variance parameter τ is estimated in each cohort under the null model with no covariates. We use the resulting estimated effective sample size as weights for meta-analysis ($w_{jk} = N_{glmm,k}$).

Summary statistic cohorts

For cohorts contributing summary statistics rather than genotyped data, we choose to define weights using N_{cc} as if they were unrelated case/control samples. This is likely sub-optimal since

most of the summary statistic cohorts are tested using some form of linear mixed model (Supplementary Table S1). It is a pragmatic solution, however, since variance component estimates for these cohorts are generally unavailable to estimate effective sample sizes analogous to N_{glmm}.

3.3 Limitations

It's important to emphasize that these effective sample size estimates are somewhat heuristic and are only intended as an approximation for the purpose of weighting the relative power between the cohorts. This is especially true of N_{glmm}, where we largely rely upon analogy to a GEE-based derivation for effective sample size, and for the pragmatic use of N_{cc} for summary statistic cohorts. In addition, it may be noted that our effective sample size calculations do not account for:

- Differences in allele frequency between cohorts. Inverse standard error-based weights are expected to reflect these differences, but sample size-based weights do not. The impact of this is likely most notable in the trans-ancestral analysis, which is part of the motivation for the secondary trans-ancestral meta-analysis with more thorough modelling of effect sizes.
- Residual correlation structure captured by robust sandwich standard errors in the GEE model. Specifically, the above derivation of N_{gee} assumes that the working correlation structure is correctly specified. The use of robust sandwich SEs in the GEE model provides some protection for inference in genome-wide association when the working correlation is misspecified, but our estimate of the effective sample size does not have the same protection.

- Uncertainty in estimating ρ and τ. We use plug-in estimates for both parameters under the null model, but this does create uncertainty in our estimate of the effective sample size for each cohort.

Despite these limitations, the defined effective sample sizes appear to perform reasonably well. The estimated values appear consistent with expectations given the sample size, case/control balance, and degree of relatedness in each cohort. In addition, informal simulations suggest that N_{gee} and N_{glmm} scale as intended with test statistics across subsamples of the COGA-fam cohort under the GEE and logistic mixed models, though with some indication that the family-based cohorts are modestly under-weighted compared to case/control cohorts (data not shown).

Importantly, the choice of these effective sample size weights w_{jk} is only expected to affect the power of the meta-analysis. The meta-analysis for null SNPs (i.e. SNPs with no true association with AD) will still have the desired null distribution and Type I error rate with sub-optimal weights. Thus any minor biases in our approximations used to define w_{jk} only serve to attenuate power in the meta-analysis. Still, we anticipate our estimated effective sample sizes are a good approximation for the relative power of each study, and thus should at least approach optimal power for the genome-wide meta-analysis in the current study.

C. References

1. Nelson, E. C. *et al.* Evidence of CNIH3 involvement in opioid dependence. *Mol. Psychiatry* **21,** 608 (2015).
2. Fergusson, D. M. & Horwood, L. J. The Christchurch Health and Development Study: review of findings on child and adolescent mental health. *Aust. N. Z. J. Psychiatry* **35,** 287–296 (2001).
3. Fergusson, D. M., Boden, J. M. & Horwood, L. J. Alcohol misuse and psychosocial outcomes in young adulthood: results from a longitudinal birth cohort studied to age 30. *Drug Alcohol Depend.* **133,** 513–519 (2013).
4. Edenberg, H. J. *et al.* Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol. Clin. Exp. Res.* **34,** 840–852 (2010).
5. Bucholz, K. K. *et al.* A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J. Stud. Alcohol* **55,** 149–158 (1994).
6. Hesselbrock, M. *et al.* A validity study of the SSAGA - A comparison with the SCAN. *Addiction* **94,** 1361–1370 (1999).
7. Begleiter, H. *et al.* The collaborative study on the genetics of alcoholism. *Alcohol Res. Heal.* **19,** 228 (1995).
8. Bierut, L. J., Strickland, J. R., Thompson, J. R., Afful, S. E. & Cottler, L. B. Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend.* **95,** 14–22 (2008).
9. Bierut, L. J. *et al.* Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum. Mol. Genet.* **16,** 24–35 (2006).

10. Bierut, L. J. *et al.* A genome-wide association study of alcohol dependence. *Proc. Natl. Acad. Sci.* **107,** 5082–5087 (2010).

