A large-scale genome-wide association study meta-analysis of cannabis use disorder

Emma C. Johnson[1]*,PhD; Ditte Demontis[2,3,4]*,PhD; Thorgeir E. Thorgeirsson[5]*,PhD; Raymond K. Walters[6,7],PhD; Renato Polimanti[8,9],PhD; Alexander S. Hatoum[1],PhD; Sandra Sanchez-Roige[10,11],PhD; Sarah E. Paul[12],BA; Frank R. Wendt[8,9],PhD; Toni-Kim Clarke[13],PhD; Dongbing Lai[14],PhD; Gunnar W. Reginsson[5],PhD; Hang Zhou[8,9],PhD; June He[1],MSc; David A.A. Baranger[15],PhD; Daniel F. Gudbjartsson[16,17],PhD; Robbee Wedow[6],PhD; Daniel E. Adkins[18],PhD; Amy E. Adkins[19,20],PhD; Jeffry Alexander[21],PhD; Silviu-Alin Bacanu[21],PhD; Tim B. Bigdeli[22],PhD; Prof Joseph Boden[23],PhD; Prof Sandra A. Brown[10],PhD; Prof Kathleen K. Bucholz[1],PhD; Jonas Bybjerg-Grauholm[3,24],MSc; Robin P. Corley[25],PhD; Prof E. Jane Costello[26],PhD; Prof Louisa Degenhardt[27],PhD; Prof Danielle M. Dick[19,28],PhD; Benjamin W. Domingue[29],PhD; Louis Fox[1],BS; Prof Alison M. Goate[30],D.Phil; Scott D. Gordon[31],PhD; Laura M. Hack[32],MD; Dana B. Hancock[33],PhD; Sarah M. Hartz[1],MD; Prof Ian B. Hickie[34],MD; David M. Hougaard[3,24],MD; Prof Kenneth Krauter[35],PhD; Penelope A. Lind[31],PhD; Jeanette N. McClintick[36],PhD; Matthew B. McQueen[37],ScD; Jacquelyn L. Meyers[22],PhD; Prof Grant W. Montgomery[38],PhD; Prof Ole Mors[3,39],PhD; Prof Preben B Mortensen[3,40,41],MD; Prof Merete Nordentoft[3,42],MD; John F. Pearson[43,61],PhD; Roseann E. Peterson[44],PhD; Maureen D. Reynolds[45],PhD; Prof John P. Rice[1],PhD; Valgerdur Runarsdottir[46],MD; Nancy L. Saccone[47,48],PhD; Richard Sherva[49],PhD; Judy L. Silberg[28,44],PhD; Prof Ralph E. Tarter[45],PhD; Thorarinn Tyrfingsson[46],MD; Prof Tamara L. Wall[10],PhD; Bradley T. Webb[21],PhD; Prof Thomas Werge[3,50,51,52],PhD; Leah Wetherill[14],PhD; Margaret J. Wright[53],PhD; Stephanie Zellers[54],MA; Mark J. Adams[13],PhD; Prof Laura J. Bierut[1],MD; Prof Jason D. Boardman[55],PhD; Prof William E. Copeland[56],PhD; Prof Lindsay A. Farrer[49],PhD; Prof Tatiana M. Foroud[14],PhD; Nathan A. Gillespie[44],PhD; Prof Richard A. Grucza[1],PhD; Prof Kathleen Mullan Harris[57,58],PhD; Prof Andrew C. Heath[1],D.Phil; Prof Victor Hesselbrock[59],PhD; Prof John K. Hewitt[25],PhD; Prof Christian J. Hopfer[60],MD; Prof John Horwood[23],MSc; Prof William G. Iacono[54],PhD; Prof Eric O. Johnson[33],PhD; Prof Kenneth S. Kendler[44],MD; Prof Martin A. Kennedy[61],PhD; Prof Henry R. Kranzler[62,63],MD; Prof Pamela A.F. Madden[1],PhD; Hermine H. Maes[28,44],PhD; Prof Brion S. Maher[64],PhD; Prof Nicholas G. Martin[31],PhD; Prof Matthew McGue[54],PhD; Prof Andrew M. McIntosh[13],MD; Prof Sarah E. Medland[31],PhD; Prof Elliot C. Nelson[1],MD; Prof Bernice Porjesz[22],PhD; Prof Brien P. Riley[21],PhD; Prof Michael C. Stallings[25],PhD; Prof Michael M. Vanyukov[45],PhD; Scott Vrieze[54],PhD; the Psychiatric Genomics Consortium Substance Use Disorders Workgroup, Lea K. Davis[11,65,66],PhD; Ryan Bogdan[11],PhD; Prof Joel

Gelernter[8,9,67,68#],MD; Prof Howard J. Edenberg[13,35#],PhD; Prof Kari Stefansson[69,70#],MD; Prof Anders D. Børglum[2,3,4#],MD; Prof Arpana Agrawal[1#],PhD

*These authors contributed equally.

#These authors jointly supervised this work.

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA
[2]Department of Biomedicine - Human Genetics and Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark
[3]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark
[4]Center for Genomics and Personalized Medicine, Aarhus, Denmark
[5]CNS Department, deCODE Genetics/Amgen, Reykjavik, Iceland
[6]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[7]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[8]Department of Psychiatry, Division of Human Genetics, Yale School of Medicine, New Haven, CT, USA
[9]Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA
[10]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA
[11]Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
[12]Department of Psychological and Brain Sciences, Washington University in Saint Louis, St. Louis, MO, USA
[13]Department of Psychiatry, University of Edinburgh, Edinburgh, United Kingdom
[14]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
[15]Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, PA, USA
[16]Statistics Department, deCODE Genetics/Amgen, Reykjavik, Iceland
[17]School of Engineering and Natural Sciences, Iceland University, Reykjavik, Iceland
[18]Department of Psychiatry, University of Utah, Salt Lake City, UT, USA
[19]Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA
[20]College Behavioral and Emotional Health Institute, Virginia Commonwealth University, Richmond, VA, USA
[21]Virginia Commonwealth University Alcohol Research Center, Virginia Commonwealth University, Richmond, VA, USA
[22]Department of Psychiatry and Behavioral Sciences, State University of New York Downstate Medical Center, Brooklyn, NY, USA
[23]Department of Psychological Medicine, University of Otago, Christchurch, New Zealand
[24]Department for Congenital Disorders, Center for Neonatal Screening, Statens Serum Institut, Copenhagen, Denmark
[25]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA
[26]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA
[27]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, New South Wales, Australia

[28]Department of Human & Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA
[29]Stanford University Graduate School of Education, Stanford University, Stanford, CA, USA
[30]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[31]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
[32]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA
[33]Center for Omics Discovery and Epidemiology, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC, USA
[34]Brain and Mind Centre, University of Sydney, Sydney, Australia
[35]Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO, USA
[36]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA
[37]Department of Integrative Physiology, University of Colorado Boulder, Boulder, CO, USA
[38]Institute for Molecular Bioscience, University of Queensland, Queensland, Australia
[39]Psychosis Research Unit, Aarhus University Hospital, Aarhus, Denmark
[40]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark
[41]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark
[42]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark
[43]Biostatistics and Computational Biology Unit, University of Otago, Christchurch, New Zealand
[44]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA
[45]School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA
[46]SAA - National Center of Addiction Medicine, Vogur Hospital, Reykjavik, Iceland
[47]Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA
[48]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA
[49]Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA
[50]Institute of Biological Psychiatry, Mental Health Services, Copenhagen University Hospital, Copenhagen, Denmark
[51]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[52]Center for GeoGenetics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
[53]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia
[54]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA
[55]Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO, USA
[56]Department of Psychiatry, University of Vermont Medical Center, Burlington, VT, USA
[57]Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[58]The Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[59]Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT, USA
[60]Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA
[61]Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand
[62]Center for Studies of Addiction, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[63]VISN 4 MIRECC, Crescenz VAMC, Philadelphia, PA, USA
[64]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[65]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

[66]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
[67]Department of Genetics, Yale School of Medicine, New Haven, CT, USA
[68]Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA
[69]deCODE Genetics/Amgen, Reykjavik, Iceland
[70]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland

Abstract word count: 849 including funding; 282 without funding

Manuscript word count: 5,134

Tables: 2

Figures: 4

Key words: GWAS; cannabis use disorder; substance use; white matter volume

ABSTRACT

Background: Variation in liability to cannabis use disorder (CUD) has a strong genetic component (estimated twin and family heritability ~ 50-70%) and is associated with negative outcomes, including increased risk of psychopathology. The aim of the current study was to conduct a well-powered GWAS to identify novel genetic variants associated with CUD.

