ARTICLE

# Genome-wide association studies of lifetime and frequency of cannabis use in 131,895 individuals

Hayley H. A. Thorpe[1], Pierre Fontanillas[2], John J. Meredith[3], Mariela V. Jennings[3], Renata B. Cupertino[3], Shreya R. Pakala[3], on behalf of the 23andMe Research Team*, Sarah L. Elson[2], Jibran Y. Khokhar[1], Lea K. Davis[4,5,6], Emma C. Johnson[7], Abraham A. Palmer[3,8] and Sandra Sanchez-Roige[3,5,8]✉



Cannabis is one of the most widely used drugs globally. We performed genome-wide association studies (**GWASs**) of lifetime ($N = 131,895$) and frequency ($N = 73,374$) of cannabis use. For lifetime cannabis use, we identified two loci, one near *CADM2* (rs35827242, $p = 4.63E-12$) and another near *GRM3* (rs12673181, $p = 6.90E-09$). For frequency of cannabis use, we identified one locus near *CADM2* (rs4856591, $p = 8.10E-09$; $r^2 = 0.76$ with rs35827242). Lifetime and frequency of cannabis use were heritable (12.88 vs. 6.63%) and genetically correlated with previous GWASs of lifetime use and cannabis use disorder (**CUD**), as well as other substance use and cognitive traits. Polygenic scores (**PGSs**) for lifetime and frequency of cannabis use predicted cannabis use phenotypes in *All of Us* participants. A phenome-wide association study using a PGS for lifetime cannabis use to interrogate a hospital cohort replicated prior associations with substance use and mood disorders, and uncovered novel associations with celiac and infectious diseases. This work demonstrates the utility of pre-addiction phenotypes in cannabis use genomic discovery.

*Molecular Psychiatry*; https://doi.org/10.1038/s41380-025-03219-2

## INTRODUCTION

Cannabis use is widespread, with approximately 209 million people globally reporting use in 2020 [1]. The number of people who use cannabis regularly is expected to increase as cannabis is decriminalized in many jurisdictions [2–4]. While people report using cannabis for medicinal purposes [5], there is increasing evidence that cannabis use has short- and long-term adverse consequences across psychiatric, cognitive, and physical health [6–14]. A prior study estimated that 27% of those who use cannabis in their lifetime will develop cannabis use disorder (**CUD**) [15], in which cannabis use becomes problematic to an individual's intra- and interpersonal wellbeing [16]. However, it is currently unclear what factors contribute most to the development of CUD.

Twin studies have estimated that problematic cannabis use is 51–78% heritable [17–19] and recent genome-wide association studies (**GWASs**) have identified hundreds of loci that are associated with CUD [20–23]. While CUD GWASs are of paramount importance, they have several limitations. First, they only examine one extreme of the addiction spectrum and do not address other substance-related behaviors such as recreational use and escalation of intake [24]. These pre-addiction phenotypes [25] precede an individual's progression to a substance use disorder (**SUD**) diagnosis [26–32] and are heritable [17, 26, 31, 33]. However, aside from GWASs of lifetime cannabis use (having *ever* versus *never* used cannabis) [34, 35], the genetics of other pre-addiction cannabis use traits are understudied [36, 37]. Second, only a portion of those engaging in frequent cannabis use seek treatment or have a CUD diagnosis [38, 39]. It is therefore unlikely that CUD GWASs and downstream analyses fully characterize the genetics of regular, potentially problematic cannabis use and its relationships with physical and mental health. Third, collecting individuals diagnosed with CUD is costly because it requires detailed assessments. In contrast, pre-addiction phenotypes can be rapidly and inexpensively collected via self-report in large population-based cohorts [40].

We collected data from 23andMe, Inc. research participants by asking if they had ever used cannabis ($N = 131,895$). Those who responded yes were asked a follow-up question about the number of days they used cannabis during their period of heaviest use (i.e., 30 days; $N = 73,374$), which provided a measure of cannabis use frequency. For both traits we performed GWASs and a battery of secondary analyses to compare biological, genetic, and phenotypic associations. Because the frequency of cannabis use distinguishes between light and heavy use whereas lifetime use does not, and because of work from a prior smoking GWAS indicating that the genetic architecture of lifetime use is distinct from that of consumption and tobacco use disorder [41], we hypothesized that the genetics of frequency of cannabis use would more closely resemble CUD compared to lifetime cannabis use.

[1]Department of Anatomy and Cell Biology, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada. [2]23andMe, Inc., Sunnyvale, CA, USA. [3]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [4]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [5]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [6]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA. [7]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [8]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. *A list of authors and their affiliations appears at the end of the paper. ✉email: sanchezroige@ucsd.edu

Received: 3 July 2024 Revised: 29 July 2025 Accepted: 28 August 2025
Published online: 13 October 2025

## METHODS

### Participants and GWASs

Lifetime and frequency of cannabis use GWASs were conducted in male and female 23andMe research participants who were genetically similar to a European reference sample, as previously described [42]. Ancestry falls along a spectrum [43, 44]; individuals were only included in the analysis if they had >97% European genetic similarity (see Supplementary Methods), as determined through local ancestry analysis [45]. Participants provided informed consent and volunteered to participate in research online under a protocol approved by the external AAHRPP-accredited Institutional Review Board (**IRB**), Ethical & Independent (**E&I**) Review Services. As of 2022, E&I Review Services is part of Salus IRB (https://www.versiticlinicaltrials.org/salusirb). During 4 months in 2015 and 14 months between 2018 to 2020, participants completed a questionnaire surveying a range of personal and behavior characteristics. Included in this survey were questions on lifetime substance use and substance use frequency. Specifically, "Yes" or "No" responses to the question *"Have you ever in your life used marijuana?"* were collected as a measure of lifetime cannabis use. If participants answered "Yes", they were prompted to answer the question *"How many days did you use marijuana during your heaviest 30 days?"* as a measure of frequency of cannabis use. Participants could respond with an integer between 0 and 30 days.

For lifetime cannabis use and frequency of cannabis use, 23andMe conducted GWASs of up to 33,419,581 imputed genetic variants using linear regression and assuming an additive genetic model. Samples were genotyped on one of five genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550 + BeadChip, including about 25,000 custom single nucleotide polymorphisms (**SNPs**) selected by 23andMe, with a total of ~560,000 SNPs. The V3 platform was based on the Illumina OmniExpress + BeadChip, with custom content to improve the overlap with our V2 array, with a total of ~950,000 SNPs. The V4 platform is a fully custom array, including a lower redundancy subset of V2 and V3 SNPs with additional coverage of lower-frequency coding variation, and ~570,000 SNPs. The V5 platform is an Illumina Infinium Global Screening Array (~640,000 variants) supplemented with ~50,000 variants of custom content. All samples included in this study reached at least a 98.5% call rate. We excluded SNPs of low genotyping quality, including those that failed a Mendelian transmission test in trios or with large allele frequency discrepancies compared to European 1000 Genomes reference data, failed Hardy-Weinberg Equilibrium testing, failed batch effects testing, or had a call rate < 90%, as well as SNPs with a minor allele frequency (**MAF**) < 0.1% and imputed variants with low imputation quality (INFO score < 0.50) or with evidence of batch effects (Supplementary Table 3). Model covariates included age, sex, the first 5 genetic principal components (**PCs**), and indicator variables for genotype platforms (see Supplementary Methods for additional details). Only unrelated participants were included. For full details on genotyping and GWASs, see Supplementary Methods.

### Functional annotation and gene-based analyses

*Functional annotation*. Using the web-based platform Functional Mapping and Annotation of Genome-Wide Association Studies (**FUMA** v1.3.8), SNPs were annotated based on ANNOVAR categories, Combined Annotation Dependent Depletion scores, RegulomeDB scores, expression quantitative trait loci (**eQTLs**), and chromatin state predicted by ChromHMM. Novel SNPs were identified as those neither in linkage disequilibrium (**LD**; $r^2 < 0.10$) nor within ±1 Mb of GWAS-significant SNPs reported by other GWASs of cannabis use traits (e.g., initiation, CUD) sourced from the literature [20–23, 34, 35, 46–51] and from the EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/). Novel genes were identified as those not identified by gene-based analyses in other cannabis-related studies [22, 34, 35, 52–57] or with start/stop positions within ±1 Mb of previously uncovered GWAS-significant SNPs.

*MAGMA gene-based and pathway analyses*. We used Multi-marker Analysis of GenoMic Annotation (**MAGMA**, v1.08, Ensembl build v92), which is included in FUMA, to annotate SNPs to protein-coding genes. LD was estimated using the 1000 Genomes European reference sample, and significance was determined by Bonferroni correction ($p < 2.53E-06$). Gene-set analysis was conducted on 10,678 gene-sets and Gene Ontology terms curated from the Molecular Signatures Database (MsigDB v7.0). Tissue-specific gene expression profiles were assessed in 54 tissue types and 30 general tissue types with average gene expression in each tissue used as a covariate. Using Genome-Tissue Expression (**GTEx**, v8) RNA-seq data, gene expression values were $\log_2$ transformed from the average Reads Per Kilobase Million (max value = 50) per tissue. Significance was determined following Bonferroni correction ($p < 9.26E-04$ for 54 tissue types; $p < 1.67E-03$ for 30 general tissue types).

*H-MAGMA*. We incorporated lifetime and frequency of cannabis use GWAS data with chromatin interaction profiles from human brain tissue using Hi-C coupled MAGMA (**H-MAGMA**) [58]. H-MAGMA assigns non-coding SNPs to genes based on chromatin interactions from fetal brain, adult brain, midbrain neuron, cortical neuron, iPSC-derived neuron, and iPSC-derived astrocyte datasets (https://github.com/thewonlab/H-MAGMA). Exonic and promoter SNPs were assigned to genes based on physical position [58]. We applied a Bonferroni correction based on the total number of gene-tissue pairs tested ($p < 9.42E-07$ to $9.45E-07$).

*S-PrediXcan*. We performed a transcriptome-wide association study using S-PrediXcan (v0.7.5) to identify eQTL-linked genes associated with lifetime and frequency of cannabis use [59]. S-PrediXcan uses genetic information to predict gene expression levels in various tissues and tests if eQTLs correlate with lifetime or frequency of cannabis use across 49 bodily tissues ($N_{genes} = 1619$ to 9949). S-PrediXcan uses precomputed tissue weights from the GTEx project database (https://www.gtexportal.org/) as the reference transcriptome dataset via Elastic net models. As input data, we included summary statistics, transcriptome tissue data, and covariance matrices of the SNPs within each gene model (HapMap SNP set available at the PredictDB Data Repository) [59] from all available tissues. We applied Bonferroni correction for each tissue type ($p < 3.09E-05$ to $5.03E-06$).

### LDSC heritability and genetic correlations across health, psychiatric, and anthropomorphic traits

Linkage Disequilibrium Score regression (**LDSC**; https://github.com/bulik/ldsc) was used to calculate SNP-based heritability ($\boldsymbol{h2_{SNP}}$) and genetic correlations ($\boldsymbol{r_g}$) [60]. $h2_{SNP}$ was calculated from pre-computed LD scores ("eur_w_ld_chr/"). $r_g$ were calculated between lifetime or frequency of cannabis use with 292 other traits across 22 health, psychiatric, and lifestyle categories (Supplementary Methods). We applied a 5% false discovery rate (**FDR**) correction to account for multiple testing.