11. Frank, J. *et al.* Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict. Biol.* **17,** 171–180 (2012).

12. Treutlein, J. *et al.* Genome-wide association study of alcohol dependence. *Arch. Gen. Psychiatry* **66,** 773–784 (2009).

13. Costello, E. J. *et al.* Genes, environments, and developmental research: methods for a multi-site study of early substance abuse. *Twin Res. Hum. Genet.* **16,** 505–515 (2013).

14. Adkins, D. E. *et al.* Genome-wide meta-analysis of longitudinal alcohol consumption across youth and early adulthood. *Twin Res. Hum. Genet.* **18,** 335–347 (2015).

15. Hartman, C. A. *et al.* Item response theory analysis of DSM-IV cannabis abuse and dependence criteria in adolescents. *J. Am. Acad. Child Adolesc. Psychiatry* **47,** 165–173 (2008).

16. Petrill, S. A., Plomin, R., DeFries, J. C. & Hewitt, J. K. *Nature, nurture, and the transition to early adolescence*. (Oxford University Press, 2003).

17. Rhea, S.-A., Gross, A. A., Haberstick, B. C. & Corley, R. P. Colorado twin registry. *Twin Res. Hum. Genet.* **9,** 941–949 (2006).

18. Stallings, M. C. *et al.* A genome-wide search for quantitative trait loci that influence antisocial drug dependence in adolescence. *Arch. Gen. Psychiatry* **62,** 1042–1051 (2005).

19. Derringer, J. *et al.* Genome-wide association study of behavioral disinhibition in a selected adolescent sample. *Behav. Genet.* **45,** 375–381 (2015).

20. Dick, D. M. *et al.* Spit for Science: launching a longitudinal study of genetic and environmental influences on substance use and emotional health at a large US university.

Front. Genet. **5,** 47 (2014).

21. Karpyak, V. M. *et al.* Genetic markers associated with abstinence length in alcohol-dependent subjects treated with acamprosate. *Transl. Psychiatry* **4,** e453 (2014).
22. Karpyak, V. M. *et al.* Gender-specific effects of comorbid depression and anxiety on the propensity to drink in negative emotional states. *Addiction* **111,** 1366–1375 (2016).
23. Olson, J. E. *et al.* The Mayo Clinic Biobank: a building block for individualized medicine. in *Mayo Clinic Proceedings* **88,** 952–962 (Elsevier, 2013).
24. Gillespie, N. A. *et al.* The Brisbane Longitudinal Twin Study: pathways to Cannabis Use, Abuse, and Dependence project—current status, preliminary results, and future directions. *Twin Res. Hum. Genet.* **16,** 21–33 (2013).
25. Couvy-Duchesne, B., Davenport, T. A., Martin, N. G., Wright, M. J. & Hickie, I. B. Validation and psychometric properties of the Somatic and Psychological HEalth REport (SPHERE) in a young Australian-based population sample using non-parametric item response theory. *BMC Psychiatry* **17,** 279 (2017).
26. Medland, S. E. *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am. J. Hum. Genet.* **85,** 750–755 (2009).
27. Maes, H. H., Silberg, J. L., Neale, M. C. & Eaves, L. J. Genetic and cultural transmission of antisocial behavior: an extended twin parent model. *Twin Res. Hum. Genet.* **10,** 136–150 (2007).
28. Vrieze, S. I. *et al.* Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biol. Psychiatry* **75,** 783–789 (2014).
29. Miller, M. B. *et al.* The Minnesota Center for Twin and Family Research Genome-Wide Association Study. *Twin Res. Hum. Genet.* **15,** 767–774 (2012).

30. Maher, B. S. *et al.* The AVPR1A gene and substance use disorders: association, replication, and functional evidence. *Biol. Psychiatry* **70,** 519–527 (2011).

31. Tarter, R. E. & Vanyukov, M. M. Introduction: Theoretical and operational framework for research into the etiology of substance use disorders. *J. Child Adolesc. Subst. Abuse* **10,** 1–12 (2001).

32. Lichtenstein, P., Floderus, B., Svartengren, M., Svedberg, P. & Pedersen, N. L. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J. Intern. Med.* **252,** 184–205 (2002).

33. Magnusson, P. K. E. *et al.* The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res. Hum. Genet.* **16,** 317–329 (2013).

34. Pierucci-Lagha, A. *et al.* Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend.* **91,** 85–90 (2007).

35. Gelernter, J. *et al.* Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. *Biol. Psychiatry* **76,** 66–74 (2014).

