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

Hayley H A Thorpe, PhD1, Pierre Fontanillas, PhD2, John J Meredith, BSc3, Mariela V Jennings, BSc3, Renata B Cupertino, PhD3, Shreya Pakala, BSc3, 23andMe Research Team2, Sarah L Elson, PhD2, Jibran Y Khokhar, PhD1, Lea K Davis4,5,6, PhD, Emma C Johnson, PhD7, Abraham A Palmer, PhD3,8, & Sandra Sanchez-Roige, PhD3,5,8*

1Department of Anatomy and Cell Biology, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada.
223andMe, Inc., Sunnyvale, CA, USA
3Department of Psychiatry, University of California San Diego, La Jolla, CA, USA
4Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
5Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
6Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
7Department of Psychiatry, Washington University School of Medicine, St. Louis, MO USA
8Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

Corresponding Author:
Sandra Sanchez-Roige, PhD
Department of Psychiatry, University of California San Diego
9500 Gilman Drive, La Jolla, CA, USA, 92093
Phone: +1-619-874-1426
Email: sanchezroige@ucsd.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Summary

Cannabis is one of the most widely used drugs globally. Decriminalization of cannabis is further increasing cannabis consumption. We performed genome-wide association studies (GWASs) of lifetime ($N=131,895$) and frequency ($N=73,374$) of cannabis use. Lifetime cannabis use GWAS identified two loci, one near CADM2 (rs11922956, $p=2.40\times10^{-11}$) and another near GRM3 (rs12673181, $p=6.90\times10^{-9}$). Frequency of use GWAS identified one locus near CADM2 (rs4856591, $p=8.10\times10^{-9}$; $r^2=0.76$ with rs11922956). Both traits were heritable 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 associated cannabis use phenotypes in AllOfUs participants. Phenome-wide association study of lifetime cannabis use PGS in a hospital cohort replicated associations with substance use and mood disorders, and uncovered associations with celiac and infectious diseases. This work demonstrates the value of GWASs of CUD transition risk factors.

Keywords: genome-wide association study, cannabis, addiction, genetic correlations, polygenic score, phenome-wide association study, CADM2, GRM3
Introduction

Approximately 209 million people globally reported using cannabis 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\(^6-14\). Up to 27% of those who use cannabis in their lifetime are estimated to 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, and thus of clinical interest to identify what makes an individual vulnerable to cannabis use and its negative effects.

Problematic cannabis use is estimated to be 51-78% heritable based on twin studies\(^17-19\) and recent genome-wide association studies (GWASs) have implicated hundreds of loci associated with CUD\(^20-23\). While CUD GWASs are of paramount importance, they come with three major caveats. First, these studies only examine one extreme of the addiction spectrum and neglect other substance-related behaviors and stages between substance initiation to substance use disorder (SUD) diagnosis (e.g., recreational use, escalating intake, dependence)\(^24\). These pre-addiction phenotypes\(^25\) are thought to dictate an individual’s progression to SUD\(^26-32\) and are heritable\(^17,26,31,33\). However, aside from GWASs of lifetime cannabis use (having ever versus never used cannabis in one’s lifetime) using data from the International Cannabis Consortium (ICC) and other sources\(^34,35\), which represents the opposing end of the addiction spectrum to CUD, the genetics of other pre-addiction cannabis 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 harmful cannabis use and its relationships with physical and mental health. Third, curating case/control SUD GWASs are costly and laborious because they require individual
psychological assessments for both cases and controls. Pre-addiction phenotypes can be rapidly and inexpensively collected in large population-based cohorts via self-report questionnaires\(^40\).

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 in their heaviest use period \((N=73,374)\) as a measure of cannabis use frequency. We performed GWASs of lifetime and frequency of cannabis use, followed by a battery of secondary analyses to compare biological, genetic, and phenotypic associations. Because the frequency of cannabis use phenotype better distinguished between light and heavy use, we hypothesized that the genetics of frequency of cannabis use would more closely resemble CUD compared to lifetime cannabis use genetics.
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 52.8±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 single nucleotide polymorphisms (SNPs) quality control, see Supplementary Table 3. Genomic control inflation factors for lifetime cannabis use (λ=1.08) and frequency of cannabis use (λ=1.03) suggested no substantial inflation due to population stratification for either GWAS. SNP-based heritability (\(h^2_{SNP}\)) was 12.88±0.97 for lifetime cannabis use, greater than the \(h^2_{SNP}\) for lifetime cannabis use from the ICC (\(h^2_{SNP}=6.63±0.43\))\(^{34}\). \(h^2_{SNP}\) for frequency of cannabis use was 4.12±0.72 (Supplementary Table 4).

