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

Emma C Johnson
Alexander S Hatoum
June He
Kathleen K Bucholz
Louis Fox

See next page for additional authors
A large-scale genome-wide association study meta-analysis of cannabis use disorder


Summary

Background Variation in liability to cannabis use disorder has a strong genetic component (estimated twin and family heritability about 50–70%) and is associated with negative outcomes, including increased risk of psychopathology. The aim of the study was to conduct a large genome-wide association study (GWAS) to identify novel genetic variants associated with cannabis use disorder.

Methods To conduct this GWAS meta-analysis of cannabis use disorder and identify associations with genetic loci, we used samples from the Psychiatric Genomics Consortium Substance Use Disorders working group, iPSYCH, and deCODE (21041 case samples, 363 884 control samples in total), contrasting cannabis use disorder cases with controls. To examine the genetic overlap between cannabis use disorder and 22 traits of interest (chosen because of previously published phenotypic correlations [eg, psychiatric disorders] or hypothesised associations [eg, chronotype] with cannabis use disorder), we used linkage disequilibrium score regression to calculate genetic correlations.

Findings We identified two genome-wide significant loci: a novel chromosome 7 locus (FOXP2, lead single-nucleotide polymorphism [SNP] rs7783012; odds ratio [OR] 1.11, 95% CI 1.07–1.15, p=1.11 × 10⁻⁹) and the previously identified chromosome 8 locus (near CHRNA2 and EPHX2, lead SNP rs4732274; OR 0.89, 95% CI 0.86–0.93, p=4.66 × 10⁻⁹). Cannabis use disorder and cannabis use were genetically correlated (r_g 0.50, p=1.50 × 10⁻²¹), but they showed significantly different genetic correlations with 12 of the 22 traits we tested, suggesting at least partially different genetic underpinnings of cannabis use and cannabis use disorder. Cannabis use disorder was positively genetically correlated with other psychopathology, including ADHD, major depression, and schizophrenia.

Interpretation These findings support the theory that cannabis use disorder has shared genetic liability with other psychopathology, and there is a distinction between genetic liability to cannabis use and cannabis use disorder.

Funding National Institute of Mental Health; National Institute on Alcohol Abuse and Alcoholism; National Institute on Drug Abuse; Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing; The European Commission, Horizon 2020; National Institute of Child Health and Human Development; Health Research Council of New Zealand; National Institute on Aging; Wellcome Trust Case Control Consortium; UK Research and Innovation Medical Research Council (UKRI MRC); The Brain & Behavior Research Foundation; National Institute on Deafness and Other Communication Disorders; Substance Abuse and Mental Health Services Administration (SAMHSA); National Institute of Biomedical Imaging and Bioengineering; National Health and Medical Research Council (NHMRC) Australia; Tobacco-Related Disease Research Program of the University of California; Families for Borderline Personality Disorder Research (Beth and Rob Elliott) 2018 NARSAD Young Investigator Grant; 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; National Institutes of Health: Genes, Environment and Health Initiative; National Institutes of Health; National Cancer Institute; The William T Grant Foundation; Australian...
Cannabis use disorders have identified variants reaching genome-wide significance, but inadequate sample sizes identified. Nicotinic acetylcholine receptor, has been robustly locus, tagged by a 2387 cases) and heterogeneity among samples have contributed to a paucity of replicable findings: only one locus, tagged by a cis-eQTL for CHRNA2 (encoding a nicotinic acetylcholine receptor), has been robustly identified. A GWAS of lifetime cannabis use (184 765 total sample size, 43 380 cases) identified eight genome-wide significant loci and 35 significant genes. Twin studies suggest high genetic correlations between early stages of cannabis experimentation and later cannabis use disorder. However, casual cannabis use is affected by a variety of socioeconomic influences and age-period-cohort effects, whereas progression to cannabis use disorder is related to other psychopathologies. Findings have suggested partially distinct genetic causes underlying alcohol consumption and alcohol use disorder, including different genetic associations with other psychiatric disorders and traits. Thus, in addition to examining the genomic liability for cannabis use disorder, we tested whether the genetic influences underlying cannabis use and cannabis use disorder diverge with respect to behavioural and brain measures.

Research in context

Evidence before this study
Cannabis use disorder is heritable (50–70% according to twin and family studies), yet identification of genomic variants associated with cannabis use disorder from genome-wide association studies (GWASs) remains sparse. We surveyed all peer-reviewed journal publications in English on GWASs of cannabis use disorder or cannabis dependence using Google Scholar and PubMed, published between Jan 1, 1990, and April 1, 2020. Search terms included “cannabis dependence”, “cannabis abuse”, “marijuana dependence”, “marijuana use disorder”, “marijuana use disorder”, and “GWAS”. The most promising finding to date is a variant that is a cis-eQTL for CHRNA2 (Demontis and colleagues), which was replicated in an independent dataset for cannabis use disorder. Independently, GWAS of cannabis use have identified multiple genetic risk loci; however, the extent to which the genetics of cannabis use correlates with liability to cannabis use disorder has not been determined. Although GWASs of cannabis use have been studied in the context of a variety of psychiatric and psychosocial correlates, it is expected that some divergent associations will be seen when looking at cannabis use disorder. Previous studies have drawn causal links between cannabis exposure and brain volume, but the relationship between genetic liability to cannabis use disorder and brain volume in individuals naive to cannabis has not yet been studied.

Added value of this study
Our study is the current largest GWAS of cannabis use disorder and the first to include a transancestral component. We found a novel risk locus on chromosome 7. The lead risk variant at this locus is an eQTL for FOXP2—a gene previously implicated in risk-taking behaviours. Contrasting cannabis use and cannabis use disorder, we found that increased liability for cannabis use disorder is genetically correlated with lower educational attainment, early age at first birth, and high body-mass index, traits that show opposite directions of association with lifetime cannabis ever-use. We also found that genetic liability for cannabis use disorder is associated with increased risk of mental health problems, infectious diseases, and respiratory illnesses in a large independent sample. Finally, we found a significant association between increased polygenic liability for cannabis use disorder and low white matter volume in cannabis-naïve children, suggesting a potential role of cannabis-related genetic predisposition in early brain development.