Methods: We conducted the largest GWAS meta-analysis of CUD to date (N_{case} = 20,916; $N_{control}$ = 363,116), and used polygenic risk score approaches to examine associations between CUD and relevant traits in independent samples: cannabis use frequency in the UK Biobank, a variety of health codes in the BioVU sample, and brain volume in adolescents.

Outcomes: We identified two genome-wide significant loci: a novel chromosome 7 locus (*FOXP2;* lead SNP rs7783012, OR = 1.11, *p* = 1.84e-09), and the previously identified chromosome 8 locus (near *CHRNA2* and *EPHX2;* lead SNP rs4732724, OR = 0.89, *p* = 6.46e-09). A phenome-wide analysis of electronic health codes in an independent sample (N = 66,915) revealed genetic overlap between CUD and mental health, respiratory illness, and infectious conditions. Although they were genetically correlated (r_g = 0.50, p =1.5e-21), CUD and cannabis use showed opposite directions of genetic correlation with education, body mass index, and age at first birth, suggesting at least partially different genetic underpinnings of cannabis use *versus* use disorder. Further, polygenic scores for CUD, but not cannabis use, were associated with less total white matter volume in cannabis-naïve children from the Adolescent Brain Cognitive Development Study (N = 4,539).

Interpretation: Collectively, these findings reinforce the conclusion that CUD is a psychiatric disorder with greater shared liability to psychopathology and potential early brain volume differences, and distinctions from genetic likelihood of cannabis use.
Funding: MH109532 (AA, HJE, JG); F32AA027435 (ECJ); 1U01MH109514-01 (ADB); 5T32DA007261 (AH); T32MH018951 (DAAB); Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to ADB); the European Commission, Horizon 2020, grant number 667302 (ADB); R01-DA034076 (deCODE, TET); R01HD060726 (BWD, KMH, JDB, MBM); Health Research Council of New Zealand 16/600 (JB, JH, MAK, JFP); Health Research Council of New Zealand 11/792 (JB, JH); AA027827, AG061162 (RB);U24DA041147 (SAB);R01HD093651, R01MH117559, P30DA023026 (WEC); DA011015, DA038504 (RPC);K02 AA018755 (DMD);P01-HD031921, R01-HD073342 (KMH);R01 DA018673, DA025109, DA024413, DA016977 (HHM); Wellcome Trust (104036/Z/14/Z, 216767/Z/19/Z), UKRI MRC (MC_PC_17209, MR/S035818/1) (AMM); K01MH113848, The Brain & Behavior Research Foundation NARSAD grant 28632 (REP); R21 DA047527, R21 DC018098 (RP); SAMHSA Grant # 1H79TI081668 (MDR); R01DA026911, R21DA046791 (NLS); R01HD050735, U54EB020403 (subaward), NHMRC Australia (496682; 1009064) (MJW); Tobacco-Related Disease Research Program of the University of California Grant Number T29KT0526,Families for Borderline Personality Disorder Research (Beth and Rob Elliott) 2018 NARSAD Young Investigator Grant #27676 (SSR);
Comorbidity and Trauma Study (CATS): R01 DA17305 & contract N01-HG-65403).
Center for Education and Drug Abuse Research (CEDAR): P50 DA005605.
The Christchurch Health and Development Study (CHDS): Health Research Council of New Zealand, the National Child Health Research Foundation (Cure Kids), the Canterbury Medical Research Foundation, the New Zealand Lottery Grants Board, the University of Otago, The Carney Centre for Pharmacogenomics, the James Hume Bequest Fund, MH077874, and R01DA024413.
The Collaborative Study on the Genetics of Alcoholism (COGA): U10AA008401, U01HG004438), and HHSN268200782096C.
Study of Addiction: Genetics and Environment (SAGE): U01 HG004422;
The Collaborative Genetic Study of Nicotine Dependence(COGEND); P01 CA089392, and the Family Study of Cocaine Dependence (FSCD); R01 DA013423, R01 DA019963 & U01HG004438, HHSN268200782096C.

Great Smoky Mountains Study (GSMS): U01DA024413, R01DA11301, R01MH063970, R01MH063671, R01MH048085, K01MH093731 and K23MH080230, NARSAD, and the William T. Grant Foundation.
Center on Antisocial Drug Dependence (CADD): DA011015, DA012845, DA021913, DA021905, DA032555, and DA035804.
Alcohol Dependence in African Americans (ADAA): R01 AA017444.
Brisbane Longitudinal Twin Study (BLTS): R00DA023549, Australian Research Council (DP0343921, DP0664638, 464914, 619667, FT110100548).
Gene-Environment-Development Initiative -GEDI – Virginia Commonwealth University (VTSABD): U01DA024413, R01DA025109, R01MH045268, R01MH055557, R01MH068521, and the Virginia Tobacco Settlement Foundation grant 8520012.
Minnesota Center for Twin and Family Research (MCTFR): DA005147, DA013240, DA024417, DA036216, AA009367, MH066140, DA042755, DA037904, HG008983, DA044283.
Substance Use Disorder Liability: Candidate System Genes: R01 DA019157 and P50 DA005605.
Yale-Penn: RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535 and the VISN 1 and VISN 4 Mental Illness Research, Education, and Clinical Centers of the U.S. Department of Veterans Affairs.
Australian Alcohol and Nicotine studies (OZ-ALC-NAG): AA07535,AA07728, AA13320, AA13321, AA14041, AA11998, AA17688,DA012854, and DA019951, by Grants from the Australian National Health and Medical Research Council (241944, 339462, 389927,389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485,and 552498); by Grants from the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, and DP0343921); and by the 5th Framework Programme (FP-5) GenomEUtwin Project (QLG2-CT-2002-01254) & AA13320.
Irish Affected Sib-Pair Study of Alcohol Dependence (IASPSAD): P20-AA-017828 and P50-AA-022537, R01-AA-011408, R01-MH-083094 and Wellcome Trust Case Control Consortium 2 grant WTCCC-084710.
Add Health: P01-HD31921. ABCD: U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025;
iPSYCH: grants from the Lundbeck Foundation R165-2013-15320, R102-A9118, R155-2014-1724 and R248-2017-2003;

deCODE: R01-DA034076
bioVU: NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at https://victr.vumc.org/biovu-funding/.

INTRODUCTION

Approximately 50-70% of cannabis use disorder (CUD) liability is attributable to genetic factors[1,2]. Three genome-wide association studies (GWAS) of CUD[3–5] have identified genome-wide significant variants, but limited sample sizes (largest N = 51,372) and heterogeneity among samples have contributed to a paucity of replicable findings: only one locus, tagged by a *cis*-eQTL for *CHRNA2* that encodes a nicotinic acetylcholine receptor, has been identified[4]. As in other polygenic disorders, particularly substance use disorders, increasing the sample size will reveal additional reliable genetic associations for CUD.

Cannabis use is common, and a recent GWAS of lifetime cannabis use (N = 184,765; N_{cases} = 43,380) identified eight genome-wide significant loci, and 35 significant genes[6]. Twin studies suggest high genetic correlations between earlier stages of cannabis experimentation and later CUD[7,8]; however, casual cannabis use is also influenced by a variety of socio-environmental influences and age-period-cohort effects, while progression to CUD likely accrues risk related to other psychopathologies. Furthermore, comparisons of alcohol consumption and alcohol use disorder support partially distinct genetic etiologies with respect to both associated variants and genetic relationships with other psychiatric disorders and traits[9–11]. Thus, genomic liability for subdiagnostic substance use measured in population-based cohorts may be partially distinct from that associated with disordered use. Whether the genetic architectures of cannabis use and CUD show divergence like alcohol has not yet been investigated.