We identified two genome-wide significant (\(p<5.00E-08\)) loci for lifetime cannabis use on chromosomes 3 and 7 (Fig 1A, Supplementary Fig. 2-3, Supplementary Table 5). The most significant association was with rs11922956 (\(p=2.40E-11\), 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 \(^{41,42}\). We also found a novel association between lifetime cannabis use and rs12673181 (\(p=6.90E-09\), chr7q21.11), which is a SNP upstream of Metabotropic glutamate receptor 3 gene (GRM3) encoding mGlu3. mGlu3 is an inhibitory group II receptor affecting a range of intracellular signaling cascades and cellular processes like glutamate neurotransmission and long-term plasticity \(^{43}\).
Frequency of cannabis use GWAS identified one significant association with rs4856591 ($p=8.10E-09$, chr3p12.1; **Fig1B, Supplementary Fig. 2, 5**), which is near to *CADM2* and is in linkage disequilibrium (LD) with rs11922956 ($r^2=0.76$, $p<1.00E-04$).
Figure 1. 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=5.00E-08). Nearest protein-coding genes (<1Mb) to significant loci (red dots) are labelled. For quantile-quantile plots and locus zoom plots, see Supplementary Fig. 2-4.
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 unique genes associated with lifetime cannabis use (Supplementary Tables 6-8), and 4 unique 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). Of the 44 unique genes associated with lifetime and frequency of cannabis use, 29 gene associations have not been previously associated with any cannabis-related trait (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

There were 115 traits genetically correlated ($r_g$) with lifetime cannabis use and 38 with frequency of cannabis use after applying a 5% false discovery rate (FDR) correction (Fig. 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 exceptions for intelligence and executive function (positively genetically correlated with frequency of use, negatively genetically correlated with lifetime use), and tense/‘highly strung’ and delay discounting (negatively genetically correlated with frequency of use, positively genetically correlated with lifetime use), as we review below (Supplementary Fig. 5).

Cannabis and Other Substance Use Traits. The genetic correlation between lifetime and frequency of cannabis use was moderate ($r_g=0.54\pm0.08$, $p=1.89E-10$), suggesting imperfect
genetic overlap between the two traits. We identified positive genetic correlations between CUD and lifetime ($r_g=0.62\pm0.04$, $p=2.44\times10^{-59}$), as well as frequency of cannabis use ($r_g=0.45\pm0.07$, $p=2.45\times10^{-10}$; Fig. 2). Compared to lifetime cannabis use from the ICC, our lifetime cannabis use trait was more strongly genetically correlated with CUD (23andMe-CUD $r_g=0.62\pm0.04$, $p=2.44\times10^{-59}$ vs. ICC-CUD $r_g=0.48\pm0.04$, $p=4.30\times10^{-33}$). Positive genetic correlations with other aspects of substance use (e.g., drug experimentation and lifetime cannabis use: $r_g=0.97\pm0.01$, $p=1.35\times10^{-161}$; frequency: $r_g=0.54\pm0.07$, $p=5.45\times10^{-14}$) and misuse (e.g., Alcohol Use Disorder Identification Test (AUDIT) problems and lifetime cannabis use: $r_g=0.46\pm0.06$, $p=1.54\times10^{-15}$; frequency of cannabis use: $r_g=0.30\pm0.10$, $p=2.46\times10^{-03}$) were among the top genetic correlations for lifetime and frequency of cannabis use (Fig. 3, Supplementary Table 13).
Figure 2. SNP-based heritability and genetic correlation analysis comparisons across cannabis-related traits. 