Implications of all the available evidence
Cannabis use disorder is a psychiatric illness that is genetically associated with many negative outcomes (including increased risk of psychiatric disorders and respiratory illnesses). Lifetime cannabis use and cannabis use disorder show at least partially divergent genetic influences and associations with relevant traits. Given increasingly permissive cannabis laws and positive perceptions of its safety, the recognition that cannabis use disorder is a serious psychiatric illness should spur prevention and treatment efforts.
**Table 1: Numbers of cases and controls in meta-analysis**

<table>
<thead>
<tr>
<th></th>
<th>European ancestry</th>
<th>African ancestry</th>
<th>European ancestry—case-control individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td>Cases</td>
</tr>
<tr>
<td>Case-control studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATS</td>
<td>958</td>
<td>453</td>
<td>...</td>
</tr>
<tr>
<td>CADD</td>
<td>397</td>
<td>699</td>
<td>59</td>
</tr>
<tr>
<td>CHDS</td>
<td>201</td>
<td>420</td>
<td>...</td>
</tr>
<tr>
<td>FSCD</td>
<td>226</td>
<td>314</td>
<td>199</td>
</tr>
<tr>
<td>COGEND Nico</td>
<td>306</td>
<td>607</td>
<td>154</td>
</tr>
<tr>
<td>COGEND SAGE</td>
<td>228</td>
<td>830</td>
<td>79</td>
</tr>
<tr>
<td>GEDI-GSMS</td>
<td>81</td>
<td>491</td>
<td>...</td>
</tr>
<tr>
<td>ADAA</td>
<td>...</td>
<td>1000</td>
<td>811</td>
</tr>
<tr>
<td>Total</td>
<td>2397</td>
<td>3814</td>
<td>1491</td>
</tr>
<tr>
<td>Family-based studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLTS</td>
<td>170</td>
<td>1216</td>
<td>...</td>
</tr>
<tr>
<td>MCTFR</td>
<td>449</td>
<td>1625</td>
<td>...</td>
</tr>
<tr>
<td>Yale Penn 1</td>
<td>916</td>
<td>833</td>
<td>1189</td>
</tr>
<tr>
<td>Yale Penn 2</td>
<td>557</td>
<td>497</td>
<td>355</td>
</tr>
<tr>
<td>bigCOGA</td>
<td>2206</td>
<td>5053</td>
<td>813</td>
</tr>
<tr>
<td>CEDAR</td>
<td>64</td>
<td>148</td>
<td>...</td>
</tr>
<tr>
<td>OZ-ALC</td>
<td>593</td>
<td>4893</td>
<td>...</td>
</tr>
<tr>
<td>VTSABD</td>
<td>99</td>
<td>734</td>
<td>...</td>
</tr>
<tr>
<td>IAPSD</td>
<td>104</td>
<td>613</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>5158</td>
<td>15612</td>
<td>2357</td>
</tr>
<tr>
<td>Summary statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Health</td>
<td>847</td>
<td>4839</td>
<td>...</td>
</tr>
<tr>
<td>PGC studies total</td>
<td>8402</td>
<td>24265</td>
<td>3848</td>
</tr>
<tr>
<td>iPSYCH</td>
<td>2758</td>
<td>53326</td>
<td>0</td>
</tr>
<tr>
<td>deCODE</td>
<td>6033</td>
<td>280396</td>
<td>0</td>
</tr>
<tr>
<td>Total (European ancestry)</td>
<td>17193</td>
<td>357987</td>
<td>...</td>
</tr>
<tr>
<td>Total (Transcanteral)</td>
<td>2141 cases; 363 B44 controls</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Methods**

**Samples**

We performed a GWAS of 20 samples in total: 18 from the Psychiatric Genomics Consortium Substance Use Disorders working group (European ancestry 8402 cases, 24265 controls; African ancestry 3848 cases, 5897 controls), one iPSYCH1 sample (European ancestry 2758 cases, 53326 controls), and one deCODE sample (European ancestry 6033 cases, 280396 controls; table 1; appendix pp 2–8).

This study was approved by the institutional review board at Washington University School of Medicine and was done 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 from deCODE were encrypted using a third-party encryption system. The iPSYCH group used pseudonymised unique identifications.

**Measures**

Psychiatric Genomics Consortium cases met criteria for a lifetime diagnosis of DSM-IV (or DSM-III-R) cannabis abuse or dependence derived from clinician ratings or semi-structured interviews. Cases from the iPSYCH sample had ICD-10 codes of F12.1 (cannabis abuse) or F12.2 (cannabis dependence), or both in the Danish Psychiatric Central Research Register; the remaining individuals in the sample were used as controls. Cases in the deCODE sample 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, whereas controls were derived from the general population of Iceland (appendix pp 2–3). Exposure data were not available for some large groups (eg, iPSYCH and deCODE); therefore, controls were defined regardless of lifetime cannabis exposure across all datasets.

**Genotyping: quality control and imputation**

For the Psychiatric Genomics Consortium, standard procedures for GWAS quality control and imputation were applied using the Ricopili pipeline for case-control groups and the Picopili pipeline for family-based samples. Briefly, variants in each group were filtered for call rate (<5% missingness), followed by individual-level filtering for call rate (<2% missingness) and heterozygosity (|F<sup>het</sup>| < 0.20). If available, chromosome X variants were checked to ensure concordance between genotype sex and reported sex. Variants were then filtered more stringently: variants with more than 2% missingness, differential missingness between cases and controls greater than 2%, invariant markers, and those departing from Hardy-Weinberg equilibrium in cases (p < 1·00×10<sup>−6</sup>) or controls (p < 1·00×10<sup>−4</sup>) were removed (appendix pp 8–10). Principal components analysis was done on a stringent quality-control set of variants using EIGENSOFT<sup>TM</sup> to exclude population outliers, infer ancestry among the retained individuals (using the 1000 Genomes Phase 3<sup>rd</sup> cosmopolitan reference panel), and derive ancestry-specific principal components for inclusion in analyses (appendix p 10). Sample and variant quality-controlled procedures, including filters for call rate, heterozygosity, and departure from Hardy-Weinberg equilibrium, were done within each ancestry group in each sample. Each group was phased using SHAPEIT<sup>™</sup> and imputed using IMPUTE2<sup>™</sup> to the 1000 Genomes Phase 3<sup>rd</sup> cosmopolitan reference panel (appendix pp 10–11).
individuals were removed and individuals who were cryptically related across groups were excluded from all but one group (appendix p 12). Single-nucleotide polymorphisms (SNPs) were filtered for INFO score of more than 0·8 and minor allele frequency of at least 0·01 before meta-analysis (appendix pp 13–14).

Quality control of iPSYCH data mirrored the process implemented in the Psychiatric Genomics Consortium, with minor deviations in thresholds for exclusion (appendix p 9). As for 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. We did variant imputation on the basis of the IMPUTE HMM model and long-range phasing. Variables were further filtered for imputation INFO score more than 0·8 and minor allele frequency at least 1% before inclusion in meta-analysis.