METHODS

Samples

Twenty samples were included: the Psychiatric Genomics Consortium (PGC) Substance Use Disorders working group (18 samples; European ancestry, N_{CUD} = 8,277, $N_{control}$ = 23,497; African ancestry N_{CUD} = 3,848, $N_{control}$ = 5,897), the iPSYCH cohort[12] (all Europeans,

N_{CUD}=2,758, $N_{control}$=53,326), and the deCODE sample (all Europeans, N_{CUD}=6,033, $N_{control}$= 280,396) (Table 1; Supplementary Information). First, the summary statistics from the 18 PGC samples were meta-analyzed together, followed by a meta-analysis that included results from the GWAS of iPSYCH and deCODE samples (between-sample genetic correlations r_g = 0.66 - 0.70).

This study was approved by the institutional review board (IRB) at Washington University School of Medicine and was conducted in accordance with all relevant ethical regulations. Investigators for each contributing study obtained informed consent from their participants and received ethics approvals from their respective review boards in accordance with applicable regulations. Personal identifiers associated with phenotypic information and samples were encrypted using a third-party encryption system[13].

Measures

CUD phenotyping

PGC cases met criteria for a lifetime diagnosis of DSM-IV (or DSM-III-R) cannabis abuse *or* dependence[14] derived from clinician ratings or semi-structured interviews[9]. iPSYCH cases had ICD-10 codes of F12.1 (cannabis abuse) and/or F12.2 (cannabis dependence) in the Danish Psychiatric Central Research Register[15], and the remaining individuals in iPSYCH were used as controls. deCODE cases met criteria for lifetime DSM-III-R or DSM-IV cannabis abuse or dependence or DSM-5 cannabis use disorder according to diagnoses made at the National Center of Addiction Medicine in Iceland, while controls were derived from the general population of Iceland (Appendix pp 1-7). Exposure data were not available for some large cohorts (e.g., iPSYCH, deCODE), therefore, controls were defined regardless of lifetime cannabis exposure across all datasets.

Genotyping: quality control and imputation

PGC: Standard procedures for GWAS quality control (QC) and imputation were applied using the Ricopili[16] pipeline (https://github.com/Nealelab/ricopili) for case-control cohorts and the Picopili pipeline (https://github.com/Nealelab/picopili) for family-based samples. Briefly, variants in each cohort were filtered for call rate (<5% missingness), followed by individual-level filtering for call rate (<2% missingness) and heterozygosity ($|F_{het}| > .20$). If available, chromosome X variants were checked to ensure concordance between genotype sex and reported sex. Variants were then filtered more stringently: 2% missingness, differential missingness between

cases and controls < 2%, invariant markers and those departing from Hardy-Weinberg equilibrium (HWE) in cases (P > 1e-10) or controls (P > 1e-6) were removed (Appendix pp 7-9). Principal components analysis (PCA) was performed on a stringently quality controlled (QC'ed) set of variants using EIGENSOFT[17,18] to exclude population outliers, infer ancestry among the retained individuals (using the 1000 Genomes Phase 3[19] cosmopolitan reference panel), and derive ancestry-specific principal components for inclusion in analyses (Appendix p 9). Final sample and variant QC procedures, including filters for call rate, heterozygosity, and departure from HWE, were then performed within each ancestry group in each cohort. Each cohort was phased using SHAPEIT[20] and imputed using IMPUTE2[21], using the 1000 Genomes Phase 3[19] cosmopolitan reference panel (Appendix pp 9-10). After imputation, duplicate individuals were removed and cryptic relatedness between cohorts was tested using PLINK[22,23] (individuals who were cryptically-related across cohorts were excluded from all but one cohort, to avoid "double-counting"), and SNPs were filtered for INFO score > 0.8 and minor allele frequency (MAF) ≥ 0.01 prior to analysis (Appendix pp 10-11);

iPSYCH: Quality control of iPSYCH data mirrored the process implemented in PGC, with minor deviations in thresholds for exclusion.

deCODE: Samples were assayed with several Illumina arrays at deCODE genetics. SNPs with low call rate (<95%), significant deviation from Hardy-Weinberg equilibrium (P<0.001), and excessive inheritance error rates (>0.001) were excluded. Variant imputation, based on the IMPUTE HMM model and long-range phasing, was performed as described previously[24]. Variants were further filtered for imputation info score >0.8 and minor allele frequency ≥1% before inclusion into meta-analysis.

Association analyses

Association analyses were conducted separately for each cohort (i.e., 18 individual PGC samples, iPSYCH, and deCODE) by ancestry (European or African – PGC only). For the seven case–control studies from PGC, imputed dosages were analyzed using logistic regression models, implemented in the Ricopili[16] pipeline. For family-based PGC samples, association analyses were conducted with imputed best-guess genotypes using generalized estimating equations (GEE) for samples that included only first-degree relatives (e.g., sibships), and logistic mixed models for complex pedigrees, in the Picopili pipeline (https://github.com/Nealelab/picopili)[9]. For calculation of SNP-heritability and genetic

correlations, subsets of genetically unrelated individuals (N_{CUD} = 5,289, $N_{control}$ = 10,004) were selected from each family-based sample from PGC (Appendix pp 11-12) and analyzed using logistic regression through Picopili; these results were then meta-analyzed along with the case-control cohorts. PGC covariates included sex and 5-10 within-ancestry principal components to account for population stratification (details in Appendix pp 11-12). Because age was not available in all samples, it was not included as a covariate in PGC analyses. However, we adjusted for age and age-squared in CATS and found it to have no impact on study-specific findings.

In the iPSYCH cohort, logistic regression was conducted with imputed dosages, covarying for five ancestral principal components, data processing waves, and the presence of another psychiatric disorder (because iPSYCH was established to study major psychiatric disorders, CUD cases and controls include comorbidity)[4]. Adding sex as a covariate to iPSYCH analyses has been shown not to alter findings[25].

The deCODE Genetics data were analyzed using logistic regression of imputed dosage data with sex, age, and county of origin as covariates[26]. To account for inflation due to population stratification and relatedness, test statistics were divided by an inflation factor estimated from LD score regression (LDSR)[27] (see Appendix p 12).

Meta-analyses within ancestry were conducted using METAL[28] (Appendix pp 12-13). First, summary statistics from case-control and family-based samples were combined, weighted by the effective sample size, because effect sizes from case-control logistic regression analyses and family-based analyses using GEE and logistic mixed models are not directly comparable. The summary statistics were filtered such that a SNP had to be present in at least two of the three contributing GWAS (deCODE, iPSYCH, and PGC). Second, a meta-analysis that excluded related individuals from the family-based PGC samples was performed with an inverse variance-weighted scheme to generate summary statistics that produced effect sizes for use in follow-up analyses (N_{cases} = 14,080, $N_{controls}$ = 343,726). We also completed a trans-ancestral meta-analysis using METAL[28] by combining results across the European (EUR) and African (AFR) ancestry cohorts, comprising 20,916 individuals with CUD (17,068 EUR, 3,848 AFR) and 363,116 controls (357,219 EUR, 5,897 AFR); see Supplemental Table 1. Conditional analyses were conducted in GCTA-COJO[29] by conditioning the meta-analysis summary statistics on the lead genome-wide significant variants.

Gene- and pathway-based tests

The FUMA web-based platform[30] v1.3.5e was used for visualization and annotation and MAGMA[31] was used within the FUMA framework to conduct gene-based association analyses, with SNPs assigned to genes based on physical position (Appendix p 13). We also used Hi-C coupled MAGMA[32] (H-MAGMA), which takes into account long-range regulatory interaction effects to assign non-coding SNPs (intergenic and intronic) to genes based on their chromatin interactions (exonic and promoter SNPs are still assigned to genes based on genomic location; Appendix p 13). Pathway analyses were conducted using PASCAL[33] to test canonical pathways from MSigDB[34] in the EUR sample. All variants within all genes were tested, using default settings, with LD structure estimated using the 1000 Genomes European sample as a reference. We also used S-PrediXcan[35] to examine gene expression differences associated with case-control status, using our CUD summary statistics and transcriptome data from the PredictDB Data Repository (http://predictdb.org) for 11 brain regions, liver tissue, whole blood, and two types of adipose tissue. We included these tissues because CUD is a psychiatric disorder and tetrahydrocannabinol (THC), a key psychoactive cannabis component, accumulates in adipose.[36,37] Analyses were restricted to the EUR-ancestry meta-analysis because the prediction models were trained on reference transcriptome data from GTEx v8[38] using only individuals of European ancestry. The significance threshold was corrected for the total number of gene-tissue pairs tested (75,684 gene-tissue pairs tested; α = 6.69e-7).