### Polygenic score analyses

*Polygenic scores of lifetime, daily, and problematic cannabis use in AoU*. We tested the associations between lifetime or frequency of cannabis use polygenic scores (**PGSs**) with cannabis traits available for *All of Us* (**AoU**) participants clustering within a European ($N = 29,523–120,529$) or African ($N = 14,201–52,577$) genetic ancestry panel (for details, see *All of Us* Research Program Genomics Investigators [61]). AoU is a diverse health database currently including survey responses, physical measurements, genotyping data, and electronic health records (**EHR**) for over 400,000 individuals living in the United States [61, 62]. Using survey and EHR data, participants were assigned binary identifiers for lifetime cannabis use (concept id: 1585636), daily cannabis use among those who reported cannabis use in their lifetime (concept id: 1585650), and problematic cannabis use (concept ids: 434327, 440387, 440996, 433452, 437838, 4323639, 4103419, 435231, 434019, 434328; Supplementary Methods).

We calculated PGSs in male and female participants who had available phenotypes and genotypes in the Allele Count/Allele Frequency (**ACAF**) threshold SNP callset curated by AoU, which includes SNPs of MAF > 1% or allele counts over 100 for each ancestral subpopulation. Using PRS-CS "auto" v1.1.0 [63], the SNP set was filtered to biallelic SNPs present in the HapMap3 European ancestry set and SNPs were weighted. Lifetime and frequency of cannabis use PGSs were created from 782,975 weighted SNPs using the allelic-scoring function *score* in PLINK (v1.9) [63]. The base R function *glm* was used to fit logistic regression models for each cannabis use trait using PGS(s), as well as the additional covariates of age, sex, and the first 10 global PCs provided by AoU. Models included single PGS models (lifetime or frequency PGS + additional covariates), a joint-PGS model (lifetime PGS + frequency PGS + additional covariates), and a null model (additional covariates only). For the joint-PGS model, Bonferroni correction was applied for two tests (lifetime PGS and frequency PGS) and three outcomes (lifetime, daily, and problematic cannabis use) for a total of $N = 6$ comparisons ($p < 8.33E-03$); single PGS models were corrected for one test and three outcomes ($N = 3$, $p < 1.67E-02$). Joint-PGS liability scale $R^2$ values were calculated as previously described [64] using the *NagelkerkeR2* function in the R package *fmsb* (v0.7.6) and the estimated prevalence of cannabis use traits in US adults (Supplementary Methods).

PGS $\Delta R^2$ was calculated by subtracting $R^2$ calculated with models including PGS from the $R^2$ of the null model.

*Phenome- and laboratory-wide association analyses in a hospital cohort (BioVU)*. We tested associations between lifetime or frequency of cannabis use PGSs and medical condition liability using data from the Vanderbilt University Medical Center (**VUMC**; IRB #160302, #172020, #190418) [65]. The BioVU cohort, a subset of VUMC biobank participants ($N = 72{,}821$), provided genotyping data and EHR containing clinical data and laboratory-assessed biomarkers [63, 65, 66]. For each unrelated European ($N = 66{,}917$) and African ($N = 12{,}383$) BioVU participant based on genetic similarity, we computed lifetime and frequency of cannabis use PGSs using PRS-CS (v1.1.0) [63].

For our phenome-wide association study (**PheWAS**), we fitted a logistic regression model to each case/control disease phenotypes ("phecodes") to estimate the log odds of each diagnosis given lifetime cannabis use/frequency of cannabis use PGS, while adjusting for sex, median age of the longitudinal EHR, and the first 10 PCs with the PheWAS v0.12R package [63]. At least two International Disease Classification (**ICD**) codes mapping to a PheWAS disease category (Phecode Map 1.2; https://phewascatalog.org/phecodes) and a minimum of 100 cases were required for phecode inclusion. We also conducted additional sensitivity analyses using tobacco use disorder (**TUD**; phecode 318) and CUD (see Supplementary Table 12 for CUD ICD codes) as covariates. We calculated the 5% FDR for all associations performed ($N = 1405$).

For our laboratory-wide association study (**LabWAS**), we implemented a pipeline as previously described [66]. LabWAS associates PGS with laboratory biomarkers (i.e., measurements) evaluated in BioVU participants. LabWAS uses the median, inverse normal quantile transformed age-adjusted values from the QualityLab pipeline in a linear regression to determine the association between lifetime or frequency of cannabis use PGSs with 314 phenotypes. We controlled for the same covariates as for the PheWAS analyses, excluding median age because the pipeline corrects for age using cubic splines with 4 knots. We applied 5% FDR correction across all LabWAS associations performed ($N = 314$).

## RESULTS

### GWASs of lifetime cannabis use and frequency of cannabis use uncover associations with CADM2 and GRM3

Participant demographics are described in Supplementary Table 1. The cohort was 65.2% female with a mean age of 52.79 ± 0.04 years old. Participant responses to surveys about lifetime and frequency of cannabis use are available in Supplementary Table 2 and Supplementary Fig. 1.

For SNP quality control, see Supplementary Table 3. Genomic control inflation factors for lifetime cannabis use ($\lambda = 1.08$) and frequency of cannabis use ($\lambda = 1.03$) suggested no substantial inflation due to population stratification for either GWAS. $h2_{SNP}$ was 12.88% ± 0.97 for lifetime cannabis use, greater than the $h2_{SNP}$ for lifetime cannabis use from the International Cannabis Consortium (**ICC**; $h2_{SNP} = 6.63\% \pm 0.43$) [34]. $h2_{SNP}$ for frequency of cannabis use was 4.12% ± 0.72 (Supplementary Table 4).

We identified two genome-wide significant ($p < 5.00\text{E-}08$) loci for lifetime cannabis use on chromosomes 3 and 7 (Fig. 1A, Supplementary Figs. 2–3, Supplementary Table 5). The most significant association was with rs35827242 ($p = 4.63\text{E-}12$, chr3p12.1) located upstream the Cell adhesion molecule 2 gene (***CADM2***), replicating findings from previous lifetime use [34] and CUD [22, 23] GWASs. *CADM2* encodes a glycoprotein primarily expressed in the brain with functions in cell-cell adhesion, synaptic formation, excitatory neurotransmission, and energy homeostasis [67, 68]. We also found a novel association between lifetime cannabis use and rs12673181 ($p = 6.90\text{E-}09$, chr7q21.11), which is a SNP upstream of the Metabotropic glutamate receptor 3 gene (***GRM3***) encoding $mGlu_3$. $mGlu_3$ is an inhibitory group II receptor affecting a range of intracellular signaling cascades and cellular processes like glutamate neurotransmission and long-term plasticity [69].

Frequency of cannabis use GWAS identified one significant association with rs4856591 ($p = 8.10\text{E-}09$, chr3p12.1; Fig. 1B, Supplementary Figs. 2, 5), which is near *CADM2* and is in LD with rs11922956 ($r^2 = 0.76$, $p < 1.00\text{E-}04$).

### Secondary analysis identifies 40 lifetime and 4 frequency of cannabis use genes

Mapping SNPs to genes via gene-based (i.e., MAGMA, H-MAGMA) and transcriptome-wide association study (**TWAS**; i.e., S-PrediXcan) analyses identified 40 candidate genes associated with lifetime cannabis use (Supplementary Tables 6–8), and 4 candidate genes associated with frequency of cannabis use (Supplementary Tables 9). None of the 4 genes associated with frequency of cannabis use (i.e., *MMS22L*, *DSCC1*, *CPSF7*, *RP11-51J9.6*) were implicated in lifetime cannabis use. The only gene to overlap across gene-based and TWAS analyses was *CADM2* (Supplementary Table 10). The 44 genes associated with lifetime and frequency of cannabis use clustered together in approximately 18 chromosomal regions (Supplementary Table 10); several of these genes clustered at the same locus, suggesting that some associations may reflect LD rather than independent signals. 29 of these genes have not been identified in prior cannabis-related GWASs (Supplementary Table 10).

Gene-set and tissue-based enrichment analyses yielded no significant results (Supplementary Tables 11–12).

### Lifetime and frequency of cannabis use are genetically correlated with psychiatric, cognitive, and physical health traits

Out of 292 traits, we identified 115 traits that were genetically correlated with lifetime cannabis use and 38 that were genetically correlated with frequency of cannabis use after applying a 5% FDR correction (Figs. 2–3, Supplementary Table 13). We identified 29 traits that were significantly genetically correlated with both lifetime and frequency of cannabis use (10 anthropomorphic traits; 19 psychiatric traits), which were usually consistent in their direction of effect, with four exceptions: "intelligence" and "executive function", which were positively genetically correlated with frequency of use but negatively genetically correlated with lifetime use, and "tense/highly strung" and "delay discounting", which were negatively genetically correlated with frequency of use but positively genetically correlated with lifetime use (Supplementary Fig. 5).

*Cannabis and other substance use traits*. The genetic correlation between lifetime and frequency of cannabis use was moderate ($r_g = 0.54 \pm 0.08$, $p = 1.89\text{E-}10$), suggesting imperfect genetic overlap between the two traits. We identified positive genetic correlations between CUD and lifetime ($r_g = 0.62 \pm 0.04$, $p = 2.44\text{E-}59$), as well as frequency of cannabis use ($r_g = 0.45 \pm 0.07$, $p = 2.45\text{E-}10$; Fig. 2). Compared to lifetime cannabis use from the ICC, our lifetime cannabis use trait was more strongly genetically correlated with CUD (23andMe-CUD $r_g = 0.62 \pm 0.04$, $p = 2.44\text{E-}59$ vs. ICC-CUD $r_g = 0.48 \pm 0.04$, $p = 4.30\text{E-}33$). We identified positive genetic correlations with other aspects of substance use such as drug experimentation and lifetime cannabis use $r_g = 0.97 \pm 0.01$, $p < 1.35\text{E-}161$) and frequency of cannabis use ($r_g = 0.54 \pm 0.07$, $p = 5.45\text{E-}14$). We also observed a genetic correlation between the Alcohol Use Disorder Identification Test (**AUDIT**) problems and lifetime cannabis use ($r_g = 0.46 \pm 0.06$, $p = 1.26\text{E-}16$) and frequency of cannabis use ($r_g = 0.30 \pm 0.10$, $p = 2.46\text{E-}03$). Additional genetic correlations are shown in Fig. 3 and Supplementary Table 13.

*Psychiatric disorders*. Lifetime cannabis use was modestly genetically correlated with schizophrenia ($r_g = 0.15 \pm 0.03$, $p = 7.33\text{E-}07$); however, frequency of cannabis use was not ($r_g = 0.02 \pm 0.05$, $p = 0.73$). We also identified associations with other psychiatric traits and lifetime cannabis use like attention-deficit hyperactivity disorder ($r_g = 0.31 \pm 0.05$, $p = 5.20\text{E-}12$), depression ($r_g = 0.22 \pm 0.04$, $p = 3.52\text{E-}10$), and cross-disorder ($r_g = 0.30 \pm 0.05$, $p = 3.91\text{E-}$