Statistical analysis
We did separate association analyses for each sample (ie, 18 individual samples from Psychiatric Genomics Consortium, iPSYCH, and deCODE) by ancestry. For the eight case-control studies from the Psychiatric Genomics Consortium, imputed dosages were analysed using logistic regression models, implemented in the Picopili pipeline.11 For family-based samples of the Psychiatric Genomics Consortium, we did association analyses with imputed dosages, covarying for five ancestral covariates included sex and five to ten within-ancestry subsets of genetically unrelated individuals were selected from each family-based sample from the Psychiatric Genomics Consortium and analysed using logistic regression through Picopili (5289 cases, 10 004 controls). These results were then meta-analysed along with the case-control groups. Psychiatric Genomics Consortium covariates included sex and five to ten within-ancestry principal components to account for population stratification (appendix pp 12–13). Because age was not available in all samples, it was not included as a covariate in the Psychiatric Genomics Consortium analyses. Sensitivity analyses in one representative sample showed this to have no impact on study-specific findings.

In the iPSYCH cohort, logistic regression was done 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, cases of cannabis use disorder and controls include comorbidity).11 Adding sex as a covariate to iPSYCH analyses has been shown not to alter findings.10 deCODE data were analysed using logistic regression of imputed dosage data with sex, age, and county of origin as covariates.27 To account for inflation due to population stratification and relatedness, test statistics were divided by an inflation factor estimated from linkage disequilibrium score regression (LDS; appendix p 13).28 Effective sample size-weighted meta-analyses across case-control and family-based samples within ancestry were done using METAL (appendix pp 13–14).29 First, summary statistics of case-control and family-based samples from the Psychiatric Genomics Consortium were combined and weighted by the effective sample size, because effect sizes from case-control logistic regression analyses and family-based analyses using generalised estimating equations and logistic mixed models are not directly comparable. Then the Psychiatric Genomics Consortium results were meta-analysed with those from the iPSYCH and deCODE samples (between-sample genetic correlations [r] 0·66–0·70). The summary statistics were filtered such that an SNP had to be present in at least two of the three contributing GWASs (deCODE, iPSYCH, and the Psychiatric Genomics Consortium).

We also did a meta-analysis that excluded related individuals from the family-based samples of the Psychiatric Genomics Consortium, using an inverse variance-weighted scheme, to generate summary statistics that produced effect sizes for use in follow-up analyses (14080 cases, 343726 controls). A transancestral meta-analysis using METAL31 combined results across the European and African ancestry cohorts, comprising 21041 individuals with cannabis use disorder (17193 from European ancestry, 3848 from African ancestry) and 363884 controls (357987 from European ancestry, 5897 African ancestry; appendix pp 13–14, 17). Conditional analyses were done in GCTA-COJO24 by conditioning the meta-analysis summary statistics on the lead variants of genome-wide significance.

The FUMA web-based platform32 version 1.3.5e was used for visualisation and annotation, and MAGMA26 was used within the FUMA framework to do genome-based association analyses, with SNPs assigned to genes on the basis of physical position (appendix pp 14–15). We also used Hi-C coupled MAGMA to assign non-coding SNPs (intergenic and intronic) to genes on the basis of their chromatin interactions (exonic and promoter SNPs are still assigned to genes on the basis of their genomic location; appendix p 15).27 Pathway analyses were done using PASCAL to test canonical pathways in the European ancestry sample.28 All variants within all genes were tested, using default settings, with the structure of linkage disequilibrium estimated using the 1000 Genomes European sample as a reference. We used S-PrediXcan39 to examine gene expression differences associated with case-control status, using our summary statistics of cannabis use disorder and transcriptome data from the PredictDB Data Repository for 11 brain regions, liver tissue, whole blood, and two types of adipose tissue. We included these tissues because cannabis use disorder is a psychiatric disorder and

University, Reykjavik, Iceland (O F Goddardsson); Department of Psychiatry, University of Utah, Salt Lake City, UT, USA (D E Adkins PhD); Department of Psychology (A E Adkins PhD, Prof D M Dick PhD), and Department of Human & Molecular Genetics (Prof D M Dick, J I Silberg PhD, H H Maes PhD), College Behavioral and Emotional Health Institute (A E Adkins), Virginia Commonwealth University Alcohol Research Center (J Alexander PhD, S A Bacaro PhD, B T Webb PhD), Prof B P Riley PhD, and Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics (R E Peterson PhD, J I Silberg, N A Gillespie PhD, Prof K S Kendler MD, H H Maes), Virginia Commonwealth University, Richmond, VA, USA; Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA (T B Bigeli PhD, J I Meyers PhD, Prof B Porjesz PhD), Department of Psychological Medicine (Prof I Boden PhD, Prof J Howard MSc), Biostatistics and Computational Biology Unit (J F Pearson PhD), and Department of Pathology and Biomedical Science (J F Pearson, Prof M A Kennedy PhD), University of Otago, Christchurch, New Zealand; Department for Congenital Disorders, Center for Neonatal Screening, Statens Serum Institut, Copenhagen, Denmark (J Ryberg-Grauholm, D M Hougaard); Department of Molecular, Cellular, and Developmental Biology (Prof K Krauter PhD), Department of Integrative Physiology (P A McQueen ScD), and Institute for Behavioral Genetics, (R P Corley PhD, Prof J Hewitt PhD, Prof M C Stallings PhD), University of Colorado Boulder, Boulder, CO, USA (Prof K Krauter); National Drug and Alcohol Research Centre, University of New South Wales, Sydney, NSW, Australia (Prof L Degenhardt PhD); Stanford University Graduate School of Education (B W Domingue PhD), and Department of Psychiatry
tetrahydrocannabinol, a key psychoactive cannabis component, accumulates in adipose tissue. Analyses were restricted to the European ancestry meta-analysis because the prediction models were trained on reference transcriptome data from GTEx version 8 using only individuals of European ancestry. The significance threshold was corrected for the total number of gene-tissue pairs tested (75,684 tested, $\alpha=6\times10^{-7}$).