Heritability and genetic correlation analyses

Heritability explained by common variants (h^2_{SNP}) and genetic correlations with 23 other traits chosen because of previous findings or hypothesized relationships (Appendix pp 13-14 and Supplemental Table 2) were estimated using LDSR[27,39] on the results of the meta-analysis of case-control subjects of EUR-ancestry. The number of unrelated AFR ancestry cases was below the acceptable sample size threshold for LDSR. Conversion of h^2_{SNP} estimates from observed scale to liability scale was performed using a range of estimated population prevalences from 1% (used by Demontis et al.[4]) to 8.5% (because in some samples we used DSM-IV cannabis abuse *or* dependence[40]). Significance of genetic correlations with other traits was determined using a Bonferroni correction for 23 tests (α = 0.002). Finally, we examined whether the genetic correlations for CUD were significantly different than those for cannabis use using the jackknife procedure implemented through LDSC[39].

Confounding effects of cannabis exposure and smoking behaviors

We used mtCOJO[41] to condition the CUD summary statistics analysis on loci associated with cannabis use at $p < 0.001$[6] (this threshold was chosen to adjust for as many SNPs as possible while retaining computational efficiency). Adjusted summary statistics were used to recompute genetic correlations. Due to the high co-occurrence of cannabis use and tobacco smoking, mtCOJO analyses conditioning the CUD summary statistics for loci significantly associated ($p < 5e-8$) with smoking initiation and cigarettes smoked per day[42] (excluding 23andMe data, due to limited access) were also performed. Given long-standing interest in the comorbidity of schizophrenia and cannabis misuse, we also used mtCOJO to condition the CUD summary statistics on significant schizophrenia loci.

Examining cannabis use in the UK Biobank

We first tested whether CUD and cannabis use in the UK Biobank were genetically correlated by running LDSR on CUD and a broad measure of maximum cannabis use frequency (derived from the Neale lab GWAS of the UK Biobank; https://github.com/Nealelab/UK_Biobank_GWAS). Linear regression was then used to examine the extent to which CUD PRS predicted a pseudo-continuous measure of self-reported cannabis use frequency, while co-varying for age, sex and 20 ancestral principal components (Appendix p 14). PRSice-2 was used to perform gene-set enrichment using gene sets and pathways from the Molecular Signatures Database (MSigDB[43]): (H) hallmark biological processes or states, (C1) positional sets from cytogenetic maps, (C2) chemical or genetic perturbations and canonical pathways, (C3) regulatory processes, (C4) computationally derived gene sets of cancer gene neighborhoods and modules, (C5) biological process, cellular component, and molecular function gene ontologies, (C6) oncogenic signatures, and (C7) immunologic signatures.

Phenome-wide association study (PheWAS) in BioVU biobank

Polygenic scores for CUD were computed using PRS-CS[44] (Appendix pp 14-15) for each of the 66,915 genotyped individuals of European descent in BioVU. Genotyping and QC of this sample have been described elsewhere[45,46]. A logistic regression model was fitted to each of 1,335 case/control phenotypes that had at least 100 cases to estimate the odds of each diagnosis given the CUD polygenic score, after adjustment for sex, median age of the longitudinal EHR measurements, and the top 10 principal components of ancestry. To explore whether pleiotropic effects of the CUD PRS were mediated by smoking behaviors, we conducted two additional

PheWAS sensitivity analyses: (1) a PheWAS on CUD summary statistics that had been conditioned on the top smoking initiation loci using mtCOJO[41], and (2) a PheWAS using a diagnosis of tobacco use disorders (TUD) as an additional covariate in the regression model, which is likely a conservative over-correction given the extremely high comorbidity expected between CUD and TUD. We used a Bonferroni-corrected phenome-wide significance threshold of $p < .05 / 1335 = 3.74e\text{-}5$, given the 1,335 phecodes; this is likely over-conservative because it incorrectly assumes independence between phecodes. PheWAS analyses were run using the PheWAS R package v0.12.[47]

Polygenic risk and brain structure among children in the Adolescent Brain Cognitive Development (ABCD) study

Data from the ongoing Adolescent Brain Cognitive Development (ABCD) study[48] (data release 2.0.1; https://abcdstudy.org/) were used to test whether CUD PRS are associated with brain structure among 4,539 cannabis-naïve (via self-report or hair toxicology) children of European ancestry (mean age = 9.93±0.63 years; 46.8% girls). Total bilateral white matter volume, gray matter volume, and intracranial volume were estimated using FreeSurfer[49] 5.3. Genotyping and quality control are described in the Appendix p 15. PRS from the CUD GWAS were generated at nine p-value thresholds (i.e., P_T = 0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, and 1), as were PRS for cannabis use[6]. Linear mixed-effects models were used to include scanner (for imaging analyses) and family as nested random effects, conducted using the lme4 package in R[50], version 3.6.0. All analyses included the following fixed effect covariates: first 20 ancestral principal components, age, sex, age by sex, parents combined income, caregiver education, genotyping batch, caregiver's marital status, prenatal cannabis exposure before and after knowledge of pregnancy, and twin status. Multiple testing within each brain structure phenotype was accounted for by applying random field theory correction[51] across p-value thresholds, as this method directly models the overlap across the different PRS thresholds and corrects for the statistical dependence among them.

ROLE OF THE FUNDING SOURCE

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

RESULTS

We identified two genome-wide significant loci in the trans-ancestral meta-analysis (AFR+EUR, N_{case} = 20,916; $N_{control}$ = 363,116; Figure S1, Supplemental Table 3). Both loci were significant in the European-ancestry meta-analysis (N_{CUD} = 17,068, $N_{controls}$ = 357,219) but did not reach significance in the much smaller African-ancestry GWAS (N_{case} = 3,848; $N_{control}$ = 5,897). No additional ancestry-specific loci were observed. Inflation in the test statistics (λ = 1.10) most likely reflects the polygenic architecture of CUD, a conclusion that is supported by LDSR (LDSR intercept = 0.99). Conditioning the CUD summary statistics on the lead SNP in each genome-wide significant locus, rs7783012 and rs4732724, revealed no evidence of additional independent significant findings.

Based on effect sizes and LD from the European-ancestry meta-analysis that excluded related individuals (N_{CUD} = 14,080, $N_{controls}$ = 343,726), the genome-wide significant locus on chromosome 8 contains a single association (independent at $R^2 < 0.1$) with lead SNP rs4732724 (OR = 0.89, SE = 0.02, *p* = 6.46e-09; Figures 1, S2 & S3). This locus was previously associated with CUD in the iPSYCH cohort[4], and includes eQTLs for *CHRNA2* (cholinergic receptor nicotinic alpha 2 subunit) in cerebellum and cerebellar hemisphere and *EPHX2* (epoxide hydrolase 2) in cerebellum and adipose tissue (Supplemental Table 4). One genome-wide significant variant in the chromosome 8 locus (rs1565735) had a CADD score of 13.28, indicating high probability of deleteriousness (Supplemental Table 5). There were additional eQTL signals at this chromosome 8 locus, for *CCDC25* (coiled-coil domain containing 25; in nucleus accumbens; multiple SNPs), *CLU* (adipose; rs2640724), and *STMN4* (stathmin 4; in prefrontal cortex; rs78875955 and rs72477506) (Figure S5).

The chromosome 7 locus is located in an intron of *FOXP2* (Forkhead box protein P2; index SNP: rs7783012, OR = 1.11, SE = 0.02, *p* = 1.84e-09; see Figures 1, S2 & S4). The index variant was an eQTL for *FOXP2* in brain (prefrontal cortex, anterior cingulate cortex) and adipose tissue, and demonstrated chromatin interactions with *FOXP2, MDFIC, MIR3666,* and *AC073626.2* (Figure S6).