36. Gelernter, J. *et al.* Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol. Psychiatry* **19,** 717–723 (2014).

37. Gelernter, J. *et al.* Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol. Psychiatry* **19,** 41–49 (2014).

38. Sherva, R. *et al.* Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. *JAMA Psychiatry* **73,** 472–480 (2016).

39. Wang, J.-C. *et al.* A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53. *Mol. Psychiatry* **18,** 1218–1224 (2013).

40. Wetherill, L. *et al.* Association of substance dependence phenotypes in the COGA sample. *Addict. Biol.* **20,** 617–627 (2015).

41. Heath, A. C. *et al.* A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biol. Psychiatry* **70,** 513–518 (2011).

42. Prescott, C. A. *et al.* The Irish Affected Sib Pair Study of Alcohol Dependence: study methodology and validation of diagnosis by interview and family history. *Alcohol. Clin. Exp. Res.* **29,** 417–429 (2005).

43. Penninx, B. W. J. H. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* **17,** 121–140 (2008).

44. Boomsma, D. I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res. Hum. Genet.* **9,** 849–857 (2006).

45. Willemsen, G. *et al.* The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Res. Hum. Genet.* **16,** 271–281 (2013).

46. Kaprio, J. The Finnish twin cohort study: an update. *Twin Res. Hum. Genet.* **16,** 157–162 (2013).

47. Loukola, A. *et al.* Genome-wide association study on detailed profiles of smoking behavior and nicotine dependence in a twin sample. *Mol. Psychiatry* **19,** 615–624 (2014).

48. Mullan Harris, K., Lee, H. & DeLeone, F. Y. Marriage and health in the transition to adulthood: Evidence for African Americans in the Add Health Study. *J. Fam. Issues* **31,** 1106–1143 (2010).

49. Harris, K. M., Halpern, C. T., Haberstick, B. C. & Smolen, A. The national longitudinal

study of adolescent health (Add Health) sibling pairs data. *Twin Res. Hum. Genet.* **16,** 391–398 (2013).

50. Eriksson, J. G., Osmond, C., Kajantie, E., Forsén, T. J. & Barker, D. J. P. Patterns of growth among children who later develop Type 2 Diabetes or its risk factors. *Diabetologia* **49,** 2853–2858 (2006).

51. Borodulin, K. *et al.* Forty-year trends in cardiovascular risk factors in Finland. *Eur. J. Public Health* **25,** 539–546 (2015).

52. Meyers, J. L. *et al.* An endophenotype approach to the genetics of alcohol dependence: A genome wide association study of fast beta EEG in families of African ancestry. *Mol. Psychiatry* **22,** 1767–1775 (2017).

53. Boyd, A. *et al.* Cohort profile: The 'Children of the 90s'-The index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42,** 111–127 (2013).

54. Fraser, A. *et al.* Cohort profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42,** 97–110 (2013).

55. Harris, P. a. *et al.* Research Electronic Data Capture (REDCap) - A metadata driven methodology and workflow process for providing translational research informatict support. *J. Biomed. Inform.* **42,** 377–81 (2009).

56. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Publishing, 2013).

57. Smith, B. H. *et al.* Cohort profile: Generation Scotland: Scottish family health study (GS: SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42,** 689–700 (2013).

58. Navrady, L. B. *et al.* Cohort profile: Stratifying Resilience and Depression Longitudinally

(STRADL): A questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS: SFHS). *Int. J. Epidemiol.* **47,** 13–14g (2018).

59. Ewing, J. A. Detecting Alcoholism: The CAGE Questionnaire. *JAMA J. Am. Med. Assoc.* **25,** 1905–1907 (1984).

60. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4,** (2015).

61. Mbarek, H. *et al.* Genome-Wide Significance for PCLO as a Gene for Major Depressive Disorder. *Twin Res. Hum. Genet.* 1–4 (2017).

62. Galinsky, K. J. *et al.* Fast Principal-Component Analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98,** 456–472 (2016).

63. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).

64. Price, A. L. *et al.* Long-Range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83,** 132–135 (2008).

65. Staples, J., Nickerson, D. A. & Below, J. E. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet. Epidemiol.* **37,** 136–141 (2013).

66. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet. Epidemiol.* **39,** 276–293 (2015).

67. Thornton, T. *et al.* Estimating kinship in admixed populations. *Am. J. Hum. Genet.* **91,** 122–138 (2012).

68. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).

69. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

70. Delaneau, O. *et al.* Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5,** (2014).

71. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34,** D590–D598 (2006).

72. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods* **10,** 5–6 (2013).

73. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5,** e1000529 (2009).

74. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 Genes, Genomes, Genet.* **1,** 457–470 (2011).

75. O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* **10,** e1004234 (2014).

76. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48,** 1284–1287 (2016).

77. Price, A. L. *et al.* Effects of cis and trans genetic ancestry on gene expression in African Americans. *PLoS Genet.* **4,** e1000294 (2008).

78. Duncan, L. E. *et al.* Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol. Psychiatry* (2018). doi:10.1038/mp.2017.77

79. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS*

Genet. **2,** 2074–2093 (2006).

80. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized Llnear models. *Biometrika* **73,** 13 (1986).

81. Højsgaard, S., Halekoh, U. & Yan, J. The R Package geepack for generalized estimating equations. *J. Stat. Softw.* **15,** 1–11 (2006).

82. Breslow, N. E. & Clayton, D. G. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* **88,** 9–25 (1993).

83. Chen, H. *et al.* Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am. J. Hum. Genet.* **98,** 653–666 (2016).

84. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* **46,** 100–106 (2014).

85. Chen, M. H. *et al.* A comparison of strategies for analyzing dichotomous outcomes in genome-wide association studies with general pedigrees. *Genet. Epidemiol.* **35,** 650–657 (2011).

86. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11,** 499–511 (2010).

87. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44,** 821–824 (2012).

88. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

89. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).

90. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11,** (2015).

91. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

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

93. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35,** 809–822 (2011).

94. Wang, X. *et al.* Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **22,** 2303–2311 (2013).

95. Cochran, W. G. The Combination of Estimates from Different Experiments. *Biometrics* **10,** 101 (1954).

96. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–585 (2013).

97. Machiela, M. J. & Chanock, S. J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31,** 3555–3557 (2015).

98. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45,** D896–D901 (2017).

99. Schmitt, A. D. *et al.* A compendium of chromatin contact caps reveals spatially active regions in the human genome. *Cell Rep.* **17,** 2012–2059 (2016).

100. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24,** 974–988 (2014).

101. Martin, J. S. *et al.* HUGIn: Hi-C unifying genomic interrogator. *Bioinformatics* **33,** 3793–3795 (2017).

102. Fritz, D. & Stefanovic, B. RNA-binding protein RBMS3 is expressed in activated hepatic stellate cells and liver fibrosis and increases expression of transcription factor Prx1. *J. Mol. Bol.* **371,** 585–595 (2007).

103. Li, Y. *et al.* Downregulation of RBMS3 is associated with poor prognosis in esophageal squamous cell carcinoma. *Cancer Res.* **71,** 6106–6115 (2011).

104. Chen, J. *et al.* RBMS3 at 3p24 inhibits nasopharyngeal carcinoma development via inhibiting cell proliferation, angiogenesis, and inducing apoptosis. *PLoS One* **7,** e44636 (2012).

105. Liang, Y.-N. *et al.* RBMS3 is a tumor suppressor gene that acts as a favorable prognostic marker in lung squamous cell carcinoma. *Med. Oncol.* **32,** 30 (2015).

106. Liu, P. *et al.* Role of glutamate decarboxylase-like protein 1 (GADL1) in taurine biosynthesis. *J. Biol. Chem.* **287,** 40898–40906 (2012).

107. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46,** 543–550 (2014).

108. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* **38,** 209–213 (2006).

109. Zhang, J. & Yu, K. F. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA J. Am. Med. Assoc.* **280,** 1690–1 (1998).

110. 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* **64,** 830–842 (2007).

111. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

112. Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45,** 501–512 (2013).

113. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50,** 668–681 (2018).

114. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88,** 294–305 (2011).

115. Finucane, H. K. *et al.* Partitioning heritability by functional category using GWAS summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).

116. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49,** 1421–1427 (2017).

117. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99,** 76–88 (2016).

118. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8,** 1826 (2017).

119. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).

120. Zheng, J. *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* **33,** 272–279 (2017).

121. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–7 (2010).

122. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48,** 624–633 (2016).

123. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).

124. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43,** 977–83 (2011).

125. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–1458 (2013).

126. Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet* **48,** 1462–1472 (2016).

127. Pilling, L. C. *et al.* Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging (Albany. NY).* **8,** 547–560 (2016).

128. the CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47,** 1121–1130 (2015).

129. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44,** 981–990 (2012).

130. Den Hoed, M. *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat. Genet.* **45,** 621–631 (2013).

131. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466,** 707–713 (2010).

132. Kilpeläinen, T. O. *et al.* Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. *Nat. Commun.* **7,** (2016).

133. Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7,** 10023 (2016).

134. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. *Diabetes* **59,** 3229–3239 (2010).

135. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).

136. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42,** 937–948 (2010).

137. Jones, S. E. *et al.* Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS Genet.* **12,** (2016).

138. Stringer, S. *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* **6,** e769 (2016).

139. Hancock, D. B. *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

140. Schumann, G. *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* **113,** 14372–14377 (2016).

141. Clarke, T.-K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112,117). *Molecular Psychiatry* (2017). doi:10.1038/mp.2017.153

142. Sanchez-Roige, S. *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

143. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv* 145581 (2017). doi:10.1101/145581

144. Duncan, L. *et al.* Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **174,** 850–858 (2017).

145. Anney, R. J. L. *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* **8,** (2017).

146. Arnold, P. D. *et al.* Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23,** 1181–1188 (2018).

147. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50,** 1112–1121 (2018).

148. Sanchez-Roige, S. *et al.* Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. *Nat. Neurosci.* **21,** 16–18 (2018).

149. Strawbridge, R. J. *et al.* Genome-wide analysis of self-reported risk-taking behaviour and

cross-disorder genetic correlations in the UK Biobank cohort. *Transl. Psychiatry* **8,** 39 (2018).

150. Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet.* **43,** 1131–1138 (2011).

151. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520,** 224–229 (2015).

152. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* (2017). doi:10.1101/166298

153. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31,** 1466–1468 (2015).

154. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4,** 133–142 (2013).

155. Martin, A. R. *et al.* Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100,** 635–649 (2017).

156. Schuckit, M. A. *et al.* A comparison of DSM-III-R, DSM-IV and ICD-10 substance use disorders diagnoses in 1922 men and women subjects in the COGA study. *Addiction* **89,** 1629–1638 (1994).

157. Rounsaville, B. J., Bryant, K., Babor, T., Kranzler, H. & Kadden, R. Cross system agreement for substance use disorders: DSM-III-R, DSM-IV and ICD-10. *Addiction* **88,** 337–348 (1993).

158. Pollock, N. K., Martin, C. S. & Langenbucher, J. W. Diagnostic concordance of DSM-III, DSM-III-R, DSM-IV and ICD-10 alcohol diagnoses in adolescents. *J. Stud. Alcohol* **61,** 439–446 (2000).

159. Agrawal, A., Heath, A. C. & Lynskey, M. T. DSM-IV to DSM-5: The impact of proposed revisions on diagnosis of alcohol use disorders. *Addiction* **106,** 1935–1943 (2011).

160. Takahashi, T. *et al.* Comparison of DSM-IV and DSM-5 criteria for alcohol use disorders in VA primary care patients with frequent heavy drinking enrolled in a trial. *Addict. Sci. Clin. Pract.* **12,** 17 (2017).

161. Abdellaoui, A. *et al.* Population structure, migration, and diversifying selection in the Netherlands. *Eur. J. Hum. Genet.* **21,** 1277–1285 (2013).

162. Williams, A. L., Patterson, N., Glessner, J., Hakonarson, H. & Reich, D. Phasing of many thousands of genotyped samples. *Am. J. Hum. Genet.* **91,** 238–251 (2012).

163. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93,** 278–288 (2013).

164. Lee, C. H., Cook, S., Lee, J. S. & Han, B. Comparison of two meta-analysis methods: Inverse-variance-weighted average and weighted sum of Z-scores. *Genomics Inf.* **14,** 173–180 (2016).

165. Hedges, L. V. & Olkin, I. *Statistical Methods for Meta-Analysis*. (Academic Press, 1985).

166. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using family history of disease. *Nat. Genet.* **49,** 325–331 (2017).

167. Pan, W. Sample size and power calculations with correlated binary data. *Control. Clin. Trials* **22,** 211–227 (2001).

168. Shih, W. J. Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations. *Biometrical J.* **39,** 899–908 (1997).

169. Dang, Q., Mazumdar, S. & Houck, P. R. Sample size and power calculations based on generalized linear mixed models with correlated binary outcomes. *Comput. Methods Programs Biomed.* **91,** 122–127 (2008).

170. Pruim, R. J. *et al.* LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–2337 (2010).