Heritability explained by common variants ($h^2_{\text{SNP}}$) and genetic correlations with 23 other traits chosen because of previous findings or hypothesised relationships (including cannabis use; appendix pp 15–17) were estimated using LDSR$^{35}$ on the results of the meta-analysis of case-control individuals of European ancestry (the number of unrelated cases of African ancestry was less than the acceptable sample size threshold for LDSR). Conversion of $h^2_{\text{SNP}}$ estimates from observed scale to liability scale was done using a range of estimated population prevalences from 16% to 8.5% (because in some samples we used DSM-IV cannabis abuse or dependence).$^{35}$ Significance of genetic correlations with other traits was determined using a Bonferroni correction for 23 tests (including with cannabis use; $\alpha=0.002$). Finally, we examined whether the genetic correlations for cannabis use disorder were significantly different than those for cannabis use using the jackknife procedure implemented through LDSR.$^{32}$

To investigate potential causal relationships, we did latent causal variable analyses on cannabis use disorder and the top genetically correlated traits: educational attainment, age at first birth, Townsend Deprivation Index, smoking initiation, and ADHD (appendix p 16).$^{34}$

We used mtCOJO$^{35}$ to condition the summary statistics of cannabis use disorder on loci associated with cannabis use at p<0.001 to adjust for as many SNPs as possible while retaining computational efficiency. Adjusted summary statistics were used to recompute genetic correlations. Because of the high co-occurrence of cannabis use and tobacco smoking, we also did mtCOJO analyses to condition the summary statistics of cannabis use disorder for loci significantly associated with smoking initiation and cigarettes smoked per day$^{38}$ (p<5.00×10⁻⁸; excluding 23andMe data, because of restricted access). Moreover, given long-standing interest in the comorbidity of schizophrenia and cannabis misuse, we used mtCOJO to condition the summary statistics of cannabis use disorder on significant schizophrenia loci.$^{37}$

LDSR was used to estimate the genetic correlation between cannabis use disorder and a broad measure of maximum cannabis use frequency. Linear regression was then used to examine the extent to which polygenic risk scores (PRS) for cannabis use disorder predicted a pseudocontinuous measure of self-reported cannabis use frequency, while covarying for age, sex, and 20 ancestral principal components (appendix p 16). PRSice-Z$^{29}$ was also used to do gene-set enrichment using gene sets and pathways from the Molecular Signatures Database.$^{39}$

PRS for cannabis use disorder were computed using PRS-CS$^{30}$ for each of the 66,915 genotyped individuals of European descent in BioVU (appendix pp 16–17). Genotying and quality control of this sample have been described elsewhere.$^{41}$ A logistic regression model was fitted to each of 1335 case or control phenotypes that had at least 100 cases to estimate the odds of each diagnosis given the PRS for cannabis use disorder, after adjustment for sex, median age of the longitudinal electronic health record measurements, and the top ten ancestral principal controls. To explore whether pleiotropic effects of the PRS for cannabis use disorder were mediated by smoking behaviors, we did two phenotype-wide association study (PheWAS) sensitivity analyses: a PheWAS on summary statistics of cannabis use disorder that had been conditioned on the top smoking initiation loci using mtCOJO,$^{35}$ and a PheWAS using a diagnosis of tobacco use disorders as an additional covariate in the regression model, which is a conservative over-correction given the extremely high comorbidity expected between cannabis use disorder and tobacco use disorder. We used a Bonferroni-corrected phenome-wide significance threshold of 3.74×10⁻⁸; this is overly conservative because it incorrectly assumes independence between phenotypes. PheWAS analyses were run using the PheWAS R package, version 0.12.$^{42}$

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

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 transancestral meta-analysis of cannabis use disorder (African and European ancestries, 21041 cases, 363 884 controls; appendix pp 17, 20). These loci were significant in the European ancestry meta-analysis but did not reach significance in the much smaller African ancestry analysis (17 193 cases, 357 987 controls vs 3848 cases, 5 897 controls; table 2). The lead SNPs were rs4732724 on chromosome 8 (\(p_{\text{transancestral}} = 2.64 \times 10^{-9}\), \(p_{\text{European}} = 2.43 \times 10^{-9}\)) and rs57783012 on chromosome 7 (\(p_{\text{transancestral}} = 5.97 \times 10^{-8}\), \(p_{\text{European}} = 5.95 \times 10^{-8}\)). The lead SNP rs4732724 on chromosome 8 contains a single association (independent at \(R^2 < 0.1\)) with lead SNP rs4732724 (odds ratio [OR] 1.11, 95% CI 1.07–1.15, SE 0.02; \(p=1.84 \times 10^{-9}\), figure 1, appendix pp 21–24). The index variant was the iPSYCH sample and includes eQTLs for CHRNA2 (cholinergic receptor nicotinic α2 subunit) in the cerebellum and cerebellar hemisphere and EPHX2 (epoxide hydrolase 2) in the cerebellum and adipose tissue.

Based on effect sizes and linkage disequilibrium from the case-control European ancestry meta-analysis (cases 14 080, 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 (odds ratio [OR] 0.89, 95% CI 0.86–0.93, SE 0.02; \(p=6.46 \times 10^{-9}\); figure 1, appendix pp 21–23). This locus was previously associated with cannabis use disorder in the iPSYCH sample and includes eQTLs for CHRNA2 (cholinergic receptor nicotinic α2 subunit) in the cerebellum and cerebellar hemisphere and EPHX2 (epoxide hydrolase 2) in the cerebellum and adipose tissue.

Inflation in the test statistics (\(k=1.10\)) probably reflects the polygenic architecture of cannabis use disorder, a conclusion supported by LDSR (LDSR intercept 0.99). Conditioning the summary statistics of cannabis use disorder on the lead SNP in each genome-wide significant locus, rs7783012 and rs4732724, did not reveal additional independent significant findings.
The gene-wise association analysis of European ancestry summary statistics identified three significant genes (\(\alpha=2.664 \times 10^{-9}\)): FOXP2 (\(p=7.31 \times 10^{-9}\)), PDE4B (\(p=6.66 \times 10^{-8}\)), and ENO4 (\(p=3.51 \times 10^{-9}\), appendix pp 17, 27). No pathways were significant (appendix p 18). Three genes, NAT6 (amygdala, cortex, frontal cortex), HYAL3 (both adipose tissues, whole blood, cerebellum, frontal cortex, hippocampus, nucleus accumbens, and spinal cord), and IFRD2 (cerebellum) were significantly related to cannabis use disorder through genetically regulated gene expression (appendix pp 18, 28). Connecting SNPs to genes via chromatin interaction data revealed significant associations in adult brain tissue (ten genes), fetal brain tissues (12 genes), iPSC-derived astrocytes (11 genes), and iPSC-derived neurons (eight genes); these genes included HYAL3, ENO4, CHRNA2, and FOXP2 (appendix pp 18, 29).