The gene-wise association analysis of EUR-ancestry summary statistics identified three significant genes (α = 2.664e-6): *FOXP2 (p*=7.31e-08*)*, *PDE4B (p*=6.66e-07*),* and *ENO4* (p=3.51e-08; Figure S7, Supplemental Table 6). No pathways were significant (Supplemental Table 7) were identified. Three genes (*NAT6* (amygdala, cortex, frontal cortex), *HYAL3* (both adipose tissues, whole blood, cerebellum, frontal cortex, hippocampus, nucleus accumbens,

spinal cord), and *IFRD2* (cerebellum)) were significantly related to CUD via genetically-regulated gene expression (Figure S8, Supplemental Tables 8 & 9). Connecting SNPs to genes via chromatin interaction data revealed significant associations in the adult brain tissue (10 genes), fetal brain tissues (12 genes), iPSC-derived astrocytes (11 genes), and iPSC-derived neurons (8 genes); these genes included *HYAL3*, *ENO4, CHRNA2, FOXP2* (Supplemental Tables 10-13, Figure S9).

The SNP-heritability (h^2_{SNP}) for CUD ranged from 0.067 - 0.121 (SE = 0.006 - 0.011) on the liability scale, depending on the estimated population prevalence (h^2_{SNP} = 0.02 (SE = 0.002) on the raw scale). CUD showed significant positive genetic correlations (r_g) with 16 of the 23 studied phenotypes (Figure 2, Supplemental Table 14). The strongest relationships were observed with smoking initiation[42] (r_g = 0.66, p = 3.2e-83), Townsend Deprivation Index (a measure of regional poverty[52]; r_g = 0.58, p = 3.3e-37), educational attainment[53] (r_g = -0.39, p = 6.7e-34), and age at first birth (r_g = -0.49, p = 5.4e-28). Thus, increased CUD risk is genetically correlated with living in an area of greater material poverty, having children at an earlier age, and lower levels of educational attainment. Liability to CUD was positively genetically correlated with alcohol use and tobacco smoking[42], nicotine dependence[54], psychiatric disorders (e.g., ADHD[25], schizophrenia[55], major depression[56]), and body-mass index[57] (BMI).

The r_g between cannabis use[6] and CUD was 0.50 (SE = 0.05, p = 1.5e-21; genetic covariance intercept = 0.014 (SE = 0.005); Supplemental Table 14). Of the eight genome-wide significant SNPs associated with cannabis use[6], only four had p <0.05 in the CUD meta-analysis (Supplemental Table 15; there was modest sample overlap between the two studies). Conditioning the CUD summary statistics for loci associated with cannabis use neither substantially modified the effect sizes of the genome-wide significant loci (rs4732724, Beta = -0.11, SE = 0.02, *p* = 8.25e-09; rs7783012, Beta = 0.10, SE = 0.02, *p* = 2.62e-09) nor identified additional novel loci (see Supplemental Table 16). The heritability of CUD adjusted for cannabis use loci (using mtCOJO[41]) was 0.095 (SE = 0.01) on the liability scale (estimated population prevalence = 8.5%).

A comparison of genetic correlations with other phenotypes revealed similarities and distinctions between cannabis use and CUD (Supplemental Table 17, Figure 2). The genetic correlations with CUD and cannabis use were significantly different for 12 of the 22 traits tested. Both cannabis use[6] and CUD were genetically correlated in the same direction with liability to

smoking initiation, schizophrenia, major depressive disorder, risk tolerance, and the Townsend Deprivation Index. Cannabis use[6] was positively genetically correlated with educational achievement and later age at first birth, and negatively with BMI. In contrast, CUD was genetically correlated with *lower* education attainment, *earlier* age at first birth, and *higher* BMI (i.e., in the *opposite direction*). Liability to CUD was genetically correlated with nicotine dependence (r_g = 0.48, p = 1.35e-09), while the genetic correlation of this trait with cannabis use was not significant (p = 0.44). In contrast, cannabis use was significantly genetically correlated with chronotype (r_g = -0.24, p = 6.40e-19), while CUD showed no significant correlation with this trait (p = 0.18). Conditioning the genetic correlations of CUD on cannabis use loci (with p < 0.001) made little difference in the magnitude of the r_gs (Supplemental Table 18).

Liability to CUD and maximum cannabis frequency were genetically correlated (r_g = 0.75, *p* = 1.80e-6) in the UK Biobank. CUD PRS were significantly associated with our pseudo-continuous measure of cannabis use frequency in the UKB (maximum R^2 = 0.04%, Z = 7.42, p = $1.15x10^{-13}$, P_T = 0.3; Figure S10, Supplemental Table 19). A total of 65/12,461 gene-sets/pathways were significantly enriched (Supplemental Table 20) at P_T = 0.3, highlighting involvement of central nervous system morphogenesis (transcription factor Nkx-2.2 target genes, R^2 = 0.02%, Z = 4.46, p = $8.22x10^{-6}$) and immune responses to exogenous compounds (ZFP91 target genes R^2 = 0.01%, Z = 4.41, p = $1.01x10^{-5}$; $CD4^+$ T-cell R^2 = 0.02%, Z = 4.41, p = $3.79x10^{-6}$; and macrophage gene sets R^2 = 0.01%, Z = 4.62, p = $1.04x10^{-5}$)

Of 1,335 phenotypes in the BioVU biobank, 46 were significantly associated with a PRS for CUD (p < 3.74 x 10^{-5}, Figure 4, Supplemental Table 21). The phenotype groups with the most abundant associations were mental disorders (n=12), the strongest associations being with tobacco use disorder (N_{cases} = 5,280, OR = 1.18, SE = .02, p = 2.66 × 10^{-27}) and substance use disorders (N_{cases} = 6,155, OR=1.18, SE=0.01, p = 1.24 × 10^{-30}), mood disorders (N_{cases} = 9,588, OR=1.09, SE=0.01, p = 2.38 × 10^{-12}) and suicidal ideation or attempt (N_{cases} = 689, OR=1.27, SE=0.04, p = 1.81 × 10^{-9}); respiratory diseases (n=12), such as respiratory failure (N_{cases} = 4,485 , OR=1.11, SE=0.02, p = 4.45 × 10^{-10}) or chronic airway obstruction (N_{cases} = 4,436, OR=1.13, SE=0.02, p = 5.64 × 10^{-14}), endocrine/metabolic conditions (n=3), such as disorders of fluid (N_{cases} = 12,562, OR=1.06, SE=0.01, p = 5.77 × 10^{-8}); infectious diseases (n=4), such as viral hepatitis (N_{cases} = 135, OR=1.3, SE=0.03, p = 3.34 × 10^{-20}); and digestive diseases (n=3), including cirrhosis of liver (e.g. N_{cases} = 1,928, OR=1.14, SE=0.02, p = 2.49 × 10^{-8}).

The secondary pheWAS analysis in BioVU with CUD summary statistics conditioned on smoking initiation revealed attenuated findings, with only ten codes now passing Bonferroni corrections; anxiety disorder, viral hepatitis, and several respiratory codes were still significant (Supplemental Table 22). When we conditioned the pheWAS on tobacco use disorder (TUD) diagnosis, some associations remained significant (respiratory conditions, viral hepatitis), whereas other associations (e.g. anxiety disorder) were no longer significantly associated with CUD PRS (Supplemental Table 23).

CUD PRS were significantly associated with reduced total white matter volume in cannabis-naïve children from the ABCD study (standardized βs ~ -0.04; p = 0.002 to 0.004; Figure 3), explaining up to 0.17% of the variance in white matter volume at the most predictive threshold of $P_T < 0.5$ (Supplemental Table 24). Children in the highest quartile of PRS, on average, had white matter volume that was 1% lower than those in the lowest quartile. Results remained significant when including intracranial volume as a covariate (standardized β = -0.08, p = 0.01) and when excluding children who used any substance (N = 3,282 ; standardized β = -0.05, p = 0.001), or who used any substance *or* were prenatally exposed to any substance (N = 2,057; standardized β = -0.05, p = 0.03). The cannabis use PRS was not correlated with white matter volume (Figure 3). There was no association between CUD PRS or cannabis use PRS and gray matter volume (all ps > 0.01; Figure S11, Supplemental Table 25).