A) Genetic correlations and $h^2_{SNP}$ across 23andMe lifetime cannabis use and frequency of cannabis use with ICC lifetime cannabis use and CUD from Levey et al. $h^2_{SNP}$ ± standard error shown in matrix diagonal (gray boxes), $r_g$ ± standard error in off-diagonal (white boxes). Correlation coefficients 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.
Psychiatric Disorders. Lifetime cannabis use was genetically correlated with schizophrenia \( (r_g=0.15\pm0.03, \ p=7.33\times10^{-7}) \); however, frequency of cannabis use was not \( (r_g=0.02\pm0.05, \ p=0.73) \). We also identified associations with other psychiatric traits and lifetime cannabis use like ADHD \( (r_g=0.31\pm0.05, \ p=5.20\times10^{-12}) \), depression \( (r_g=0.22\pm0.04, \ p=3.52\times10^{-10}) \), and cross-disorder \( (r_g=0.30\pm0.05, \ p=3.91\times10^{-10}) \). We identified significant genetic correlations between frequency of cannabis use and the psychiatric related traits “depression possibly related to stressful or traumatic events” \( (r_g=-0.54\pm0.16, \ p=9.22\times10^{-4}) \), stress-related disorder \( (r_g=0.33\pm0.10, \ p=1.44\times10^{-3}) \), and anxiety/panic attacks \( (r_g=-0.38\pm0.14, \ p=6.06\times10^{-3}) \), though only stress-related disorder was also genetically correlated with lifetime cannabis use \( (r_g=0.25\pm0.06, \ p=1.42\times10^{-4}) \).

Externalizing and Risk-Taking Traits. Among the strongest associations for lifetime cannabis use were positive genetic correlations with externalizing behavior \( (r_g=0.84\pm0.03, \ p=5.65\times10^{-208}) \), and two traits that were used to construct externalizing behavior\(^{44}\): number of sexual partners \( (r_g=0.69\pm0.03, \ p=6.16\times10^{-115}) \) and age at first sex (reverse-coded; \( r_g=0.60\pm0.03, \ p=1.08\times10^{-83}) \). We found similar positive genetic correlations with frequency of cannabis use and externalizing and risk-taking (externalizing: \( r_g=0.45\pm0.06, \ p=1.68\times10^{-15}) \); number of sexual partners: \( r_g=0.42\pm0.06, \ p=3.17\times10^{-12}) \).

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 \( (r_g=0.16\pm0.04, \ p=3.51\times10^{-4}) \) and other impulsivity-related measures \( (r_g=0.27\pm0.05 \text{ to } 0.46\pm0.05, \ p=1.02\times10^{-22} \text{ to } 3.20\times10^{-4}) \), and negative genetic correlations with childhood intelligence \( (r_g=-0.29\pm0.08, \ p=3.20\times10^{-4}) \), educational years \( (r_g=-0.17\pm0.03, \ p=1.84\times10^{-7}) \), common executive function \( (r_g=-0.13\pm0.03, \ p=3.63\times10^{-5}) \), and intelligence \( (r_g=-0.12\pm0.03, \ p=3.04\times10^{-5}) \).
For frequency of cannabis use, we identified positive genetic correlations with intelligence ($r_g=0.40\pm0.05$, $p=4.18\times10^{-14}$) and common executive function ($r_g=0.34\pm0.06$, $p=7.86\times10^{-9}$). There was also a negative genetic correlation with delay discounting ($r_g=-0.23\pm0.07$, $p=1.62\times10^{-3}$), indicating those who use cannabis more frequently may devalue delayed rewards. Consistent with lifetime cannabis use, we found positive genetic correlation with the impulsivity-related measure perseverance ($r_g=0.28\pm0.09$, $p=1.48\times10^{-3}$).

**Physical Health Traits.** We identified genetic correlations between lifetime cannabis use and 17 physical health traits, including chronic pain ($r_g=0.21\pm0.04$, $p=5.59\times10^{-9}$), back pain ($r_g=0.22\pm0.05$, $p=2.19\times10^{-6}$), and coronary artery disease with angina ($r_g=0.17\pm0.04$, $p=2.59\times10^{-5}$). For frequency of cannabis use, there was a positive genetic correlation with diabetes ($r_g=0.20\pm0.07$, $p=5.96\times10^{-3}$) and a negative genetic correlation with irritable bowel syndrome ($r_g=-0.27\pm0.10$, $p=6.55\times10^{-3}$).
**Figure 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.
Lifetime and Frequency of Cannabis Use Polygenic Scores Associate with Cannabis Use Phenotypes