The SNP-heritability (\(h^2_{SNP}\)) for cannabis use disorder was \(0.067 - 0.121\) (SE 0.006 - 0.011) on the liability scale, depending on the estimated population prevalence and \(h^2_{SNP}=0.02\) (SE 0.002) on the raw scale. Cannabis use disorder showed significant \(r_g\) with 16 of the 23 studied phenotypes, for which the strongest relationships were observed with smoking initiation (\(r_g=0.66\), \(p=3.20 \times 10^{-8}\)), Townsend Deprivation Index (a measure of regional poverty; \(r_g=0.58\), \(p=3.30 \times 10^{-7}\)), educational attainment (\(r_g=0.39\), \(p=6.70 \times 10^{-3}\)), and age at which first child is born (\(r_g=0.49\), \(p=5.40 \times 10^{-28}\), figure 2, appendix p 18). Thus, increased risk of cannabis use disorder is genetically correlated with increased liability for smoking initiation, living in an area of high material poverty, having children at an early age, and low levels of educational attainment.

Liability to cannabis use disorder was also positively genetically correlated with alcohol use, nicotine dependence, psychiatric disorders (eg, ADHD, schizophrenia, major depression), and body-mass index (BMI).

The \(r_g\) between cannabis use and cannabis use disorder was \(0.50\) (SE 0.05, \(p=1.50 \times 10^{-7}\)). Of the eight genome-wide significant SNPs associated with cannabis use, only four had \(p<0.05\) in the meta-analysis of cannabis use disorder (modest sample overlap between the two studies; genetic covariance intercept 0.014 [SE 0.005]). Conditioning the summary statistics of cannabis use disorder for loci associated with cannabis use either substantially modified the effect sizes of the genome-wide significant loci (\(rs\text{4732724, }\beta=0.11, SE=0.02, p=8.25 \times 10^{-9}\), \(rs\text{7783012, }\beta=0.10, SE=0.02, p=2.62 \times 10^{-9}\)) or identified additional novel loci (appendix p 18). The heritability of cannabis use disorder adjusted for cannabis use loci (using \(mt\text{COJO}^{10}\)) was 0.095 (SE 0.01) on the liability scale (estimated population prevalence 8.5%).

The \(r_g\)s with cannabis use disorder and cannabis use were significantly different for 12 of the 22 traits compared (figure 2, appendix p 18). Cannabis use\(^{1}\) and cannabis use disorder were positively genetically correlated with liability to smoking initiation, schizophrenia, major depressive disorder, risk tolerance, and the Townsend Deprivation Index. Cannabis use\(^{3}\) was positively genetically correlated with educational achievement and later age at birth of first child, and negatively with BMI. In contrast, cannabis use disorder was genetically correlated with low education attainment, early age at birth of first child, and high BMI. Liability to cannabis use disorder was genetically correlated with nicotine dependence (\(r_g=0.48, SE=1.35 \times 10^{-9}\)), whereas 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 (\(r_g=0.24, SE=6.40 \times 10^{-9}\)), whereas cannabis use disorder showed no significant correlation with this trait (\(p=0.18\)). Conditioning the \(r_g\) of cannabis use disorder on cannabis use loci (with \(p<0.001\)) made little difference in the magnitude of the \(r_g\)s (appendix p 18).

We found no evidence of genetically causal relationships between liability to cannabis use disorder and to any of the most highly correlated traits (ie, educational attainment, age at first birth, Townsend Deprivation...
Figure 3: PheWAS associations between polygenic risk for CUD and phenotypes in the BioVU biobank.

The 46 phenotypes shown are significantly associated with CUD ($p<3.74 \times 10^{-5}$, corrected for 1335 phenotypes tested). CUD=cannabis use disorder. PheWAS=phenotype-wide association study. NOS=not otherwise specified. SIRS=systemic inflammatory response syndrome.
Articles

Index, smoking initiation, or ADHD; genetic causality proportion 0.05–0.27, p=0.128–0.856; appendix p 18).

Liability to cannabis use disorder and maximum cannabis use frequency in the UK Biobank were genetically correlated (r = 0.75, p=1.80×10⁻⁶). PRS for cannabis use disorder were significantly associated with our pseucontinuous measure of cannabis use frequency in the UK Biobank (maximum R² 0.04%, Z 7.42, p=1.15×10⁻¹³, threshold p<0.3; appendix p 18, 30).

65 of 12 461 gene sets and pathways were significantly enriched, highlighting involvement of CNS morphogenesis (transcription factor Nkx–2.2 target genes, R² 0.02%, Z 4.46, p=8.22×10⁻⁹) and immune responses to exogenous compounds (ZFPI91 target genes R² 0.01%, Z 4.41, p=1.01×10⁻⁵; CD4⁺ T-cell R² 0.02%, Z 4.41, p=3.79×10⁻⁶; and macrophage gene sets R² 0.01%, Z 4.62, p=1.04×10⁻⁵; appendix p 18).

Of 1335 phenotypes in the BioVU biobank, 46 were significantly associated with the PRS for cannabis use disorder (p=3.74×10⁻⁵; figure 3, appendix p 18). The phenotype groups with the most abundant associations were mental disorders (n=12), the strongest associations being with tobacco use disorder (cases 5280, OR 1.18, 95% CI 1.10–1.19; SE 0.01, 95% CI 1.07–1.15; p=4.45×10⁻¹²) and suicidal ideation or attempt (cases 689, OR 1.27, SE 0.04, 95% CI 1.17–1.37; p=1.81×10⁻⁹); respiratory disorders (n=12), such as respiratory failure (cases 4485, OR 1.11, SE 0.02, 95% CI 1.07–1.15; p=4.45×10⁻¹⁰) or chronic airway obstruction (cases 4436, OR 1.13, SE 0.02, 95% CI 1.09–1.18; p=5.64×10⁻¹⁰); endocrine or metabolic conditions (n=3), such as disorders of fluid (cases 12 562, OR 1.06, SE 0.01, 95% CI 1.04–1.08; p=5.77×10⁻⁹); infectious diseases (n=4), such as viral hepatitis (cases 135, OR 1.3, SE 0.03, 95% CI 1.23–1.38; p=3.34×10⁻¹⁰); and digestive diseases (n=3), including cirrhosis of liver (cases 1928, OR 1.14, SE 0.02, 95% CI 1.10–1.19; p=2.49×10⁻⁹).

A secondary pheWAS analysis in BioVU using summary statistics of cannabis use disorder 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. When we conditioned the pheWAS on tobacco use disorder, some associations remained significant (respiratory conditions, viral hepatitis), whereas other associations (eg, anxiety disorder) were no longer associated with PRS for cannabis use disorder (appendix p 18).