DISCUSSION

This GWAS meta-analysis extended support for one previously identified locus on chromosome 8 and identified a novel variant on chromosome 7. The lead variant (rs7783012) at the chromosome 7 locus is a *cis*-eQTL for *FOXP2* expression in brain and adipose tissue. *FOXP2* was also significantly implicated in gene-based tests that incorporated information about chromatin interactions in iPSC-derived astrocytes (Figure S9). rs7783012 has also been associated with a variety of measures related to externalizing behaviors (e.g., ADHD[25], age at first sexual intercourse[58], generalized risk tolerance[59]) and with educational attainment[53]. *FOXP2* is essential to synaptic plasticity and has been implicated in the normal development of speech and language acquisition[60–62]. However, due to the prominence of the protein product of *FOXP2* as a regulator of numerous genes, indirect pathways of vulnerability beyond risk taking are also possible.

Individual SNPs on chromosome 8 are eQTLs for *CHRNA2* and *EPHX2*, extending prior work by Demontis et al.[4] in iPSYCH, which along with the deCODE data are included in the present analysis and continue to be the main contributors to this finding (p_{iPSYCH} = 5.73e-08, p_{deCODE} = 3.03e-4, $p_{PGC\ (including\ relateds)}$ = 0.06; see Figure S3). The large GWAS of schizophrenia[63] has also implicated this variant (p = 3.68e-6), but conditioning for top schizophrenia loci did not modify the association with CUD (p = 4.33e-8; Supplemental Table 26). Given the role of *CHRNA2* variants in tobacco smoking[42,64,65], it is plausible that the finding for both CUD and schizophrenia are partially driven by the high rates of tobacco use in those populations[66]. However, conditioning on the GWAS of cigarettes per day actually increased the significance of the lead variant rs4732724 (p_{mtcojo_CPD} = 4.16e-09; Supplemental Table 27), although a new SNP was identified as the "lead" SNP (rs11783093). When this new lead SNP was conditioned for the GWAS of smoking initiation, there was an attenuation of the signal ($p_{mtcojo\_smkinitation}$ = 1.55e-06; Supplemental Table 28). These findings suggest that the chromosome 8 signal may be partly driven by smoking initiation, or indicative of a pleiotropic effect with a stronger impact on CUD than on smoking initiation[42]. Despite the plausibility of *CHRNA2* in the etiology of CUD, it is worth noting that *EPHX*, which is involved in the metabolism of cannabinoids[67–69], was also identified in eQTL analyses but not supported by other post-hoc studies.

Cannabis use and CUD were genetically correlated (r_g = 0.50) but conditioning for cannabis use loci did not substantially reduce the heritability of CUD, and although it reduced the significance of the top loci, the effect sizes remained consistent. This is an imperfect method of accounting for possible index-event bias, but we are reassured that conditioning the genetic associations with CUD on genetic loci for cannabis use did not meaningfully change our results. Importantly, cannabis use and CUD show *divergent* genetic relationships with educational attainment[53], body-mass index[57], and age at first birth, with CUD indexing greater impairment in these psychosocial and anthropometric indices. This divergence is similar to that found between alcohol intake and alcohol use disorder[9,10].

Cannabis use frequency in the UK Biobank was genetically correlated with CUD, but, similarly to other psychiatric and behavioral traits[70], the CUD polygenic risk scores explained only a small proportion of variance in cannabis use frequency (R^2 = 0.04%). We also found genetic overlap between CUD and several mental health phenotypes, respiratory illnesses, and infectious diseases in the BioVU biobank. The strongest association was with tobacco use disorder, but conditioning for loci associated with smoking initiation retained many of the pheWAS

associations at significant levels, including anxiety, phobic, and dissociative disorders, respiratory failure, and viral hepatitis. An even more stringent analysis that co-varied for TUD revealed independent associations with viral hepatitis, type 1 diabetes, respiratory measures and pain, but not mental health. These associations could reflect genuine pleiotropy (e.g., with risk-taking behaviors and injection drug use) or index putatively causal peripheral effects of cannabis.

We identified an association between polygenic risk for CUD and lower white matter volume in drug-naïve children. Some prior cross-sectional studies have linked differences in gray matter volume with cannabis use and dependence[71–73]; however, a large mega-analysis did not find reductions in global or regional volumes in cannabis-dependent adults compared with controls[74]. In our study, the association between CUD PRS and white matter volume persisted in the subset of children who were not exposed to any substance, including prenatally. We recapitulate a relationship between CUD and white matter volume with PRS between CUD and cannabis use frequency by detecting enrichment of Nkx-2.2 transcription factor targets. Nkx-2.2 is highly expressed in the brain and plays a critical role in myelin gene expression[75]. This suggests that polygenic liability for CUD might index differences in white matter volume in the developing brain, independently of the onset of substance use behaviors. Still, the CUD PRS white matter association was small (% R^2 ranging from 0.15% to 0.18%), and additional studies are needed to confirm this association.

Some limitations are noteworthy. Our AFR sample was under-powered and warrants greater data collection[76–78]. We had little or no information regarding comorbid psychiatric disorders for the majority of PGC samples, however, we conducted conditional analyses to account for these in a subset of studies and it made little difference. Another limitation is our lack of information regarding lifetime cannabis exposure and the potency of cannabis used by our samples. Our estimates of genome-wide SNP-h^2 (0.07 - 0.12) were far lower than the h^2 estimated from twin and family studies (0.5 - 0.7). This is very common across essentially all psychiatric disorders, and might be due to low power and some heritability residing in variants too rare to be included in our GWAS. An additional limitation is that we did not conduct formal Mendelian Randomization analyses; to do this, we would have needed to remove sample overlap between our CUD GWAS and the other GWAS of interest, which would have greatly decreased our statistical power. However, using latent causal variable (LCV) analyses, we briefly examined the evidence for bidirectional causality between CUD and the top genetically correlated traits:

educational attainment, age at first birth, Townsend Deprivation Index (TDI), smoking initiation, and ADHD. LCV[79] is an approach related to MR but can account for sample overlap among the input GWAS. There was no significant evidence of causal relationships between CUD and any of these traits, with the absolute value of the genetic causality proportion ranging from 0.05 - 0.27, and p-values for the null hypothesis that there is no genetic causality ranging from 0.13 - 0.86 (see Appendix pp 15-16). Estimates of genetic overlap may be sensitive to sample characteristics, e.g., older volunteers in the UK Biobank cohort[80] and some younger registry-based cohorts in our CUD GWAS. In addition, imbalance between cases and controls could have impacted our findings, although we don't observe substantial genetic heterogeneity (Supplemental Figures S3-S4).

In conclusion, our findings provide further evidence that CUD is a serious, psychiatric illness with neurobiological influences that diverge at least partially from cannabis use. From a public health perspective, the recognition that CUD is a serious form of psychopathology should spur efforts to identify and aid at-risk individuals in the face of escalating cannabis use worldwide, especially among adolescents.

ACKNOWLEDGEMENTS

This research was conducted using the UK Biobank Resource, application numbers 4844 and 58146. The iPSYCH team was supported by grants from the Lundbeck Foundation (R165-2013-15320, R102-A9118, R155-2014-1724 and R248-2017-2003), the European Commission (Horizon 2020, grant number 667302) and the universities and university hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. Data handling and analysis on the GenomeDK HPC facility was supported by NIMH (1U01MH109514-01 to ADB). High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to ADB). Work at deCODE genetics was partly funded by the National Institute on Drug Abuse (NIDA) (R01-DA034076). The BioVU dataset(s) used for the PheWAS analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618,

P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at https://victr.vumc.org/biovu-funding/. Dr Degenhardt is funded by the National Institutes of Health (R01 DA144740) and the Australian National Health and Medical Research Council Research [1135991].