Lifetime and frequency PGS associations with cannabis use traits in All of Us (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 PGS in the European cohort, based on genetic similarity (see Methods), lifetime cannabis use PGS associated with lifetime cannabis use (β=0.19±0.01, p<2.00E-16), daily cannabis use (β=0.09±0.03, p=5.09E-04), and problematic cannabis use (β=0.22±0.02, p<2.00E-16; Table 1, Supplementary Table 16). Frequency of cannabis use PGS was associated with lifetime cannabis use (β=0.06±0.01, p<2.00E-16), and nominally associated with problematic cannabis use (β=0.06±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 (β=0.08±0.01, p=2.76E-12) and the frequency PGS (β=0.04±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).
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. Bold PGS results are significant following Bonferroni correction \((p<8.33E-03)\). For full analysis variables, see Supplementary Table 16.

| Variable  | European Cohort | | | | African Cohort | | | |
|-----------|-----------------|---|---|---|---|---|---|---|---|
|           | Lifetime Use \((N_{case}=64,711,\ N_{control}=49,595)\) | Daily Use \((N_{case}=1,411,\ N_{control}=28,112)\) | Problematic Use \((N_{case}=1,825,\ N_{control}=118,704)\) | Lifetime Use \((N_{case}=26,064,\ N_{control}=21,610)\) | Daily Use \((N_{case}=2,483,\ N_{control}=11,718)\) | Problematic Use \((N_{case}=2,315,\ N_{control}=50,262)\) |
| 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 |
Figure 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) shown above data label in its corresponding legend color (n.s. $p>0.05$, *$p<0.05$, **$p<0.01$, ***$p<0.001$).
Lifetime Cannabis Use Polygenic Score Associates with Psychiatric and Infectious Disease

Diagnoses

Our phenome- and laboratory-wide association studies (PheWAS/LabWAS) uncovered 15 FDR-significant PheWAS associations and 9 FDR-significant LabWAS associations with lifetime cannabis use in the BioVU European cohort, as described below (Fig. 5; Supplementary Tables 19-20). When CUD was included as a covariate, 8 PheWAS and 6 LabWAS associations remained, as we discuss below. Tobacco smoking is prevalent among cannabis users\textsuperscript{45}; 4 PheWAS and 4 LabWAS associations persisted when adjusting for tobacco use disorder (TUD), and 1 PheWAS and 5 LabWAS associations persisted when CUD and TUD were jointly included as covariates, described further below. We found no significant associations with cannabis use frequency in the European cohort, nor any 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 (β=0.09±0.01, \( p=2.44\times10^{-15} \)), substance addiction and disorders (β=0.14±0.02, \( p=8.56\times10^{-13} \)), CUD (β=0.21±0.03, \( p=1.24\times10^{-10} \)), alcohol-related disorders (β=0.10±0.02, \( p=2.43\times10^{-5} \)), mood disorder (β=0.05±0.01, \( p=3.38\times10^{-7} \)), two anxiety traits (anxiety disorders: β=0.05±0.01, \( p=8.85\times10^{-6} \); anxiety disorder: β=0.04±0.01, \( p=2.55\times10^{-4} \)), depression (β=0.05±0.01, \( p=1.73\times10^{-5} \)), bipolar (β=0.09±0.02, \( p=1.59\times10^{-4} \)), and suicide ideation or attempt (β=0.12±0.03, \( p=2.64\times10^{-4} \)). 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 (β=0.02±0.06, \( p=0.68 \)), schizophrenia and other psychotic disorders (β=0.03±0.03, \( p=0.29 \)), or psychosis (β=0.08±0.04, \( p=0.07 \)).
Infectious Diseases. We found significant positive associations between lifetime cannabis use and infectious diseases, including human immunodeficiency virus (HIV) disease ($\beta=0.21\pm0.04$, $p=1.14E-07$), symptomatic HIV infection ($\beta=0.21\pm0.04$, $p=1.26E-07$), viral hepatitis C ($\beta=0.13\pm0.03$, $p=3.99E-06$), and viral hepatitis ($\beta=0.11\pm0.03$, $p=6.17E-06$). All associations persisted following control for CUD, 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 ($\beta=-0.34\pm0.05$, $p=1.55E-11$). This association persisted with following control for CUD, TUD, and combined CUD and TUD.