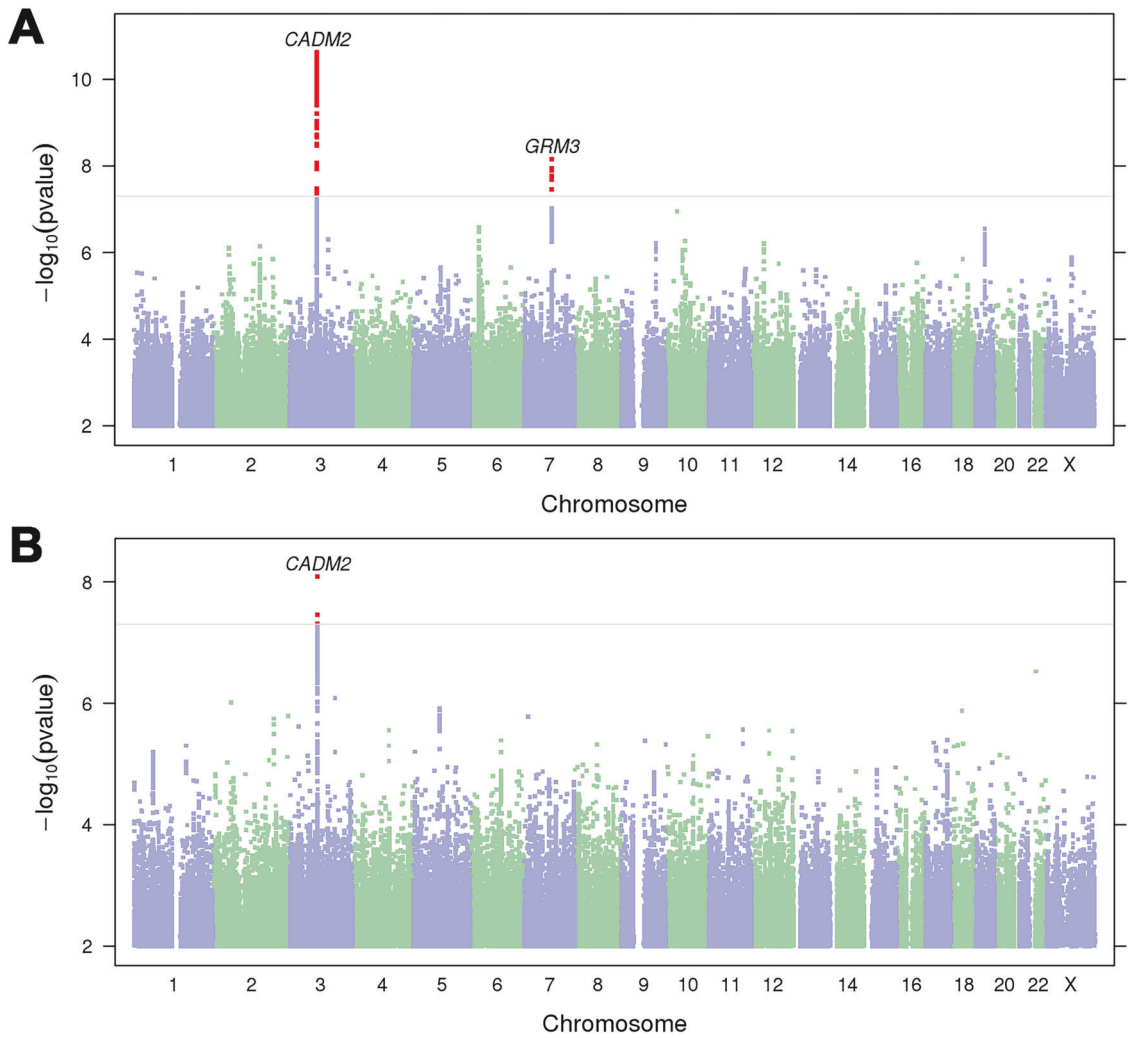


**Fig. 1 GWASs of lifetime and frequency of cannabis use.** Manhattan plots of **A** lifetime cannabis use ($N = 131{,}895$) and **B** frequency of cannabis use ($N = 73{,}374$). The horizontal line represents the significance threshold ($p = 5.00\text{E-}08$). Nearest protein-coding genes (< 1 Mb) to significant loci (red dots) are labelled. For quantile-quantile plots and locus zoom plots, see Supplementary Figs. 2–4.

10). We identified significant genetic correlations between frequency of cannabis use and the psychiatric-related traits "depression possibly related to stressful or traumatic events" ($r_g = -0.54 \pm 0.16$, $p = 9.22\text{E-}04$), stress-related disorder ($r_g = 0.33 \pm 0.10$, $p = 1.44\text{E-}03$), and anxiety/panic attacks ($r_g = -0.38 \pm 0.14$, $p = 6.06\text{E-}03$), though only stress-related disorder was also genetically correlated with lifetime cannabis use ($r_g = 0.25 \pm 0.06$, $p = 3.10\text{E-}05$).

*Externalizing and risk-taking traits.* Among the strongest associations for lifetime cannabis use were positive genetic correlations with externalizing behavior ($r_g = 0.84 \pm 0.03$, $p = 5.65\text{E-}208$), and traits that were used to construct externalizing behavior [70]: the number of sexual partners ($r_g = 0.69 \pm 0.03$, $p = 6.16\text{E-}115$) and the age at first sex (reverse-coded; $r_g = 0.60 \pm 0.03$, $p = 1.08\text{E-}83$). We found similar positive genetic correlations with frequency of cannabis use and externalizing ($r_g = 0.45 \pm 0.06$, $p = 1.68\text{E-}15$), and traits that were used to construct externalizing, such as risk-taking ($r_g = 0.24 \pm 0.07$, $p = 5.48\text{E-}4$) and the number of sexual partners ($r_g = 0.42 \pm 0.06$, $p = 3.17\text{E-}12$).

*Cognitive traits.* We identified significant genetic correlations between lifetime cannabis use and 11 cognitive and executive function-related traits; these included positive genetic correlations with delay discounting ($r_g = 0.16 \pm 0.04$, $p = 3.51\text{E-}04$) and other impulsivity-related measures ($r_g = 0.27 \pm 0.05$ to $0.46 \pm 0.05$, $p = 1.02\text{E-}22$ to $3.20\text{E-}04$), and negative genetic correlations with childhood intelligence ($r_g = -0.29 \pm 0.08$, $p = 3.20\text{E-}04$), intelligence ($r_g = -0.12 \pm 0.03$, $p = 3.04\text{E-}05$), educational years ($r_g = -0.17 \pm 0.03$, $p = 1.84\text{E-}07$), and executive function ($r_g = -0.13 \pm 0.03$, $p = 3.63\text{E-}05$).

For frequency of cannabis use, we identified positive genetic correlations with intelligence ($r_g = 0.40 \pm 0.05$, $p = 4.18\text{E-}14$) and common executive function ($r_g = 0.34 \pm 0.06$, $p = 7.86\text{E-}09$). There was also a negative genetic correlation with delay discounting ($r_g = -0.23 \pm 0.07$, $p = 1.62\text{E-}03$), indicating those who use cannabis more frequently may devalue delayed rewards. Consistent with lifetime cannabis use, we found a positive genetic correlation with the impulsivity-related measure perseverance ($r_g = 0.28 \pm 0.09$, $p = 1.48\text{E-}03$).

*Physical health traits.* We identified mostly positive genetic correlations between lifetime cannabis use and 17 physical health traits, including chronic pain ($r_g = 0.21 \pm 0.04$, $p = 5.59\text{E-}09$), back pain ($r_g = 0.22 \pm 0.05$, $p = 2.19\text{E-}06$), and coronary artery disease with angina ($r_g = 0.17 \pm 0.04$, $p = 2.59\text{E-}05$). For frequency of cannabis use, there was a positive genetic correlation with diabetes ($r_g = 0.20 \pm 0.07$, $p = 5.96\text{E-}03$) and a negative genetic

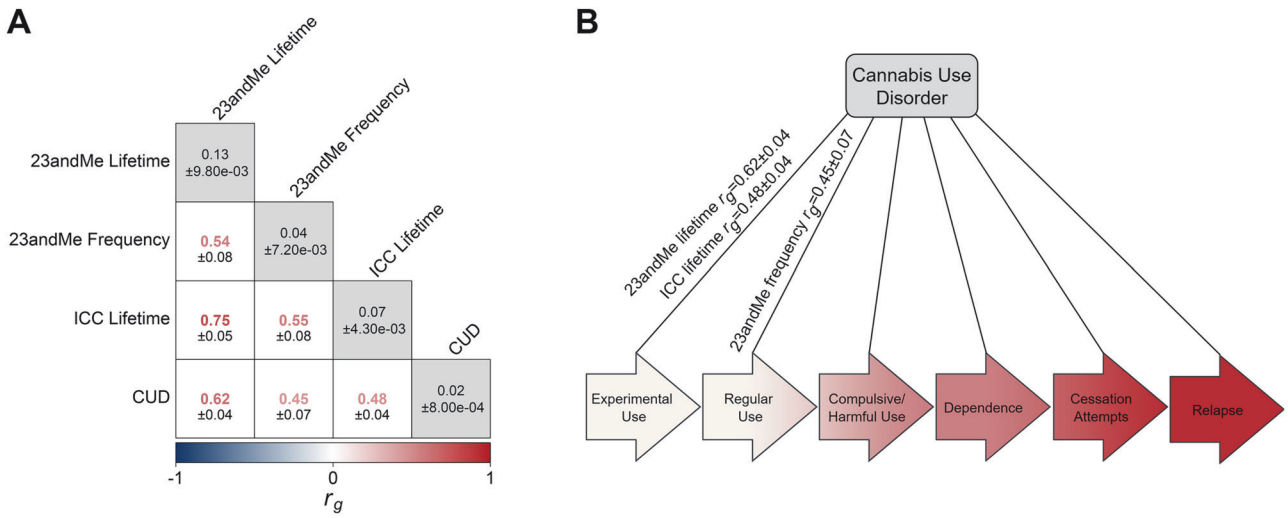


**Fig. 2 SNP-based heritability and genetic correlation analysis comparisons across cannabis-related traits. A** Genetic correlations and $h2_{SNP}$ across 23andMe lifetime cannabis use and frequency of cannabis use with ICC lifetime cannabis use and CUD from Levey et al. [22]. $h2_{SNP}$ ± standard error are shown in matrix diagonal (gray boxes), $r_g$ ± standard error are shown in off-diagonal (white boxes). Correlation coefficients are shown in heatmap color, with $p$ value underneath in black. **B** CUD requires progression through multiple pre-addiction stages, including experimental use, regular use, compulsive/harmful use, dependence, cessation attempts, and relapse. Aside from lifetime cannabis use as a proxy for experimental use and frequency of cannabis use as a proxy for regular use, which positively genetically correlate with CUD, most of these stages have not been genetically explored with GWAS.

correlation with irritable bowel syndrome ($r_g = -0.27 \pm 0.10$, $p = 6.55E-03$).

### Lifetime and frequency of cannabis use PGSs associate with cannabis use phenotypes

Lifetime and frequency PGS associations with cannabis use traits in AoU were considered in single (i.e., models only incorporating lifetime *or* frequency of cannabis use PGS as variables) and joint (i.e., models incorporating lifetime *and* frequency of cannabis use PGS as variables) PGS models (Supplementary Tables 14–16). In the joint-PGS model simultaneously accounting for lifetime and frequency PGSs in the European cohort, based on genetic similarity (see **Methods**), lifetime cannabis use PGS associated with lifetime cannabis use ($\beta = 0.19 \pm 0.01$, $p < 2.00E-16$), daily cannabis use ($\beta = 0.09 \pm 0.03$, $p = 5.09E-04$), and problematic cannabis use ($\beta = 0.22 \pm 0.02$, $p < 2.00E-16$; Table 1, Supplementary Table 16). Frequency of cannabis use PGS was associated with lifetime cannabis use ($\beta = 0.06 \pm 0.01$, $p < 2.00E-16$), and nominally associated with problematic cannabis use ($\beta = 0.06 \pm 0.03$, $p = 0.01$), which did not survive multiple testing correction. Lifetime and frequency PGSs were estimated to explain 0.31–1.52% of the phenotypic variance in cannabis use traits (Fig. 4). In the African cohort, based on genetic similarity (see **Methods**), lifetime cannabis use was predicted by the lifetime PGS ($\beta = 0.08 \pm 0.01$, $p = 2.76E-12$) and the frequency PGS ($\beta = 0.04 \pm 0.01$, $p = 1.88E-04$), which contributed an estimated 0.20% to phenotypic variance. In both populations, phenotypic variance was primarily attributable to the lifetime cannabis use PGS versus the frequency of cannabis use PGS.

In all models, age was a significant negative predictor and being a male was a significant positive predictor of problematic, daily, and lifetime cannabis use (Supplementary Tables 14–17).