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/scientists/workgroups/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.The ABCD data repository grows and changes over time. The ABCD data used in this report came from NIMH Data Archive 2.0.1.

Funding support for the Comorbidity and Trauma Study (CATS)(dbGAP accession number: phs000277.v1.p1) was provided by the National Institute on Drug Abuse (R01 DA17305); GWAS genotyping services at the CIDR at The Johns Hopkins University were supported by the National Institutes of Health (contract N01-HG-65403). Funding support for the Center for Education and Drug Abuse Research (CEDAR)(dbGAP accession number: phs001649.v1.p1 was provided by the National Institute on Drug Abuse (P50 DA005605). The Christchurch Health and Development Study(CHDS) has been supported by funding from the Health Research Council of New Zealand, the National Child Health Research Foundation (Cure Kids), the Canterbury Medical Research Foundation, the New Zealand Lottery Grants Board, the University of Otago, the Carney Centre for Pharmacogenomics, the James Hume Bequest Fund, US NIH grant MH077874, and NIDA grant ‘‘A developmental model of gene-environment interplay in SUDs’’ (R01DA024413) 2007–2012. The Collaborative Study on the Genetics of Alcoholism (COGA) Principal Investigators B. Porjesz, V. Hesselbrock, T. Foroud; Scientific Director, A. Agrawal; Translational Director, D. Dick, includes eleven different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, T. Foroud, J.

Nurnberger Jr., Y. Liu); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz, J. Meyers, C. Kamarajan, A. Pandey); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, A. Brooks, R. Hart); The Children's Hospital of Philadelphia, University of Pennsylvania (L. Almasy); Virginia Commonwealth University (D. Dick, J. Salvatore); Icahn School of Medicine at Mount Sinai (A. Goate, M. Kapoor, P. Slesinger); and Howard University (D. Scott). Other COGA collaborators include: L. Bauer (University of Connecticut); L. Wetherill, X. Xuei, D. Lai, S. O'Connor, M. Plawecki, S. Lourens (Indiana University); L. Acion (University of Iowa); G. Chan (University of Iowa; University of Connecticut); D.B. Chorlian, J. Zhang, S. Kinreich, G. Pandey (SUNY Downstate); M. Chao (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, S. Saccone (Washington University); F. Aliev, P. Barr (Virginia Commonwealth University); H. Chin and A. Parsian are the NIAAA Staff Collaborators. We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting- Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). The authors thank Kim Doheny and Elizabeth Pugh from CIDR and Justin Paschall from the NCBI dbGaP staff for valuable assistance with genotyping and quality control in developing the dataset available at dbGaP (phs000125.v1.p1; also: phs000763.v1.p1; phs000976.v1.p1). Support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI; U01 HG004422; dbGaP study accession phs000092.v1.p1]. SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center [U01 HG004446]. Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism [COGA; U10 AA008401], the Collaborative Genetic Study of Nicotine Dependence [COGEND; P01 CA089392; see also phs000404.v1.p1], and the Family Study of Cocaine Dependence [FSCD; R01 DA013423, R01 DA019963]. Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research (CIDR), was provided by the NIH GEI [U01HG004438], the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High

throughput genotyping for studying the genetic contributions to human disease" [HHSN268200782096C]. The GSMS project (phs000852.v1.p1) was supported by the National Institute on Drug Abuse (U01DA024413, R01DA11301), the National Institute of Mental Health (R01MH063970, R01MH063671, R01MH048085, K01MH093731 and K23MH080230), NARSAD, and the William T. Grant Foundation. We are grateful to all the GSMS and CCC study participants who contributed to this work. The following grants supported data collection and analysis of CADD (dbGAP in progress): DA011015, DA012845, DA021913, DA021905, DA032555, and DA035804. ADAA was funded by NIH grant R01 AA017444. Brisbane Longitudinal Twin Study (BLTS) was supported by the United States National Institute on Drug Abuse (R00DA023549), and by the Australian Research Council (DP0343921, DP0664638, 464914, 619667, FT110100548). Gene-Environment-Development Initiative -GEDI – Virginia Commonwealth University (VTSABD; dbGAP in progress) was supported by the National Institute on Drug Abuse (U01DA024413, R01DA025109), the National Institute of Mental Health (R01MH045268, R01MH055557, R01MH068521), and the Virginia Tobacco Settlement Foundation grant 8520012. We are grateful to all the VTSABD-YAFU-TSA study participants who contributed to this work. Minnesota Center for Twin and Family Research (MCTFR;phs000620.v1.p1) support contributing to this publication was supported by the National Institutes of Health under award number DA005147, DA013240, DA024417, DA036216, AA009367, MH066140, DA042755, DA037904, HG008983, DA044283. Funding for Substance Use Disorder Liability: Candidate System Genes study was supported by R01 DA019157 and P50 DA005605. Yale-Penn (phs000425.v1.p1; phs000952.v1.p1) was supported by National Institutes of Health Grants RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535 and the Veterans Affairs Connecticut and Philadelphia Veterans Affairs Mental Illness Research, Educational, and Clinical Centers. Australian Alcohol and Nicotine studies (OZ-ALC-NAG; phs000181.v1.p1) were supported by National Institutes of Health Grants AA07535,AA07728, AA13320, AA13321, AA14041, AA11998, AA17688,DA012854, and kennedyDA019951; by Grants from the Australian National Health and Medical Research Council (241944, 339462, 389927,389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485,and 552498); by Grants from the Australian Research Council(A7960034, A79906588, A79801419, DP0770096, DP0212016, and DP0343921); and by the 5th Framework Programme (FP-5) GenomEUtwin Project (QLG2-CT-2002-01254). Genome-wide association study genotyping at Center for Inherited Disease Research was supported by a Grant to the late Richard Todd, M.D., Ph.D., former Principal Investigator of Grant AA13320. Irish Affected Sib-Pair Study of Alcohol Dependence

(IASPSAD)GWAS data collection and analysis was supported by National Institute on Alcohol Abuse and Alcoholism grants P20-AA-017828 and P50-AA-022537. Sample collection was supported by R01-AA-011408. Control genotyping was supported by National Institute of Mental Health grant R01-MH-083094 and Wellcome Trust Case Control Consortium 2 grant WTCCC-084710. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

We additionally thank the groups who directly shared GWAS results. We would like to acknowledge all participating groups of the International Cannabis Consortium, and in particular the members of the working group including Sven Stringer, Camelia Minica, Karin Verweij, Hamdi Mbarek, Eske Derks, Nathan Gillespie and Jacqueline Vink. Thanks also to the ENIGMA consortium for providing GWAS results on subcortical brain volumes (available from http://enigma.ini.usc.edu/research/download-enigma-gwas-results/).Finally, we acknowledge the valuable contribution of groups who have publicly released summary statistics from their respective GWAS. Specifically, thanks to researchers from Schumann et al. (2016) including the CHARGE+ and AlcGen consortia (results available at https://grasp.nhlbi.nih.gov/FullResults.aspx) and to all members of Psychiatric Genomics Consortium (PGC; results available for download at http://www.med.unc.edu/pgc/results-and-downloads). Individual-level data from the genotyped cohorts and cohort-level summary statistics will be made available to researchers following an approved analysis proposal through the PGC Substance Use Disorder group with agreement of the cohort PIs; contact the corresponding authors for details. Cohort data are also available from dbGaP except where prohibited by IRB or European Union data restrictions (accession numbers to be available before publication), in particular the working groups on Attention Deficit/Hyperactivity Disorder (ADHD; chaired by Dr. Stephen Faraone), Autism Spectrum Disorder (ASD; chaired by Drs. Mark Daly and Bernard Devlin), and Eating Disorders (ED; chaired by Drs. Cynthia Bulik and Gerome Breen).

DISCLOSURES

Dr. Werge has acted as advisor and lecturer to H. Lundbeck A/S. Dr. Bierut and the spouse of Dr. Saccone are listed as inventors on Issued U.S. Patent 8,080,371,“Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Dr. McIntosh has received research support from Eli Lilly, Janssen, Pfizer, and the Sackler Foundation. Dr. Kranzler is a member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, which was supported in the last three years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. Drs. Kranzler and Gelernter are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018. Dr Degenhardt reports untied educational grant funding to conduct studies of new opioid medications in Australia from Indivior, Mundipharma, Seqirus and Reckitt Benckiser.