Blood Laboratory Biomarkers. LabWAS revealed associations with lifetime cannabis use across four blood biomarkers: mean corpuscular hemoglobin (MCH; $\beta=0.02\pm3.53E-03$, $p=1.60E-07$), carbon dioxide serum/plasma ($\beta=-0.02\pm3.47E-03$, $p=1.92E-06$), MCH concentration ($\beta=0.02\pm3.85E-03$, $p=9.41E-05$), and mean corpuscular volume ($\beta=0.01\pm3.53E-03$, $p=7.77E-04$). 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 ($\beta=0.02\pm3.51E-03SE$, $p=2.778E-09$) and complement C4 in serum or plasma ($\beta=0.06\pm0.02$, $p=6.84E-05$), 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 ($\beta=-0.02\pm3.90E-03$, $p=1.02E-04$), endocrine biomarker parathyrin intact in serum or plasma ($\beta=-0.04\pm0.01$, $p=1.25E-03$), and the metabolic biomarker calcitriol in serum and plasma ($\beta=-0.02\pm0.01$, $p=1.37E-03$) 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.
Figure 5. Forest plot of FDR-significant phenome associations with lifetime cannabis use PGS unconditioned (UC), 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.
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. We identified 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.

Pre-addiction phenotypes are intended to capture prodromal symptoms of SUD; lifetime and frequency of cannabis use are heritable risk factors for CUD development\textsuperscript{15,27,29,30,32,34}. They can be easily self-reported in large cohorts, making them attractive targets for GWAS. 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 ($r_g=0.45\pm0.07$ vs. $0.62\pm0.04$), potentially due to lower power ($N=73,374$ vs. $131,895$). However, lifetime and frequency of cannabis use was genetically correlated with each other and their associations with other complex traits were almost always directionally consistent. This included positive genetic correlations between lifetime and frequency of use with other...
substance use, misuse, and behavioral traits thought to be substance use risk factors like externalizing, impulsivity, and risk-taking\textsuperscript{21,44,46-50}, consistent with ICC lifetime cannabis use genetic correlations\textsuperscript{34} and reports of a general “addiction risk factor” or externalizing factor accounting for genetic overlap across substances\textsuperscript{23,44,51-53}. We previously demonstrated that consumption and problematic use phenotypes (i.e., alcohol\textsuperscript{24,54,55}, tobacco\textsuperscript{56}) 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\textsuperscript{36}, as we and others have previously demonstrated for other substance use traits\textsuperscript{23,54,55,57-59}.

One of our most notable findings was a novel association between lifetime cannabis use and rs12673181, which is located upstream of the \textit{GRM3} gene that encodes the group II inhibitory glutamate receptor mGlu\textsubscript{3}. There are no known associations with this or other \textit{GRM3} SNPs with cannabis-related traits, and while GWASs implicate \textit{GRM3} variants in other substance use (i.e., alcohol, smoking)\textsuperscript{60}, schizophrenia\textsuperscript{61-65}, neuroticism\textsuperscript{66,67}, educational attainment\textsuperscript{68}, and other phenotypes\textsuperscript{69-71}, these variants are not in LD with rs12673181. Recent studies also suggest that mGlu\textsubscript{3} potentiates activity of mGlu\textsubscript{5}\textsuperscript{72}, which has also garnered attention for its potential role in addictive-like behaviors and endocannabinoid synthesis\textsuperscript{73,74}. While rs12673181 lies upstream of \textit{GRM3}, it is not a known expression quantitative trait locus (eQTL) of \textit{GRM3} (Supplementary Table 5)\textsuperscript{75}. Further functional work, especially pertaining to the regulation of \textit{GRM3}, is required to characterize its association with cannabis use vulnerability.

Through multiple lines of evidence, we found lifetime and frequency of cannabis use associated with the \textit{CADM2} gene, replicating prior GWASs of lifetime cannabis use and CUD\textsuperscript{23,34}. Other GWASs have found an association between SNPs in \textit{CADM2} and other substance use traits\textsuperscript{23,48,51,60,76-91}, risk-taking\textsuperscript{76,85,89,92-94}, impulsivity\textsuperscript{48}, and externalizing behaviors\textsuperscript{44}. 
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 substance use PGS analyses, including by Hodgson et al. who estimated that ICC lifetime cannabis use PGS predicted 0.82% of variance in lifetime cannabis use and 1.2% of variance in continued cannabis use in UK Biobank participants. Although it is improbable that cannabis use PGS alone will be sufficient for clinical utility, lifetime and frequency of cannabis use PGS could be useful for models predicting problematic cannabis use risk.