### Lifetime cannabis use PGS associates with psychiatric and infectious disease diagnoses

PheWAS uncovered 15 FDR-significant associations and LabWAS uncovered 9 FDR-significant associations with lifetime cannabis use in the BioVU European cohort (Fig. 5; Supplementary Tables 19–20). When CUD was included as a covariate, 8 PheWAS and 6 LabWAS associations remained. Tobacco smoking is prevalent among cannabis users [71, 72]; 4 PheWAS and 4 LabWAS associations persisted when adjusting for TUD, and 1 PheWAS and 5 LabWAS associations persisted when CUD and TUD were jointly included as covariates. We found no significant associations with cannabis use frequency in the European cohort. There were no significant associations for lifetime or frequency of cannabis use in the African cohort.

*Psychiatric disorders*. Our PheWAS identified positive associations between lifetime cannabis use PGS and seven psychiatric disorders: TUD ($\beta = 0.09 \pm 0.01$, $p = 2.44E-15$), substance addiction and disorders ($\beta = 0.14 \pm 0.02$, $p = 8.56E-13$), CUD ($\beta = 0.21 \pm 0.03$, $p = 1.24E-10$), alcohol-related disorders ($\beta = 0.10 \pm 0.02$, $p = 2.43E-05$), mood disorder ($\beta = 0.05 \pm 0.01$, $p = 3.38E-07$), two anxiety traits (anxiety disorders: $\beta = 0.05 \pm 0.01$, $p = 8.85E-06$; anxiety disorder: $\beta = 0.04 \pm 0.01$, $p = 2.55E-04$), depression ($\beta = 0.05 \pm 0.01$, $p = 1.73E-05$), bipolar ($\beta = 0.09 \pm 0.02$, $p = 1.59E-04$), and suicide ideation or attempt ($\beta = 0.12 \pm 0.03$, $p = 2.64E-04$). TUD, substance addiction and disorders, and mood disorders persisted following adjustment for CUD, only substance addiction and disorders persisted following control for TUD, and no psychiatric disorders were significant following control for both CUD and TUD. We did not find evidence of an association with schizophrenia ($\beta = 0.02 \pm 0.06$, $p = 0.68$), schizophrenia and other psychotic disorders ($\beta = 0.03 \pm 0.03$, $p = 0.29$), or psychosis ($\beta = 0.08 \pm 0.04$, $p = 0.07$).

*Infectious diseases*. We found significant positive associations between lifetime cannabis use and infectious diseases, such as human immunodeficiency virus (**HIV**; $\beta = 0.21 \pm 0.04$, $p = 1.14E-07$), symptomatic HIV infection ($\beta = 0.21 \pm 0.04$, $p = 1.26E-07$), and viral hepatitis C ($\beta = 0.13 \pm 0.03$, $p = 3.99E-06$). All associations persisted following control for CUD, and both HIV associations persisted following control for TUD, but no infectious disease associations persisted following control for both CUD and TUD.

*Other diagnoses*. Lifetime cannabis use PGS was negatively associated with one digestive trait, celiac disease ($\beta = -0.34 \pm 0.05$, $p = 1.55E-11$). This association persisted with following control for CUD, TUD, and combined CUD and TUD.

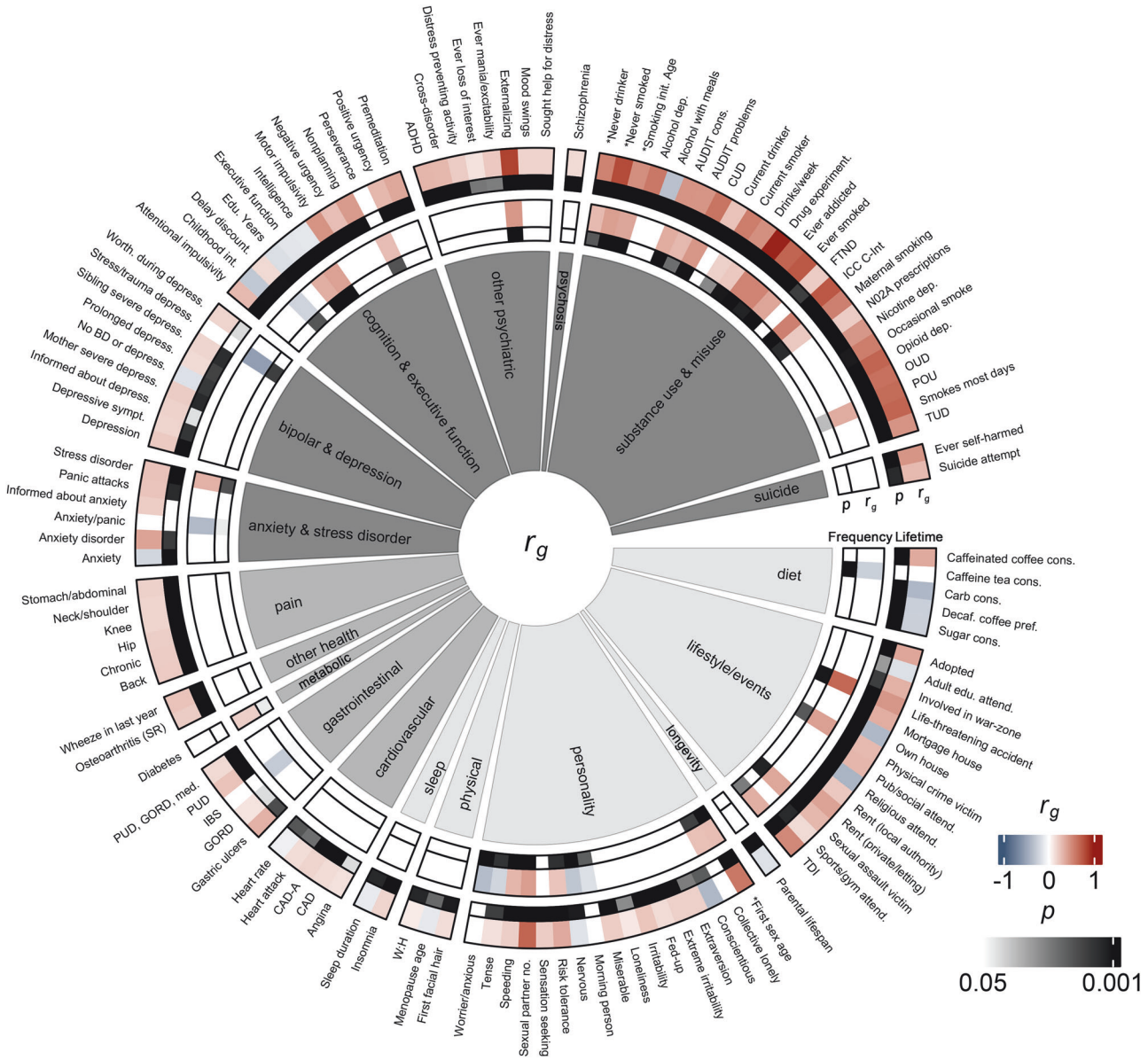


**Fig. 3 Comparison of genetic correlations across anthropometric (light gray), health (medium gray), and psychiatric (dark gray) traits between lifetime cannabis use (lanes 1 and 2) and frequency of cannabis use (lanes 3 and 4).** Lanes 1 and 3 show $r_g$ values calculated by LDSC, and lanes 2 and 4 show FDR-corrected *p* values. Only traits for which at least one cohort was FDR-significant are displayed. For a full list of correlations and trait names, see Supplementary Table 13. *reverse coded traits.

*Blood laboratory biomarkers*. LabWAS revealed associations with lifetime cannabis use across four blood biomarkers: mean corpuscular hemoglobin (**MCH**; $\beta = 0.02 \pm 3.53\text{E-}03$, $p = 1.60\text{E-}07$), carbon dioxide serum/plasma ($\beta = -0.02 \pm 3.47\text{E-}03$, $p = 1.92\text{E-}06$), MCH concentration ($\beta = 0.02 \pm 3.85\text{E-}03$, $p = 9.41\text{E-}05$), and mean corpuscular volume ($\beta = 0.01 \pm 3.53\text{E-}03$, $p = 7.77\text{E-}04$). Following CUD adjustment, all but mean corpuscular volume remained significant; following adjustment for TUD alone or alongside CUD, carbon dioxide serum/plasma and MCH remained significant.

*Immune laboratory biomarkers*. Two immune biomarkers, leukocytes in blood ($\beta = 0.02 \pm 3.51\text{E-}03$, $p = 2.77\text{E-}09$) and complement C4 in serum or plasma ($\beta = 0.06 \pm 0.02$, $p = 6.84\text{E-}05$), were positively associated with lifetime cannabis use. Both remained significant following control with TUD and CUD independently or together.

*Other laboratory biomarkers*. The kidney biomarker creatinine in blood ($\beta = -0.02 \pm 3.90\text{E-}03$, $p = 1.02\text{E-}04$), endocrine biomarker parathyrin intact in serum or plasma ($\beta = -0.04 \pm 0.01$, $p = 1.25\text{E-}03$), and the metabolic biomarker calcitriol in serum and plasma ($\beta = -0.02 \pm 0.01$, $p = 1.37\text{E-}03$) were negatively associated with lifetime cannabis use; none were significant following control for TUD, but creatinine in blood remained significant when CUD, and when CUD and TUD were used as covariates.

## DISCUSSION

This study contributes to the growing body of cannabis use genetics literature by providing new GWASs of 131,895 individuals of European genetic similarity assessed for lifetime cannabis use and, for the first time, 73,374 individuals assessed for frequency of cannabis use. Both GWASs replicated the robust associations with variants nearby *CADM2* and lifetime cannabis use GWAS identified

**Table 1.** Joint-PGS regression analysis associating lifetime cannabis use PGS, frequency of cannabis use PGS, and select covariates with lifetime, daily, and problematic cannabis use in AoU cohorts.

| | European Cohort | | | | | | | | African Cohort | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lifetime Use ($N_{case}$ = 64,711, $N_{control}$ = 49,595) | | | Daily Use ($N_{case}$ = 1411, $N_{control}$ = 28,112) | | | Problematic Use ($N_{case}$ = 1825, $N_{control}$ = 118,704) | | | Lifetime Use ($N_{case}$ = 26,064, $N_{control}$ = 21,610) | | | Daily Use ($N_{case}$ = 2483, $N_{control}$ = 11,718) | | | Problematic Use ($N_{case}$ = 2315, $N_{control}$ = 50,262) | | |
| **Variable** | **β** | **StdErr** | ***p*** | **β** | **StdErr** | ***p*** | **β** | **StdErr** | ***p*** | **β** | **StdErr** | ***p*** | **β** | **StdErr** | ***p*** | **β** | **StdErr** | ***p*** |
| Lifetime PGS | **0.19** | **0.01** | **<2.00E-16** | **0.09** | **0.03** | **5.09E-04** | **0.22** | **0.02** | **<2.00E-16** | **0.08** | **0.01** | **2.76E-12** | −0.02 | 0.03 | 0.52 | 3.62E-03 | 0.02 | 0.88 |
| Frequency PGS | **0.06** | **0.01** | **<2.00E-16** | −0.03 | 0.03 | 0.38 | 0.06 | 0.03 | 0.01 | **0.04** | **0.01** | **1.88E-04** | 0.03 | 0.03 | 0.23 | 0.01 | 0.03 | 0.61 |

Bold PGS results are significant following Bonferroni correction ($p < 8.33E-03$). For full analysis variables, see Supplementary Table 16.

one novel locus near *GRM3*. We found that lifetime and frequency of cannabis use reliably genetically correlated with substance use-related traits, including CUD, and PGSs for both traits associated with cannabis use phenotypes in AoU. Polygenic analysis of lifetime cannabis use also revealed positive associations with substance use and mood disorders consistent with the literature, and novel phenotypic associations with anxiety disorders, infectious diseases, and red blood cell biomarkers. Overall, these results support the value of cannabis use phenotypes spanning the addiction spectrum in the exploration of genetic factors influencing cannabis use vulnerability and health risk.