DATA AVAILABILITY

Upon publication, summary statistics will be made available for download at https://www.med.unc.edu/pgc/download-results/.

1. Kendler, K. S. *et al.* A population-based Swedish Twin and Sibling Study of cannabis, stimulant and sedative abuse in men. *Drug Alcohol Depend.* **149**, 49–54 (2015).
2. Verweij, K. J. H. *et al.* Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. *Addiction* **105**, 417–430 (2010).
3. Agrawal, A. *et al.* A genome-wide association study of DSM-IV: Cannabis dependence. *Addict. Biol.* **16**, 514–518 (2011).
4. Demontis, D. *et al.* Genome-wide association study implicates CHRNA2 in cannabis use disorder. *Nat. Neurosci.* 1 (2019).
5. 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).
6. Pasman, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21**, 1161–1170 (2018).
7. Agrawal, A., Neale, M. C., Jacobson, K. C., Prescott, C. A. & Kendler, K. S. Illicit drug use and abuse/dependence: modeling of two-stage variables using the CCC approach. *Addict. Behav.* **30**, 1043–1048 (2005).
8. Gillespie, N. A., Neale, M. C. & Kendler, K. S. Pathways to cannabis abuse: a multi-stage model from cannabis availability, cannabis initiation and progression to abuse. *Addiction* **104**, 430–438 (2009).
9. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
10. Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am. J. Psychiatry* (2018).

doi:https://doi.org/10.1176/appi.ajp.2018.18040369
11. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, 1499 (2019).
12. Pedersen, C. B. *et al.* The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6 (2018).
13. Gulcher, J. R., Kristjánsson, K., Gudbjartsson, H. & Stefánsson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
14. Association, A. P. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)*. (Arlington, VA: American Psychiatric Publishing, Inc., 2000).
15. Mors, O., Perto, G. P. & Mortensen, P. B. The Danish psychiatric central research register. *Scand. J. Public Health* **39**, 54–57 (2011).
16. Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *bioRxiv* 587196 (2019).
17. 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).
18. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904 (2006).
19. Consortium, T. 1000 G. P. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
20. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nature Methods* **9**, 179–181 (2011).
21. Hancock, D. B. *et al.* Assessment of Genotype Imputation Performance Using 1000 Genomes in African American Studies. *PLoS One* **7**, e50610 (2012).
22. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* (2007). doi:10.1086/519795
23. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
24. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435 (2015).
25. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63 (2019).
26. Price, A. L. *et al.* The Impact of Divergence Time on the Nature of Population Structure: An Example from Iceland. *PLOS Genet.* **5**, e1000505 (2009).
27. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
28. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
29. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369 (2012).
30. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
31. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, (2015).
32. Nancy, Y., Fauni, H., Ma, W. & Won, H. Connecting gene regulatory relationships to neurobiological mechanisms of brain disorders. *bioRxiv* 681353 (2019).
33. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
34. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740

(2011).
35. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
36. Kreuz, D. S. & Axelrod, J. Delta-9-tetrahydrocannabinol: localization in body fat. *Science (80-.).* **179**, 391–393 (1973).
37. Dinis-Oliveira, R. J. Metabolomics of Δ9-tetrahydrocannabinol: implications in toxicity. *Drug Metab. Rev.* **48**, 80–87 (2016).
38. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
39. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236 (2015).
40. STINSON, F. S., RUAN, W. J., PICKERING, R. & GRANT, B. F. Cannabis use disorders in the USA: prevalence, correlates and co-morbidity. *Psychol. Med.* **36**, 1447–1460 (2006).
41. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
42. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
43. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545 LP – 15550 (2005).
44. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1–10 (2019).
45. Dennis, J. *et al.* Genetic risk for major depressive disorder and loneliness in sex-specific associations with coronary artery disease. *Mol. Psychiatry* 1–11 (2019).
46. Ruderfer, D. M. *et al.* Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *Mol. Psychiatry* 1–9 (2019).
47. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).
48. Lisdahl, K. M. *et al.* Adolescent brain cognitive development (ABCD) study: overview of substance use assessment methods. *Dev. Cogn. Neurosci.* **32**, 80–96 (2018).
49. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
50. Bates, D., Sarkar, D., Bates, M. D. & Matrix, L. The lme4 package. *R Packag. version* **2**, 74 (2007).
51. Nichols, T. E. Multiple testing corrections, nonparametric methods, and random field theory. *Neuroimage* **62**, 811–815 (2012).
52. Morris, R. & Carstairs, V. Which deprivation? A comparison of selected deprivation indexes. *J. Public Health (Bangkok).* **13**, 318–326 (1991).
53. 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).
54. 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* **23**, 1911–1919 (2018).
55. Consortium, S. W. G. of the P. G. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421 (2014).
56. 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).
57. 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).
58. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits.

Nat. Genet. (2019). doi:10.1038/s41588-019-0481-0
59. Linnér, R. K. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245 (2019).
60. Lai, C. S. L., Gerrelli, D., Monaco, A. P., Fisher, S. E. & Copp, A. J. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* **126**, 2455–2462 (2003).
61. MacDermot, K. D. *et al.* Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am. J. Hum. Genet.* **76**, 1074–1080 (2005).
62. Hurst, J. A., Baraitser, M., Auger, E., Graham, F. & Norell, S. An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.* **32**, 352–355 (1990).
63. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381 (2018).
64. Bierut, L. J. *et al.* Variants in nicotinic receptors and risk for nicotine dependence. *Am. J. Psychiatry* **165**, 1163–1171 (2008).
65. Rollema, H. *et al.* Rationale, pharmacology and clinical efficacy of partial agonists of α4β2 nACh receptors for smoking cessation. *Trends Pharmacol. Sci.* **28**, 316–325 (2007).
66. Fowler, I. L., Carr, V. J., Carter, N. T. & Lewin, T. J. Patterns of current and lifetime substance use in schizophrenia. *Schizophr. Bull.* **24**, 443–455 (1998).
67. McDougle, D. R. *et al.* Anti-inflammatory ω-3 endocannabinoid epoxides. *Proc. Natl. Acad. Sci.* **114**, E6034–E6043 (2017).
68. Snider, N. T., Nast, J. A., Tesmer, L. A. & Hollenberg, P. F. A cytochrome P450-derived epoxygenated metabolite of anandamide is a potent cannabinoid receptor 2-selective agonist. *Mol. Pharmacol.* **75**, 965–972 (2009).
69. Wagner, K., Inceoglu, B. & Hammock, B. D. Soluble epoxide hydrolase inhibition, epoxygenated fatty acids and nociception. *Prostaglandins Other Lipid Mediat.* **96**, 76–83 (2011).
70. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* **12**, 44 (2020).
71. Battistella, G. *et al.* Long-Term Effects of Cannabis on Brain Structure. *Neuropsychopharmacology* **39**, 2041–2048 (2014).
72. Demirakca, T. *et al.* Diminished gray matter in the hippocampus of cannabis users: Possible protective effects of cannabidiol. *Drug Alcohol Depend.* **114**, 242–245 (2011).
73. Chye, Y. *et al.* Cannabis-related hippocampal volumetric abnormalities specific to subregions in dependent users. *Psychopharmacology (Berl).* **234**, 2149–2157 (2017).
74. Mackey, S. *et al.* Mega-Analysis of Gray Matter Volume in Substance Dependence: General and Substance-Specific Regional Effects. *Am. J. Psychiatry* **176**, 119–128 (2019).
75. Zhang, C. *et al.* The transcription factor NKX2-2 regulates oligodendrocyte differentiation through domain-specific interactions with transcriptional co-repressors. *J. Biol. Chem.* jbc-RA119 (2020).
76. Martin, A. R. *et al.* Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
77. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584 (2019).
78. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* (2019).
79. O'Connor, L. J. & Price, A. L. *Distinguishing genetic correlation from causation across 52 diseases and complex traits*. (Nature Publishing Group, 2018).
80. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am. J. Epidemiol.* (2017).