The PRS for cannabis use disorder were significantly associated with reduced total white matter volume in cannabis-naive children from the ABCD Study (standardised β=0.04; p=0.001; figure 4), explaining up to 0.18% of the variance in white matter volume at the most predictive threshold of p<0.5 (appendix p 18). Children in the highest quartile of PRS, on average, had a white matter volume that was 1% lower than those in the lowest quartile. Results remained significant when including intracranial volume as a covariate (standardised β=0.08, p=0.01) and when excluding 1246 (27%) of 4539 children who used any substance (standardised β=0.05, p=0.001), or when excluding 2482 (54%) of 4539 who used any substance or were prenatally exposed to any substance (standardised β=0.05, p=0.03). The PRS for cannabis use were not significantly correlated with white matter volume (figure 3). After correction for multiple testing, there was no association between PRS for cannabis use disorder or cannabis use and grey matter volume (all p>0.01; appendix p 18, 31).

**Discussion**

This GWAS meta-analysis confirmed one previously identified locus on chromosome 8 as associated with cannabis use disorder and identified a new locus on chromosome 7. The lead variant at the chromosome 7 locus (rs7783012) 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 (appendix p 29). rs7783012 has also been associated with measures related to externalising behaviours (eg, ADHD,⁴¹ age at first sexual intercourse,⁴¹ generalised risk tolerance)⁴¹ and with educational attainment.⁴¹ FOXP2 is essential to synaptic plasticity and has been implicated in the normal development of speech and language acquisition.⁴¹ However, because of 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 previous work by

---

**Figure 4:** Polygenic risk score associations with white matter volume in drug-naive children

Total white matter volume was regressed on polygenic risk scores for CUD and cannabis use (in separate models). CUD=cannabis use disorder.
Demontis and colleagues5 in iPSYCH, with replication in the deCODE data. Note that iPSYCH and deCODE are the main contributors to this finding in the meta-analysis (PFOREST=5.73×10−8, PdeCODE=0.0003, PRGC=0.06; appendix p 23). A large GWAS of schizophrenia5 has also implicated this variant (P=3.68×10−9), but conditioning for top schizophrenia loci did not modify the association with cannabis use disorder (P=4.33×10−8; appendix p 18). The role of CHRNA2 variants in tobacco smoking,6 it is plausible that the findings for cannabis use disorder and schizophrenia are partially driven by the high rates of tobacco use in those populations.7 However, conditioning cannabis use disorder on the GWAS of cigarettes per day increased the significance of the lead variant rs4732724 (P=4.16×10−9; appendix p 18), although a different SNP was identified as the lead SNP (rs11783093). When rs11783093 was conditioned for the GWAS of smoking initiation, the signal was attenuated (P=0.55×10−6; appendix p 18). These findings suggest that the chromosome 8 signal might be partly driven by smoking initiation, or indicative of a pleiotropic effect with a stronger impact on cannabis use disorder than on smoking initiation.8 Despite the plausibility of CHRNA2 in the cause of cannabis use disorder, it is worth noting that EPHX2, which is involved in the metabolism of cannabinoids,9,10 was also identified in eQTL analyses but not supported by other post-hoc analyses (appendix p 29).

Cannabis use and cannabis use disorder were modestly genetically correlated (r² 0.50) but conditioning for cannabis use loci did not substantially reduce the heritability of cannabis use disorder, and although it reduced the significance of the top loci, the effect sizes remained consistent. Although this does not fully account for possible index-event bias,11 it suggests that the findings are not due to cannabis exposure alone. Cannabis use and cannabis use disorder also show divergent genetic relationships with educational attainment,12 BMI,13 and age at birth of first child, with cannabis use disorder indexing greater impairment in these psychosocial and anthropometric indices than cannabis use disorder. This divergence is similar to that found between alcohol intake and alcohol use disorder.14

We found genetic overlap between cannabis use disorder 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 covaried for tobacco use disorder revealed independent associations with viral hepatitis, type 1 diabetes, respiratory measures, and pain, but not mental health. These associations could reflect genuine pleiotropy (eg, with risk-taking behaviours and injection drug use) or index putatively causal peripheral effects of cannabis.

Cannabis use frequency in the UK Biobank was genetically correlated with cannabis use disorder as well, but, similarly to other psychiatric and behavioural traits,15 the PRS for cannabis use disorder explained only a small proportion of variance in cannabis use frequency (R² 0.04%).

Some previous cross-sectional studies have linked differences in grey matter volume with cannabis use and dependence;16 however, a large mega-analysis did not find reductions in global or regional volumes in cannabis-dependent adults compared with controls.17 In our study, the association between PRS for cannabis use disorder and white matter volume persisted in the subset of children who were not exposed to any substance, including prenatally. This finding suggests that polygenic liability for cannabis use disorder might index differences in white matter volume in the developing brain, independently of the onset of substance use behaviours. Still, the association between PRS for cannabis use disorder and white matter was small (R² 0.15–0.18%), and additional studies are needed to confirm this association.

Some limitations are noteworthy. Our African ancestry sample was under-powered; more data are needed, particularly in light of potential disparities that result from a majority of genetic studies focusing on European-ancestry populations.18,19 We had little or no information regarding comorbid psychiatric disorders for the majority of PGC samples; however, we did conditional analyses to account for these and it made little difference. Information regarding lifetime cannabis exposure and the potency of cannabis used was scarce. Our estimates of genome-wide SNP-h² were far lower than the h² estimated from twin and family studies (0.07–0.12 vs 0.5–0.7). This discrepancy between pedigree-estimated heritability and SNP-heritability is common across essentially all substance use disorders, and might be due to low power, some heritability residing in variants too rare to be included in our GWAS, and insufficient coverage of optimal common-variant genomic coverage in available microarray data even after imputation. An additional limitation is that we did not do formal Mendelian randomisation analysis. To do this analysis, we would have needed to remove sample overlap between our cannabis use disorder GWAS and the other GWASs of interest, which would have greatly decreased our statistical power. However, after doing latent causal variable analyses,17 an approach related to mendelian randomisation that can account for sample overlap among the input GWAS, there was no significant evidence of causal relationships between liability to cannabis use disorder and to any of the top genetically correlated traits: educational attainment, age at first birth, Townsend Deprivation Index, smoking initiation, or ADHD (appendix p 16). Overall, estimates of genetic overlap might also be sensitive to sample characteristics, such as older volunteers in the UK Biobank cohort and some younger registry-based cohorts in our GWAS.

---

www.thelancet.com psychiatry Vol 7 December 2020
addition, imbalance between cases and controls could have affected our findings, although we did not observe substantial genetic heterogeneity (appendix pp 23–24).

In conclusion, our findings provide further evidence that cannabis use disorder is a serious, psychiatric illness with genetic and neurobiological influences that diverge at least partially from cannabis use. From a public health perspective, the recognition that cannabis use disorder 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.