Lifetime cannabis use captures both experimental/occasional and heavy use; despite the simplicity of this phenotype, we uncovered multiple novel genetic associations with lifetime cannabis use (i.e., *GRM3* locus, genetic correlations, polygenic associations), and found it reliably associated with CUD and multiple other important traits. Although frequency of use may better account for regular cannabis use, this trait did not associate with CUD to a greater degree compared to lifetime cannabis use ($r_g = 0.45 \pm 0.07$ vs. $0.62 \pm 0.04$). This is different from tobacco GWASs, where cigarette consumption was more strongly associated with nicotine dependence than lifetime tobacco use was [41]. This could be attributed to lower power ($N = 73{,}374$ vs. 131,895) or fundamental differences between the two drugs and how they are used. Lifetime and frequency of cannabis use were genetically correlated with each other and their associations with other complex traits were almost always directionally consistent. We previously demonstrated that consumption and problematic use phenotypes (i.e., alcohol [24, 73, 74], tobacco [41]) are correlated but non-identical traits; this is likely true for cannabis. Future multivariate analyses incorporating lifetime, frequency, and other cannabis use GWASs (e.g., CUD, dependence, craving, etc.) could effectively boost locus discovery, identify novel relationships between CUD behaviors and health, and parse genomic factors pertaining to the stages of CUD [36], as we and others have previously demonstrated for other substance use traits [23, 73–77].

One of our most notable findings was a novel association between lifetime cannabis use and rs12673181, which is located upstream of the *GRM3* gene that encodes the group II inhibitory glutamate receptor $mGlu_3$. There are no known associations with this or other *GRM3* SNPs with cannabis-related traits, and while GWASs implicate *GRM3* variants in other substance use (i.e., alcohol, smoking) [78], schizophrenia [79–83], neuroticism [84, 85], educational attainment [86], and other phenotypes [87–89], those associations are with SNPs that are not in LD with rs12673181. Recent studies suggest that $mGlu_3$ potentiates activity of $mGlu_5$ [90], which has also garnered attention for its potential role in addictive-like behaviors and endocannabinoid synthesis [91, 92]. While rs12673181 lies upstream of *GRM3*, it is not a known eQTL of *GRM3* (Supplementary Table 5) [93]. Further functional work, especially pertaining to the regulation of *GRM3*, is required to better understand rs12673181's association with cannabis use vulnerability.

Through multiple lines of evidence, we found lifetime and frequency of cannabis use associated with the *CADM2* gene, replicating prior GWASs of lifetime cannabis use and CUD [23, 34]. Other GWASs have found an association between SNPs in *CADM2* and other substance use traits [23, 42, 46, 78, 94–109], risk-taking [94, 103, 107, 110–112], impulsivity [42], and externalizing behaviors [70].

Supporting the genetic correlation observed across cannabis GWAS data, PGSs for lifetime and, to a lesser degree, frequency of cannabis use, associated with phenotypes across the CUD progression spectrum (i.e., lifetime, daily, and problematic use). More variance was explained by lifetime (0.29–1.40%) rather than frequency of use PGS (0.12–0.19%), and together they explained up to 1.6% of phenotypic variance. This is on par with recent

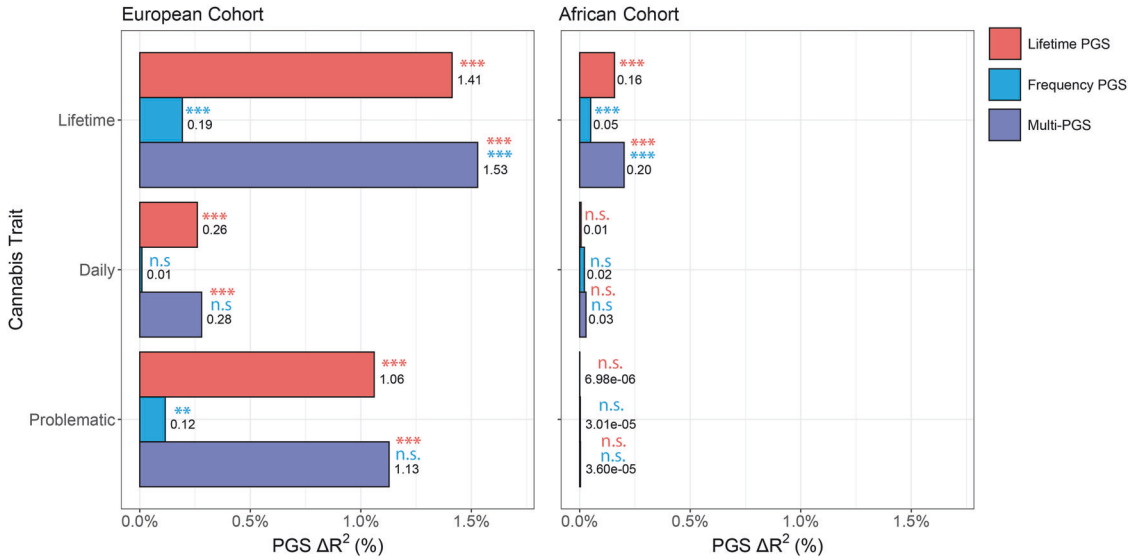


**Fig. 4 Percent proportion of lifetime, daily, and problematic cannabis use variance attributable to lifetime cannabis use PGS, frequency of cannabis use PGS, or both (joint-PGS) in European and African AoU cohorts.** Bonferroni-corrected significance of PGS contribution for single- and joint-PGS models (see Table 1, Supplementary Tables 15–16) are shown above data label in its corresponding legend color (n.s. $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

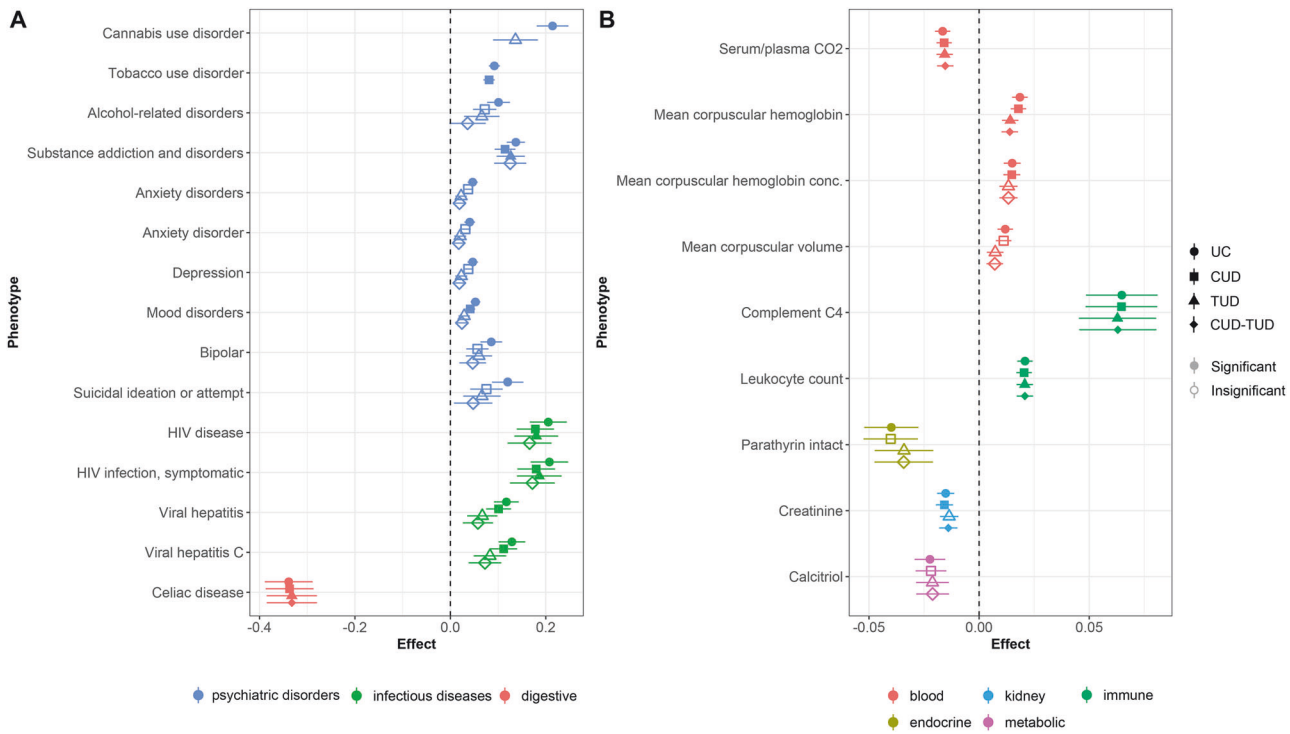


**Fig. 5 Forest plot of FDR-significant phenome associations with unconditioned (UC) lifetime cannabis use PGS, or with adjustment for cannabis use disorder (CUD), tobacco use disorder (TUD), or both (CUD-TUD). A** PheWAS results. **B** LabWAS results. For full trait information, see Supplementary Tables 19–20.

substance use PGS analyses [113–117], including by Hodgson and colleagues [118], who estimated that ICC lifetime cannabis use PGS predicted 0.82% of the variance in lifetime cannabis use and 1.2% of the variance in continued cannabis use in UK Biobank participants. Although it is improbable that cannabis use PGS alone will provide much clinical utility [119], lifetime and frequency of cannabis use PGS could be useful for models predicting the risk for problematic cannabis use.

Largely consistent with the genetic correlations we observed, PheWAS uncovered positive associations between lifetime cannabis use PGS with substance use, depression, anxiety, bipolar, and suicidality in the BioVU cohort ($N < 66{,}917$). To our knowledge, the positive associations with HIV and hepatitis diagnoses, negative association with celiac disease, and mixed associations with multiple blood and immune laboratory biomarkers are novel. Our findings complement a recent PheWAS conducted in the Yale-

Penn sample ($N < 10{,}610$), which is a cohort deeply phenotyped for psychiatric disorder diagnoses and related diagnostic criteria. That study found ICC lifetime cannabis use PGS positively associated with CUD, as well as traits related to other substance use (e.g., alcohol, tobacco, sedatives, stimulants) and depression [120]. That many of these relationships disappear when controlling for CUD in our PheWAS and in an independent PheWAS study [120], as well as when controlling for TUD in our study, supports the hypothesis that these associations are mediated by regular cannabis and tobacco use rather than genetic liability for lifetime cannabis use. Furthermore, like other recent studies [120], we found minimal evidence of a relationship between lifetime cannabis use genetics, schizophrenia, and psychosis (aside from bipolar), despite the genetic relationship between cannabis use and psychosis being the subject of intense interest [55, 121–123] following observations of their apparent bidirectional phenotypic relationship [124]. Epidemiological evidence supports a link between heavy or high potency cannabis use with psychosis [125–127]. Identifying more robust variant associations, especially for frequency of cannabis use, will aid future causal inference analyses that can help to resolve the role of cannabis genetics in health.