Contributors
SAB, DB, ADB, JHo, GWM, MN, AA, RB, JHE, IBH, JG, ECJ, TET, and RKW designed the study. MJA, SAB, JDB, AMG, AMM, AMH, HKR, DL, BTW, RET, S-A, VB, BSM, AEKA, KSC, WEC, ADB, FRW, SEM, JG-B, LD, NGM, DBH, JHo, RPC, LW, MAK, DMD, JPR, MDR, MCS, MMV, BPR, LJF, LB, JKB, KG, JAG, RAG, DFG, KMH, SMH, ACH, JKHE, IBH, DMH, WGI, EOJ, KJF, PAKM, HFM, MM, MBM, GWM, OM, PBM, ECN, JFP, BP, VR, NLS, BS, JLS, TET, TT, SV, TLW, TM, MJW, and RW collected and interpreted the data. DAA, AMP, AR, PSH, IBH, DD, RET, BSM, FRW, SEM, TBB, SEBP, LHM, DBH, UK, MAK, DEA, SZ, JA, GWW, AR, BB, SS-R, TKC, BWD, NAG, SDG, DFH, SMH, JHE, ECJ, EOJ, REP, PAL, CWM, GWM, TET, and RKW, and ARW analysed the data. MJA, DAA, DL, RP, ASH, DD, S-A, DBH, DEA, GWW, AR, BB, SS-R, JDB, LKD, JHE, JG, DFG, JHo, JHE, ECJ, TET, and RKW wrote the manuscript. SAB, AMG, HKR, DL, LAF, RP, ASH, DD, RET, WEC, ADB, SEP, DBH, JHo, RPC, LW, MAK, DMD, MCS, MMV, GWR, AA, RB, JMN, SS-R, JDB, KKB, LKD, JHE, JG, ECJ, TET, and RKW edited the manuscript. AMM, KSK, SEM, LD, DMD, JLM, BPR, AA, RB, JHE, JG, NAG, ACH, EOJ, HHM, ECN, and TET provided phenotype expertise. CJH, AMM, LAF, ADB, KSK, MMG, HA, AA, RB, LKD, JHE, TMF, JG, KMH, JKHE, WGI, HHM, ECN, BP, KS, TET, SV, RKW, TM, and MJW supervised the study.

Declaration of interests
TW has acted as an advisor and a lecturer to H Lundbeck A/S. LB and the spouse of NLS are listed as inventors on Issued US Patent 8 080 371, Markers for Addiction, covering the use of particular SNPs in determining the diagnosis, prognosis, and treatment of addiction. AMM has received research support from Eli Lilly, Janssen, Pfizer, and the Sackler Foundation. HKR is a member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, which was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala. Neurosciences, HKR and JG are named as inventors on PCT patent application number 15/826 460, entitled Genotype-guided dosing of opioid agonists, filed Jan 24, 2018. LD reports untied educational grant support from Indivior, Arbor and Amygdala, was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala. DR received compensation for his time working with the drug development team for the preparation of this article. The authors declare no other interests.

Data sharing
Summary statistics will be made available for download at https://www.medune.uc/pnp/download-results/.

Acknowledgments
The Psychiatric Genomics Consortium’s Substance Use Disorders working group is supported by MH109532 with funding from the National Institute of Mental Health (NIMH) and the National Institute on Drug Abuse (NIDA). We gratefully acknowledge previous support from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and thank all our contributing investigators and study participants who made this research possible. We acknowledge Dr E Jane Costello for leadership in collecting and characterising the Gene-Environment-Development Initiative–Great Smoky Mountains Study sample and her valuable insights on the development of this study and interpretation of findings. This research was done using the UK Biobank Resource, application numbers 4844 and 5846. Individual funding was provided by the following grants: MH109532 (AA, JHE, JG); F32AA027435 (ECJ); K02DA013253 (AA); U101MH109534-01 (ADB); ST21DA007451 (AH); T24MH108951 (DAAB); NIH F32 MH112598 (FRW); Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (ADB); the European Commission, Horizon 2020, grant number 667302 (ADB); R01-DA034076 (TET); R01-HD060726 (BWD, KMH, JDB, MBM); Health Research Council of New Zealand 16/600 (JB, JHo, MK, JFP); Health Research Council of New Zealand 11/792 (JB, JHo); AA287287, AG616122 (RB); U24DA041147 (SAB); R01HD099655, R01DA041207, R01DA042082 (WEC); DA01015, DA083904 (RPK); K2A AA018755 (DMD); P01-HD011703, R01-HD073342 (KMI); R01 AA081673, DA25509, DA024413, DA016977 (H1MM); Wellcome Trust (104063/Z/14/Z, 21676/7/Z/19/Z), UKRI MRC (MC_PC_1709, MR/S053818/1) (AMM); R01MH111348, The Brain & Behavior Research Foundation NARSAD grant 28612 (REP); R21 DA075272, R21 DC080989 (RP); SAMHSA Grant # H779R105168 (MFR); R01AA026911, R21DA046971 (NLS); R01HD050735, U48EB020403 (subaward), NHMC Australia (496682; 1009064) (MJW); Tobacco-Related Disease Research Program of the University of California Grant Number T29RT0626, Families for Borderline Personality Disorder Research (Beth and Rob Elliott) 2018 NARSAD Young Investigator Grant #26766 (SS-R). The iPSCYCH team was supported by grants from the Lundbeck Foundation (R165–2013–15120, R102-A9118, R155–2014–274, and R28–2017–2003), the European Commission (Horizon 2020, grant number 667302), and the universities and university hospitals of Aarhus and Copenhagen, Denmark. The Danish National BioBank resource was supported by the Novo Nordisk Foundation. Data handling and analysis 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 NIDA (R01–DA04076). The BioVU dataset(s) used for the PhEWS 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 grants include the NIH funded Shared Instrumentation Grant S10RR025141, and Clinical Translational Sciences Award grants UL1TR000243, UL1TR000445, and UL1R024975. Genomic data are also supported by investigator-led projects that include 1U01HG004798, R01NS032830, RC2GM092618, R01NS032830, R05GM115305, U101HG00678, U101HG065962, and R01HD074711, and additional funding sources https://vict.vumc.org/bovu-funding/.

For the full list of supporters see https://abctstudy.org/federal-partners.html.

For the ABCD consortium investigators see https://abctstudy.org/scientists/workgroups/.

For the full list of supporters see https://abctstudy.org/federal-partners.html.