Our results were consistent with the prior cannabis use GWAS by ICC [34]. Lifetime cannabis use measured in US-based 23andMe research participants was genetically correlated with the same trait examined in the ICC cohort, which is composed of participants across North America, Europe, and Australia [34]. Both lifetime cannabis use datasets were genetically correlated with CUD, but the magnitude of this association was stronger in the 23andMe dataset compared to ICC ($r_g = 0.62$ vs. 0.48) despite our smaller sample size. Heritability estimates for our lifetime cannabis use trait was also higher (12.88 vs. 6.63%). Heritability may decrease when meta-analyzing cohorts, possibly due to cohort-specific environmental/geocultural differences that could exist surrounding cannabis use [128–130]. Furthermore, while we found consistent positive correlations with psychiatric disorders, including schizophrenia [21, 34, 131–134], attention-deficit hyperactivity disorder [21, 34, 131, 135], bipolar disorder [34, 131], and depression [21, 131] between 23andMe and ICC lifetime cannabis use, we also observed that the genetic correlation with educational attainment was negative with 23andMe and positive with ICC lifetime cannabis use [34]. Interestingly, while most genetic correlations between lifetime and frequency of cannabis use were also mostly in agreement, lifetime cannabis use negatively genetically correlated with intelligence and common executive function and positively genetically correlated with delay discounting, while we saw the inverse with frequency of use. This is not entirely unprecedented, as relationships between cannabis use and cognitive traits can be paradoxical, especially among those with psychiatric disorders, such as those with psychosis who use cannabis exhibiting greater cognitive abilities than those who do not [136]. In sum, although most associations were consistent, the differences we observed in trait heritability and patterns of genetic correlations suggest some disunity between 23andMe and ICC lifetime cannabis use cohorts, as well as lifetime and frequency of cannabis use data, which will warrant careful consideration before attempting to meta-analyzing GWAS data.

There are several limitations to our study. The legal status of cannabis use differs across countries and even US states, and has been in flux over the last several decades. Thus, for some of our older subjects, both lifetime and frequency of use could be reflecting use decades ago, whereas younger subjects are referencing more recent use. Most studies suggest that legalizing recreational cannabis use increases lifetime and frequency of use rates [137], which may have impacted our findings in complex ways that depend on which location a given participant was in at the time of their use. In addition, frequency of cannabis use was measured by the number of use days over a 30-day window, which may not accurately reflect lifetime use intensity because it does not account for the duration of regular use or use quantity. These characteristics are important to CUD trajectory and other health and wellbeing relationships [138–141]. Lifetime and frequency of cannabis use GWASs also relied on self-reported data. Cannabis use is most common during adolescence and young adulthood [142], but participants in this study averaged in their 50s and could have been at greater risk for recall bias regarding cannabis use in early life [143]. Socioeconomic variables are also associated with cannabis use rates [144, 145], and the on-average higher socioeconomic status of 23andMe research participants may have influenced our findings [36]. Finally, GWASs were conducted using genomic information from individuals of genetically predicted European ancestry. While we extended our polygenic analyses to African cohorts, cross-population transferability of PGS is suboptimal compared to investigations where discovery and target populations are ancestrally aligned [146, 147]. This, along with lower sample numbers, may explain why we observed fewer associations in African versus European cohorts. Due to sample size constraints, we also did not explore associations in other ancestral groups, further limiting the generalizability of our results.

This project showcases the utility of pre-addiction phenotypes in cannabis use genomic discovery. Lifetime and frequency of cannabis use genetically associated with CUD and other SUDs, alongside concerning health and psychiatric problems. Increasing sample size and investigating other heritable, diverse phenotypes (e.g., drug responsivity, craving, withdrawal; Fig. 2B) will be integral to further our understanding of CUD vulnerability and the health consequences of cannabis use.

## DATA AVAILABILITY

We provide 23andMe summary statistics for the top 10,000 independent SNPs. 23andMe GWAS summary statistics will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit https://research.23andme.com/collaborate/#dataset-access/ for more information and to apply to access the data. We will share the Jupyter notebooks used for PGS analysis in AoU with registered All of Us researchers upon request.

## REFERENCES

1. United Nations Office on Drug and Crime. Booklet 3 - drug market trends of cannabis and opioids. Vienna; 2022.
2. Pacek LR, Mauro PM, Martins SS. Perceived risk of regular cannabis use in the United States from 2002 to 2012: differences by sex, age, and race/ethnicity. Drug Alcohol Depend. 2015;149:232–44.
3. Statistics Canada. Mental and substance use disorders in Canada. 2013. (Catalogue no.82-624-X).
4. Rotermann M. Looking back from 2020, how cannabis use and related behaviours changed in Canada. Health Rep. 2021;32:3–14.
5. Whiting PF, Wolff RF, Deshpande S, Di Nisio M, Duffy S, Hernandez AV, et al. Cannabinoids for medical use: a systematic review and meta-analysis. JAMA. 2015;313:2456–73.
6. Bellocchio L, Inchingolo AD, Inchingolo AM, Lorusso F, Malcangi G, Santacroce L, et al. Cannabinoids drugs and oral health-from recreational side-effects to medicinal purposes: a systematic review. Int J Mol Sci. 2021;22:8329.
7. Diep C, Tian C, Vachhani K, Won C, Wijeysundera DN, Clarke H, et al. Recent cannabis use and nightly sleep duration in adults: a population analysis of the NHANES from 2005 to 2018. Reg Anesth Pain Med. 2022;47:100–4.
8. Hasin DS, Kerridge BT, Saha TD, Huang B, Pickering R, Smith SM, et al. Prevalence and correlates of DSM-5 cannabis use disorder, 2012–2013: findings from the national epidemiologic survey on alcohol and related Conditions-III. Am J Psychiatry. 2016;173:588–99.
9. Hayley AC, Stough C, Downey LA. DSM-5 cannabis use disorder, substance use and DSM-5 specific substance-use disorders: evaluating comorbidity in a population-based sample. Eur Neuropsychopharmacol. 2017;27:732–43.
10. Howard J, Osborne J. Cannabis and work: need for more research. Am J Ind Med. 2020;63:963–72.

11. Keen L 2nd, Turner AD, George L, Lawrence K. Cannabis use disorder severity and sleep quality among undergraduates attending a Historically Black University. Addict Behav. 2022;134:107414.
12. Lo JO, Hedges JC, Girardi G. Impact of cannabinoids on pregnancy, reproductive health, and offspring outcomes. Am J Obstet Gynecol. 2022;227:571–81.
13. Pacek LR, Herrmann ES, Smith MT, Vandrey R. Sleep continuity, architecture and quality among treatment-seeking cannabis users: an in-home, unattended polysomnographic study. Exp Clin Psychopharmacol. 2017;25:295–302.
14. Page RL 2nd, Allen LA, Kloner RA, Carriker CR, Martel C, et al. Medical marijuana, recreational cannabis, and cardiovascular health: a scientific statement from the American Heart Association. Circulation. 2020;142:e131–52.
15. Feingold D, Livne O, Rehm J, Lev-Ran S. Probability and correlates of transition from cannabis use to DSM-5 cannabis use disorder: results from a large-scale nationally representative study. Drug Alcohol Rev. 2020;39:142–51.
16. American Psychiatric Association. Substance-related and addictive disorders. Diagnostic and statistical manual of mental disorders, 5th edn. Washington, D.C.; 2013. https://doi.org/10.1176/appi.books.9780890425596.dsm16.
17. Hines LA, Morley KI, Rijsdijk F, Strang J, Agrawal A, Nelson EC, et al. Overlap of heritable influences between cannabis use disorder, frequency of use and opportunity to use cannabis: trivariate twin modelling and implications for genetic design. Psychol Med. 2018;48:2786–93.
18. Kendler KS, Ohlsson H, Maes HH, Sundquist K, Lichtenstein P, Sundquist J. A population-based Swedish Twin and Sibling Study of cannabis, stimulant and sedative abuse in men. Drug Alcohol Depend. 2015;149:49–54.
19. Verweij KJ, Zietsch BP, Lynskey MT, Medland SE, Neale MC, Martin NG, et al. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. Addiction. 2010;105:417–30.
20. Demontis D, Rajagopal VM, Thorgeirsson TE, Als TD, Grove J, Leppala K, et al. Genome-wide association study implicates CHRNA2 in cannabis use disorder. Nat Neurosci. 2019;22:1066–74.
21. Johnson EC, Demontis D, Thorgeirsson TE, Walters RK, Polimanti R, Hatoum AS, et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. Lancet Psychiatry. 2020;7:1032–45.
22. Levey DF, Galimberti M, Deak JD, Wendt FR, Bhattacharya A, Koller D, et al. Multi-ancestry genome-wide association study of cannabis use disorder yields insight into disease biology and public health implications. Nat Genet. 2023;55:2094–103.
23. Xu H, Toikumo S, Crist RC, Glogowska K, Jinwala Z, Deak JD, et al. Identifying genetic loci and phenomic associations of substance use traits: a multi-trait analysis of GWAS (MTAG) study. Addiction. 2023;118:1942–52.
24. Sanchez-Roige S, Palmer AA, Clarke TK. Recent efforts to dissect the genetic basis of alcohol use and abuse. Biol Psychiatry. 2020;87:609–18.
25. McLellan AT, Koob GF, Volkow ND. Preaddiction-A missing concept for treating substance use disorders. JAMA Psychiatry. 2022;79:749–51.
26. Agrawal A, Madden PA, Bucholz KK, Heath AC, Lynskey MT. Initial reactions to tobacco and cannabis smoking: a Twin study. Addiction. 2014;109:663–71.
27. Degenhardt L, Coffey C, Carlin JB, Swift W, Moore E, Patton GC. Outcomes of occasional cannabis use in adolescence: 10-year follow-up study in Victoria, Australia. Br J Psychiatry. 2010;196:290–5.
28. Scherrer JF, Grant JD, Duncan AE, Sartor CE, Haber JR, Jacob T, et al. Subjective effects to cannabis are associated with use, abuse and dependence after adjusting for genetic and environmental influences. Drug Alcohol Depend. 2009;105:76–82.
29. Swift W, Coffey C, Carlin JB, Degenhardt L, Patton GC. Adolescent cannabis users at 24 years: trajectories to regular weekly use and dependence in young adulthood. Addiction. 2008;103:1361–70.
30. Windle M, Wiesner M. Trajectories of marijuana use from adolescence to young adulthood: predictors and outcomes. Dev Psychopathol. 2004;16:1007–27.
31. Lyons MJ, Toomey R, Meyer JM, Green AI, Eisen SA, Goldberg J, et al. How do genes influence marijuana use? The role of subjective effects. Addiction. 1997;92:409–17.
32. Leung J, Chan GCK, Hides L, Hall WD. What is the prevalence and risk of cannabis use disorders among people who use cannabis? A systematic review and meta-analysis. Addict Behav. 2020;109:106479.
33. Haberstick BC, Zeiger JS, Corley RP, Hopfer CJ, Stallings MC, Rhee SH, et al. Common and drug-specific genetic influences on subjective effects to alcohol, tobacco and marijuana use. Addiction. 2011;106:215–24.
34. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat Neurosci. 2018;21:1161–70.
35. Stringer S, Minica CC, Verweij KJ, Mbarek H, Bernard M, Derringer J, et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. Transl Psychiatry. 2016;6:e769.
36. Sanchez-Roige S, Palmer AA. Emerging phenotyping strategies will advance our understanding of psychiatric genetics. Nat Neurosci. 2020;23:475–80.
37. Thorpe HHA, Talhat MA, Khokhar JY. High genes: genetic underpinnings of cannabis use phenotypes. Prog Neuropsychopharmacol Biol Psychiatry. 2021;106:110164.
38. McBain RK, Wong EC, Breslau J, Shearer AL, Cefalu MS, Roth E, et al. State medical marijuana laws, cannabis use and cannabis use disorder among adults with elevated psychological distress. Drug Alcohol Depend. 2020;215:108191.
39. Hughes JR, Naud S, Budney AJ, Fingar JR, Callas PW. Attempts to stop or reduce daily cannabis use: an Intensive Natural History study. Psychol Addict Behav. 2016;30:389–97.
40. Mallard TT, Sanchez-Roige S. Dimensional phenotypes in psychiatric genetics: lessons from genome-wide association studies of alcohol use phenotypes. Complex Psychiatry. 2021;7:45–8.
41. Toikumo S, Jennings MV, Pham BK, Lee H, Mallard TT, Bianchi SB, et al. Multi-ancestry meta-analysis of tobacco use disorder identifies 461 potential risk genes and reveals associations with multiple health outcomes. Nat Human Behav. 2024;8:1177–93.
42. Sanchez-Roige S, Jennings MV, Thorpe HHA, Mallari JE, van der Werf LC, Bianchi SB, et al. CADM2 is implicated in impulsive personality and numerous other traits by genome- and phenome-wide association studies in humans and mice. Transl Psychiatry. 2023;13:167.
43. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet. 2015;96:37–53.
44. National Academies of Sciences, Engineering, and Medicine Health and Medicine Division, Division of Behavioral and Social Sciences and Education, Board on Health Sciences Policy, Committee on population, committee on the use of race, ethnicity, and ancestry as population descriptors in genomics research. Using population descriptors in genetics and genomics research: a new framework for an evolving field. Washington (DC): National Academies Press (US); 2023.
45. Durand EY, Do CB, Mountain JL, Macpherson JM Ancestry composition: a novel, efficient pipeline for ancestry deconvolution. bioRxiv [Preprint]. 2014.
46. Hatoum AS, Colbert SMC, Johnson EC, Huggett SB, Deak JD, Pathak G, et al. Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. Nat Ment Health. 2023;1:210–23.
47. Yin B, Wang X, Huang T, Jia J. Shared genetics and causality between decaffeinated coffee consumption and neuropsychiatric diseases: a large-scale genome-wide cross-trait analysis and mendelian randomization analysis. Front Psychiatry. 2022;13:910432.
48. Agrawal A, Chou YL, Carey CE, Baranger DAA, Zhang B, Sherva R, et al. Genome-wide association study identifies a novel locus for cannabis dependence. Mol Psychiatry. 2018;23:1293–302.
49. Sherva R, Wang Q, Kranzler H, Zhao H, Koesterer R, Herman A, et al. Genome-wide Association Study of cannabis dependence severity, novel risk variants, and shared genetic risks. JAMA Psychiatry. 2016;73:472–80.
50. Agrawal A, Lynskey MT, Hinrichs A, Grucza R, Saccone SF, Krueger R, et al. A genome-wide association study of DSM-IV cannabis dependence. Addict Biol. 2011;16:514–8.
51. Minica CC, Dolan CV, Hottenga JJ, Pool R, Genome of the Netherlands Consortium, Fedko IO, et al. Heritability, SNP- and gene-based analyses of cannabis use initiation and age at onset. Behav Genet. 2015;45:503–13.
52. Zhao Y, Han X, Zheng ZL. Analysis of the brain transcriptome for substance-associated genes: An update on large-scale genome-wide association studies. Addict Biol. 2023;28:e13332.
53. Greco LA, Reay WR, Dayas CV, Cairns MJ. Exploring opportunities for drug repurposing and precision medicine in cannabis use disorder using genetics. Addict Biol. 2023;28:e13313.
54. Carreras-Gallo N, Dwaraka VB, Caceres A, Smith R, Mendez TL, Went H, et al. Impact of tobacco, alcohol, and marijuana on genome-wide DNA methylation and its relationship with hypertension. Epigenetics. 2023;18:2214392.
55. Cheng W, Parker N, Karadag N, Koch E, Hindley G, Icick R, et al. The relationship between cannabis use, schizophrenia, and bipolar disorder: a genetically informed study. Lancet Psychiatry. 2023;10:441–51.
56. Greco LA, Reay WR, Dayas CV, Cairns MJ. Pairwise genetic meta-analyses between schizophrenia and substance dependence phenotypes reveals novel association signals with pharmacological significance. Transl Psychiatry. 2022;12:403.
57. Minica CC, Verweij KJH, van der Most PJ, Mbarek H, Bernard M, van Eijk KR, et al. Genome-wide association meta-analysis of age at first cannabis use. Addiction. 2018;113:2073–86.
58. Sey NYA, Hu B, Mah W, Fauni H, McAfee JC, Rajarajan P, et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by