For the additional funding sources see https://vict.vumc.org/biovu-funding/.
Pharmacogenomics, the James Hume Bequest Fund, US NIH grant MH077874, and NIDA grant, a developmental model of gene-environment interplay in substance use disorders in young adults (U01/DA024413) 2007–12. The Collaborative Study on the Genetics of Alcoholism (COGA): Principal Investigators B Porjesz, V Hesselbrock, and T Foroud; Scientific Director, A Agrawal; Translational Director, D Dick) includes 11 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 Krzan); 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 Tischfeld, A Brooks, R Hart); The Children’s Hospital of Philadelphia, University of Pennsylvania (L Almasy); Virginia Commonwealth University (D Dick, J Jutabia); 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 Wertherli, X Xuei, D Lai, S O’Connor, M Plawecki, S Lourens (Indiana University); L Acton (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 Sacone (Washington University); F Aliev, P Barr (Virginia Commonwealth University); and H Chin, and A Parsian (NIDA Data Collection Coordinators). We express our deepest memories of Henri Begleiter and Theodore Reich, founding Principal Investigator and Co-Principal Investigator 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 crucial contributions. This national collaborative study is supported by NIAA grant U10AA08401 from NIAA and NIDA. We thank Kim Doheny and Elizabeth Pugh from the Center for Inherited Disease Research and Justin Paschall from the NCBII dbGaP staff for valuable assistance with genotyping and quality control in developing the dataset available at dbGaP (phs000125.v1.p1, phs000097.v1.p1). Support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIAA, National Institutes of Health, and Environment Initiative (GEI); U01 HG004446; dbGaP study accession phs000092.v1.p1. SAGE is one of the GWAS funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with genotype harmonization and as well as with the general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004466). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by COGA (U10 AA008041, the Collaborative Genetic Study of Nicotine Dependence (P01 CA083912; see also phs000404.v1.p3), and the Family Study of Cocaine Dependence (R01 DA13423, R01 DA09963). Support for funding genotyping, as was done at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIAA grant N01AA-00002, and the NIDA contract. General population-based genetic resource centers were supported by the NIAA and the NIDA grant contract High throughput genotyping for studying the genetic contributions to human disease (HHSN268200728926C). The Great Smoky Mountains Study project (phs000852.v1.p1) was supported by NIDA (U01DA024413, R01DA13103), NIMH (R01MH063970, R01MH063671, R01MH048085, R01MH035731, and R23MH082033), NARSAD, and the William T Grant Foundation. We are grateful to the Great Smoky Mountains Study and Caring for Children the Community study participants who contributed to this work. The following contributions data collection and analysis of Center on Antisocial Drug Dependence (dbGaP in progress): DA010105, DA012845, DA021913, DA012195, DA012355, and DA0135804. Alcohol Dependence in African-Americans was funded by NIH grant R01 AA07744. Brisbane Longitudinal Twin Study was supported by the US NIDA (R01DA023549), and by the Australian Research Council (DP0343921, DP066468, 464914, 619667, FT1000548). Gene-Environment-Development Initiative Virginia Commonwealth University (VTSABD) dbGaP in progress was supported by NIDA (U01DA024413, R01DA023509), NIMH (R01MH045268, R01MH055557, R01MH036823), and the Virginia Tobacco Settlement Foundation, grant 8520012. We are grateful to all the Virginia Twin Studies of Adolescent Behavioral Development, Young Adult Follow-Up, Transitions to Substance use study participants who contributed to this work. Minnesota Center for Twin and Family Research (phs000620.v1.p1) contributing to this publication was supported by the NIH under award numbers DA005137, DA001240, DA002417, DA036216, AA009367, MH066140, DA042755, DA37904, HG008983, and DA046283. 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 NIH grants RC2 DA028099, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA13310, and R01 AA175345 and the Veterans Affairs Connecticut and Philadelphia Veterans Affairs Mental Illness Research, Educational, and Clinical Centers. Australian Alcohol and Nicotine studies (OZ–AAL-NAG; phs000181.v1.p1) were supported by NIH grants AA07353, AA07728, AA13320, AA13321, AA14041, AA15198, AA17688, DA012854, and DA019951, by grants from the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389398, 442915, 442981, 496739, 525485, and 552498), by grants from the Australian Research Council (A7960034, A79906588, A7980149, D0770096, D0212016, and D03P43921), and by the 5th Framework Programme (FP5) GenoMEtwin Project (L1G2-CT-2002-01254). GWAS genotyping at Center for Inherited Disease Research was supported by a grant to the late Richard Todds, MD, PhD, former Principal Investigator of grant AA13320. Irish Affected Sib-Pair Study of Alcohol Dependence GWAS data collection and analysis was supported by NIAAs grants P20- AA-017828 and P50-AA-022537. Sample collection was supported by R01-AA-01408. Control genotyping was supported by NIMH grant R01-MH-081094 and Welcome Trust Case Control Consortium grant 2WTCCC-08470. This research uses data from Add Health, a programme project directed by Kathleen Mullian Harris and designed by J Richard Udry, Peter S Bearman, and Kathleen Mullian 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. No direct support was ever 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, Hamid HBarek, Eiske Dekker, Nathan Gillespie, and Jacqueline Vine. We also wish to thank the ENIGMA consortium for providing GWAS results on subcortical brain volumes. Finally, we acknowledge the valuable contribution of groups who have publicly released summary statistics from their respective GWAS. Specifically, thanks to researchers from Schumann and colleagues (2016) including the CHARGE+ and AlzGene consortia and to all members of the Psychiatric Genomics Consortium (PGC). 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 Principal Investigators; contact the corresponding authors for details. Some cohort data are also available from dbGaP except when prohibited by Institutional Review Board, funding requirements or European Union data restrictions (accession numbers: phs000727.v1.p1, phs000092.v1.p1, phs000852.v1.p1, phs000404.v1.p1, phs000092.v1.p1, phs000620.v1.p1, phs000243.v1.p1, phs000181.v1.p1, phs000761.v1.p1, phs00025.v1.p1). References 1 Verweij KJH, Zietsch BP, Lynskey MT, et al. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. Addiction 2010; 105: 347–356. 2 Agrawal A, Lynskey MT, Hinrichs A, et al. A genome-wide association study of DSM-IV cannabis dependence. Addict Biol 2011; 16: 514–18. 3 Dennis DJ, Rajagopal VM, Thorgeirsson TE, et al. Genome-wide association study implicates CHRNA2 in cannabis use disorder. Nat Neurosci 2019; 22: 1066–74. 4 Sherwa, R, Wang Q, Kranzler H, et al. Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. JAMA Psychiatry 2016; 73: 472–80.


60 Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med 2020; 12: 44.