incorporating brain chromatin interaction profiles. Nat Neurosci. 2020;23:583–93.
59. Barbeira, Dickinson AN, Bonazzola SP, Zheng R, Wheeler HE J, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9:1825.
60. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.
61. All of Us Research Program Genomics Investigators. Genomic data in the All of Us research program. Nature. 2024;627:340–6.
62. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" research program. N Engl J Med. 2019;381:668–76.
63. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019;10:1776.
64. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. Genet Epidemiol. 2012;36:214–24.
65. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008;84:362–9.
66. Dennis J, Sealock J, Levinson RT, Farber-Eger E, Franco J, Fong S, et al. Genetic risk for major depressive disorder and loneliness in sex-specific associations with coronary artery disease. Mol Psychiatry. 2021;26:4254–64.
67. Fogel AI, Akins MR, Krupp AJ, Stagi M, Stein V, Biederer T. SynCAMs organize synapses through heterophilic adhesion. J Neurosci. 2007;27:12516–30.
68. Yan X, Wang Z, Schmidt V, Gauert A, Willnow TE, Heinig M, et al. Cadm2 regulates body weight and energy homeostasis in mice. Mol Metab. 2018;8:180–8.
69. Tyler RE, Besheer J, Joffe ME. Advances in translating mGlu(2) and mGlu(3) receptor selective allosteric modulators as breakthrough treatments for affective disorders and alcohol use disorder. Pharmacol Biochem Behav. 2022;219:173450.
70. Karlsson Linner R, Mallard TT, Barr PB, Sanchez-Roige S, Madole JW, Driver MN, et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. Nat Neurosci. 2021;24:1367–76.
71. Agrawal A, Budney AJ, Lynskey MT. The co-occurring use and misuse of cannabis and tobacco: a review. Addiction. 2012;107:1221–33.
72. Galimberti M, Levey DF, Deak JD, Zhou H, Stein MB, Gelernter J. Genetic influences and causal pathways shared between cannabis use disorder and other substance use traits. Mol Psychiatry. 2024.
73. Mallard TT, Savage JE, Johnson EC, Huang Y, Edwards AC, Hottenga JJ, et al. Item-Level Genome-Wide Association Study of the alcohol use disorders identification test in three population-based cohorts. Am J Psychiatry. 2022;179:58–70.
74. Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun. 2019;10:1499.
75. Zhou H, Kember RL, Deak JD, Xu H, Toikumo S, Yuan K, et al. Multi-ancestry study of the genetics of problematic alcohol use in over 1 million individuals. Nat Med. 2023;29:3184–92.
76. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, 23andMe Research Team, Substance Use Disorder Working Group of the Psychiatric Genomics Consortium. et al. Genome-Wide Association Study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. Am J Psychiatry. 2019;176:107–18.
77. Savage JE, Barr PB, Phung T, Lee YH, Zhang Y, Ge T, et al. Genetic heterogeneity across dimensions of alcohol use behaviors. *medRxiv*. 2023.2012.2026.23300537. 2023.
78. Saunders GRB, Wang X, Chen F, Jang SK, Liu M, Wang C, et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. Nature. 2022;612:720–4.
79. Trubetskoy V, Pardinas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature. 2022;604:502–8.
80. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet. 2019;51:1670–8.
81. Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50:381–9.
82. Li Z, Chen J, Yu H, He L, Xu Y, Zhang D, et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat Genet. 2017;49:1576–83.
83. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.
84. Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. Item-level analyses reveal genetic heterogeneity in neuroticism. Nat Commun. 2018;9:905.
85. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet. 2018;50:6–11.
86. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50:1112–21.
87. Hysi PG, Choquet H, Khawaja AP, Wojciechowski R, Tedja MS, Yin J, et al. Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. Nat Genet. 2020;52:401–7.
88. Tedja MS, Wojciechowski R, Hysi PG, Eriksson N, Furlotte NA, Verhoeven VJM, et al. Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. Nat Genet. 2018;50:834–48.
89. Watanabe K, Jansen PR, Savage JE, Nandakumar P, Wang X, 23andMe Research Team. et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. Nat Genet. 2022;54:1125–32.
90. Di Menna L, Joffe ME, Iacovelli L, Orlando R, Lindsley CW, Mairesse J, et al. Functional partnership between mGlu3 and mGlu5 metabotropic glutamate receptors in the central nervous system. Neuropharmacology. 2018;128:301–13.
91. Ibrahim KS, Abd-Elrahman KS, El Mestikawy S, Ferguson SSG. Targeting vesicular glutamate transporter machinery: implications on metabotropic glutamate receptor 5 signaling and behavior. Mol Pharmacol. 2020;98:314–27.
92. Jung KM, Mangieri R, Stapleton C, Kim J, Fegley D, Wallace M, et al. Stimulation of endocannabinoid formation in brain slice cultures through activation of group I metabotropic glutamate receptors. Mol Pharmacol. 2005;68:1196–202.
93. GTEX portal. https://www.gtexportal.org/home/. Accessed 2024.
94. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault JB, Kutalik Z. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. Nat Hum Behav. 2023;7:1216–27.
95. Deak JD, Levey DF, Wendt FR, Zhou H, Galimberti M, Kranzler HR, et al. Genome-wide investigation of maximum habitual alcohol intake in us veterans in relation to alcohol consumption traits and alcohol use disorder. JAMA Netw Open. 2022;5:e2238880.
96. Pasman JA, Demange PA, Guloksuz S, Willemsen AHM, Abdellaoui A, Ten Have M, et al. Genetic risk for smoking: disentangling interplay between genes and socioeconomic status. Behav Genet. 2022;52:92–107.
97. Xu K, Li B, McGinnis KA, Vickers-Smith R, Dao C, Sun N, et al. Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. Nat Commun. 2020;11:5302.
98. Zhou H, Sealock JM, Sanchez-Roige S, Clarke TK, Levey DF, Cheng Z, et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. Nat Neurosci. 2020;23:809–18.
99. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet. 2020;52:437–47.
100. Evangelou E, Gao H, Chu C, Ntritsos G, Blakeley P, Butts AR, et al. New alcohol-related genes suggest shared genetic mechanisms with neuropsychiatric disorders. Nat Hum Behav. 2019;3:950–61.
101. Zhong VW, Kuang A, Danning RD, Kraft P, van Dam RM, Chasman DI, et al. A genome-wide association study of bitter and sweet beverage consumption. Hum Mol Genet. 2019;28:2449–57.
102. Sanchez-Roige S, Fontanillas P, Elson SL, Gray JC, de Wit H, MacKillop J, et al. Genome-wide association studies of impulsive personality traits (BIS-11 and UPPS-P) and drug experimentation in up to 22,861 adult research participants identify loci in the CACNA1I and CADM2 genes. J Neurosci. 2019;39:2562–72.
103. Karlsson Linner R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, 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. 2019;51:245–57.
104. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, 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. 2019;51:237–44.
105. Erzurumluoglu AM, Liu M, Jackson VE, Barnes DR, Datta G, Melbourne CA, et al. Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. Mol Psychiatry. 2020;25:2392–409.
106. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet. 2019;104:65–75.
107. Clifton EAD, Perry JRB, Imamura F, Lotta LA, Brage S, Forouhi NG, et al. Genome-wide association study for risk taking propensity indicates shared pathways with body mass index. Commun Biol. 2018;1:36.

108. Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N = 112 117). Mol Psychiatry. 2017;22:1376–84.
109. Brazel DM, Jiang Y, Hughey JM, Turcot V, Zhan X, Gong J, et al. Exome chip meta-analysis fine maps causal variants and elucidates the genetic architecture of rare coding variants in smoking and alcohol use. Biol Psychiatry. 2019;85:946–55.
110. Baselmans B, Hammerschlag AR, Noordijk S, Ip H, van der Zee M, de Geus E, et al. The genetic and neural substrates of externalizing behavior. Biol Psychiatry Glob Open Sci. 2022;2:389–99.
111. Strawbridge RJ, Ward J, Cullen B, Tunbridge EM, Hartz S, Bierut L, et al. Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. Transl Psychiatry. 2018;8:39.
112. Mills MC, Tropf FC, Brazel DM, van Zuydam N, Vaez A, eQTLGen Consortium. et al. Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. Nat Hum Behav. 2021;5:1717–30.
113. Johnson EC, Sanchez-Roige S, Acion L, Adams MJ, Bucholz KK, Chan G, et al. Polygenic contributions to alcohol use and alcohol use disorders across population-based and clinically ascertained samples. Psychol Med. 2021;51:1147–56.
114. Sun Y, Chang S, Wang F, Sun H, Ni Z, Yue W, et al. Genome-wide association study of alcohol dependence in male Han Chinese and cross-ethnic polygenic risk score comparison. Transl Psychiatry. 2019;9:249.
115. Lai D, Wetherill L, Bertelsen S, Carey CE, Kamarajan C, Kapoor M, et al. Genome-wide association studies of alcohol dependence, DSM-IV criterion count and individual criteria. Genes Brain Behav. 2019;18:e12579.
116. Chang LH, Couvy-Duchesne B, Liu M, Medland SE, Verhulst B, Benotsch EG, et al. Association between polygenic risk for tobacco or alcohol consumption and liability to licit and illicit substance use in young Australian adults. Drug Alcohol Depend. 2019;197:271–9.
117. Allegrini AG, Verweij KJH, Abdellaoui A, Treur JL, Hottenga JJ, Willemsen G, et al. Genetic vulnerability for smoking and cannabis use: associations with E-cigarette and water pipe use. Nicotine Tob Res. 2019;21:723–30.
118. Hodgson K, Coleman JRI, Hagenaars SP, Purves KL, Glanville K, Choi SW, et al. Cannabis use, depression and self-harm: phenotypic and genetic relationships. Addiction. 2020;115:482–92.
119. Smeland OB, Andreassen OA. Polygenic risk scores in psychiatry - large potential but still limited clinical utility. Eur Neuropsychopharmacol. 2021;51:68–70.
120. Kember RL, Hartwell EE, Xu H, Rotenberg J, Almasy L, Zhou H, et al. Phenome-wide association analysis of substance use disorders in a deeply phenotyped sample. Biol Psychiatry. 2023;93:536–45.
121. Johnson EC, Hatoum AS, Deak JD, Polimanti R, Murray RM, Edenberg HJ, et al. The relationship between cannabis and schizophrenia: a genetically informed perspective. Addiction. 2021;116:3227–34.
122. Reginsson GW, Ingason A, Euesden J, Bjornsdottir G, Olafsson S, Sigurdsson E, et al. Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. Addict Biol. 2018;23:485–92.
123. Johnson EC, Austin-Zimmerman I, Thorpe HHA, Levey DF, Baranger DAA, Colbert SMC et al. Cross-ancestry genetic investigation of schizophrenia, cannabis use disorder, and tobacco smoking. Neuropsychopharmacology. 2024;49:1655–65.
124. Hamilton I, Monaghan M. Cannabis and psychosis: are we any closer to understanding the relationship? Curr Psychiatry Rep. 2019;21:48.
125. Di Forti M, Quattrone D, Freeman TP, Tripoli G, Gayer-Anderson C, Quigley H, et al. The contribution of cannabis use to variation in the incidence of psychotic disorder across Europe (EU-GEI): a multicentre case-control study. Lancet Psychiatry. 2019;6:427–36.
126. Oluwoye O, Monroe-DeVita M, Burduli E, Chwastiak L, McPherson S, McClellan JM, et al. Impact of tobacco, alcohol and cannabis use on treatment outcomes among patients experiencing first episode psychosis: data from the national RAISE-ETP study. Early Interv Psychiatry. 2019;13:142–6.
127. Schoeler T, Monk A, Sami MB, Klamerus E, Foglia E, Brown R, et al. Continued versus discontinued cannabis use in patients with psychosis: a systematic review and meta-analysis. Lancet Psychiatry. 2016;3:215–25.
128. Durvasula A, Price AL Distinct explanations underlie gene-environment interactions in the UK Biobank. *medRxiv* 2023.
129. Abdellaoui A, Dolan CV, Verweij KJH, Nivard MG. Gene-environment correlations across geographic regions affect genome-wide association studies. Nat Genet. 2022;54:1345–54.
130. Thorpe HHA, Fontanillas P, Pham BK, Meredith JJ, Jennings MV, Courchesne-Krak NS, et al. Genome-wide association studies of coffee intake in UK/US participants of European ancestry uncover cohort-specific genetic associations. Neuropsychopharmacology. 2024;49:1609–18.
131. Jang SK, Saunders G, Liu M, andMe Research T, Jiang Y, Liu DJ, et al. Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. Psychol Med. 2022;52:968–78.
132. Chang LH, Ong JS, An J, Verweij KJH, Vink JM, Pasman J, et al. Investigating the genetic and causal relationship between initiation or use of alcohol, caffeine, cannabis and nicotine. Drug Alcohol Depend. 2020;210:107966.
133. Abdellaoui A, Smit DJA, van den Brink W, Denys D, Verweij KJH. Genomic relationships across psychiatric disorders including substance use disorders. Drug Alcohol Depend. 2021;220:108535.
134. Waldman ID, Poore HE, Luningham JM, Yang J. Testing structural models of psychopathology at the genomic level. World Psychiatry. 2020;19:350–9.
135. Poore HE, Hatoum A, Mallard TT, Sanchez-Roige S, Waldman ID, Palmer AA, et al. A multivariate approach to understanding the genetic overlap between externalizing phenotypes and substance use disorders. Addict Biol. 2023;28:e13319.
136. Kayir H, Ruffolo J, McCunn P, Khokhar JY. The relationship between cannabis, cognition, and schizophrenia: it's complicated. Curr Top Behav Neurosci. 2023;63:437–61.
137. Farrelly KN, Wardell JD, Marsden E, Scarfe ML, Najdzionek P, Turna J, et al. The impact of recreational cannabis legalization on cannabis use and associated outcomes: a systematic review. Subst Abuse. 2023;17:11782218231172054.
138. Fergusson DM, Boden JM, Horwood LJ. Psychosocial sequelae of cannabis use and implications for policy: findings from the Christchurch Health and Development Study. Soc Psychiatry Psychiatr Epidemiol. 2015;50:1317–26.
139. van der Pol P, Liebregts N, de Graaf R, Korf DJ, van den Brink W, van Laar M. Predicting the transition from frequent cannabis use to cannabis dependence: a three-year prospective study. Drug Alcohol Depend. 2013;133:352–9.
140. Walden N, Earleywine M. How high: quantity as a predictor of cannabis-related problems. Harm Reduct J. 2008;5:20.
141. Zeisser C, Thompson K, Stockwell T, Duff C, Chow C, Vallance K, et al. A 'standard joint'? The role of quantity in predicting cannabis-related problems. Addiction Research & Theory. 2012;20:82–92.
142. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2022 National Survey on Drug Use and Health. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2022.
143. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211–7.
144. Jeffers AM, Glantz S, Byers A, Keyhani S. Sociodemographic characteristics associated with and prevalence and frequency of cannabis use among adults in the US. JAMA Netw Open. 2021;4:e2136571.
145. Karriker-Jaffe KJ. Neighborhood socioeconomic status and substance use by U.S. adults. Drug Alcohol Depend. 2013;133:212–21.
146. Atkinson EG, Bianchi SB, Ye GY, Martinez-Magana JJ, Tietz GE, Montalvo-Ortiz JL, et al. Cross-ancestry genomic research: time to close the gap. Neuropsychopharmacology. 2022;47:1737–8.
147. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51:584–91.

## ACKNOWLEDGEMENTS

HHAT is funded by a Canadian Institute of Health Research (CIHR) Postdoctoral Fellowship (#491556). JYK is funded by a CIHR Canada Research Chair in Translational Neuropsychopharmacology. PF, the 23andMe Research Team, and SLE are employed by 23andMe, Inc. AAP is supported by NIDA (P50DA037844 and P30DA060810). SSR is funded through the National Institute on Drug Abuse (NIDA DP1DA054394) and the Tobacco-Related Disease Research Program (T32IR5226). We would like to thank the research participants and employees of 23andMe for making this work possible. Participants provided informed consent and volunteered to participate in the research online, under a protocol approved by the external AAHRPP-accredited Institutional Review Board (**IRB**), Ethical & Independent (**E&I**) Review Services. As of 2022, E&I Review Services is part of Salus IRB (https://www.versiticlinicaltrials.org/salusirb). The following members of the 23andMe Research Team contributed to this study: Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Katarzyna Bryc, Emily Bullis, Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Teresa Filshtein, Kipper Fletez-Brant, Will Freyman, Karl Heilbron, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Ethan M. Jewett, Yunxuan Jiang, Katelyn Kukar, Keng-Han Lin, Maya Lowe, Jey C. McCreight, Matthew H. McIntyre, Steven J. Micheletti, Meghan E. Moreno, Joanna L. Mountain, Priyanka Nandakumar, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Morgan Schumacher, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, Alejandro Hernandez, Corinna Wong, Christophe Toukam Tchakouté. We would also like to thank The Externalizing Consortium for sharing the GWAS summary statistics of externalizing. The Externalizing Consortium: Principal Investigators: Danielle M. Dick, Philipp Koellinger, K. Paige Harden, Abraham A. Palmer. Lead Analysts: Richard Karlsson Linnér, Travis T. Mallard, Peter B. Barr, Sandra Sanchez-

Roige. Significant Contributors: Irwin D. Waldman. The Externalizing Consortium has been supported by the National Institute on Alcohol Abuse and Alcoholism (R01AA015416 -administrative supplement), and the National Institute on Drug Abuse (R01DA050721). Additional funding for investigator effort has been provided by K02AA018755, U10AA008401, P50AA022537, as well as a European Research Council Consolidator Grant (647648 EdGe to Koellinger). The content is solely the responsibility of the authors and does not necessarily represent the official views of the above funding bodies. The Externalizing Consortium would like to thank the following groups for making the research possible: 23andMe, Add Health, Vanderbilt University Medical Center's BioVU, Collaborative Study on the Genetics of Alcoholism (COGA), the Psychiatric Genomics Consortium's Substance Use Disorders working group, UK10K Consortium, UK Biobank, and Philadelphia Neurodevelopmental Cohort. We gratefully acknowledge *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us* Research Program for making available the participant data examined in this study. The *All of Us* Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276.

## AUTHOR CONTRIBUTIONS

SSR and AAP conceived the idea. PF and SLE contributed formal analyses and curation of 23andMe data. HHAT contributed to formal analyses, investigation, and data visualization. contributed to formal data analysis and data visualization. JJM, MVJ, RBC, and SP contributed to formal analyses. HHAT and SSR wrote the manuscript. HHAT, PF, JJM, MVJ, RBC, SP, SLE, JYK, LKD, ECJ, AAP and SSR reviewed and edited the manuscript.

## THE 23ANDME RESEARCH TEAM

Pierre Fontanillas[2] and Sarah L. Elson[2]

## COMPETING INTERESTS

PF, the 23andMe Research Team, and SLE were employed by 23andMe, Inc. PF and SLE hold stock or stock options in 23andMe, Inc. The remaining authors have nothing to disclose.

## ETHICS STATEMENT

All methods were performed in accordance with the relevant guidelines and regulations. Participants provided informed consent and volunteered to participate in research online under a protocol approved by the external AAHRPP-accredited Institutional Review Board (**IRB**), Ethical & Independent (**E&I**) Review Services.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41380-025-03219-2.

**Correspondence** and requests for materials should be addressed to Sandra Sanchez-Roige.