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. That many of these relationships disappear when controlling for CUD in our PheWAS and in the Kember et al. study, 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 others, 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
Epidemiological evidence supports a link between cannabis and heavy or high potency cannabis use. Identifying more robust variant associations, especially for frequency of cannabis use, will aid future causal inference analyses that can resolve the role of cannabis genetics in health.

Our results were consistent with the prior cannabis use GWAS by ICC. 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. 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. Furthermore, while we found consistent positive correlations with psychiatric disorders, including schizophrenia, ADHD, bipolar disorder, and depression 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. 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. 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 changing 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 others are referencing more recent use. Most studies suggest that legalizing recreational cannabis use increases lifetime and frequency of use rates, 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 determined by the number of use days over a 30-day window does 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. Lifetime and frequency of cannabis use GWASs also relied on self-reported data. Cannabis use is most common during adolescence and young adulthood, 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. Socioeconomic variables are also associated with cannabis use rates, and the on-average higher socioeconomic status of 23andMe research participants may have influenced our findings.

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. 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 SUD, alongside concerning health and psychiatric problems. Increasing sample size and investigating other heritable, diverse phenotypes (e.g., drug responsivity, craving, withdrawal; Figure 2B) will be integral to further our understanding of CUD vulnerability and the health consequences of cannabis use.
Acknowledgements

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.versiticlinaltrials.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.

Data Availability

We will provide 23andMe summary statistics for the top 10,000 SNPs upon publication. 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.
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. All authors reviewed and edited the manuscript.

Funding Sources

SR is funded through the National Institute on Drug Abuse (NIDA DP1DA054394) and the Tobacco-Related Disease Research Program (T32IR5226). 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).

Declaration of Interests

PF, the 23andMe Research Team, and S.L.E. are employed by and hold stock or stock options in 23andMe, Inc. The remaining authors have nothing to disclose.
Methods

Participants and GWASs

Lifetime and frequency of cannabis use GWASs were conducted in male and female 23andMe research participants of European genetic similarity, as previously described. Ancestry falls along a spectrum; individuals were only included in the analysis if they had >97% European ancestry (see Supplementary Methods), as determined through local ancestry analysis. 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.versiticlinaltrials.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 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 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, in current use, is an Illumina Infinium Global Screening Array (~640,000 variants) supplemented with ~50,000 variants of custom content. Samples that failed to reach 98.5% call rate were excluded from the study. 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<0.1% and imputed variants with low imputation quality 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). Unrelated participants categorized as of European ancestry were included in the GWASs (lifetime cannabis use N=131,895; frequency of cannabis use N=73,374; Durand et al., 2014). 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, eQTLs, and chromatin state predicted by ChromHMM. Novel SNPs were identified as those not in LD ($r^2<0.10$) or within ±1Mb of GWAS-significant SNPs uncovered by other GWASs of cannabis use traits (e.g., initiation, age of onset, CUD) sourced from the literature 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 or with start/stop positions within ±1Mb 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 determined following Bonferroni correction ($p<2.53\times10^{-6}$). 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 ($p<9.26\times10^{-4}$ for 54 tissue types; $p<1.67\times10^{-3}$ 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; Sey et al., 2020). H-MAGMA assigns non-coding SNPs to genes based on chromatin interactions from fetal brain, adult brain, midbrain neurons, cortical neurons, iPSC-derived neurons, and iPSC-derived astrocytes datasets (https://github.com/thewonlab/H-MAGMA). Exonic and promoter SNPs were assigned to genes based on physical position$^{139}$. We applied a Bonferroni correction based on the total number of gene-tissue pairs tested ($p<9.42\times10^{-7}$ to $9.45\times10^{-7}$).

**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$^{140}$. 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 ($N_{\text{genes}}=1,619 \text{ to } 9,949$). 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)\textsuperscript{140} from all available tissues. We applied Bonferroni correction for each tissue type \((p<3.09E-05 \text{ to } 5.03E-06)\).

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 \(h^2\text{SNP}\) and genetic correlations\textsuperscript{141}. \(h^2\text{SNP}\) 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% FDR correction to account for multiple testing.

Polygenic score analyses

**PGS of lifetime, daily, and problematic cannabis use in AoU.** We tested the associations between lifetime or frequency of cannabis use PGSs with cannabis traits available for AoU participants clustering within a European or African genetic ancestry panel (for details, see All of Us Research Program Genomics Investigators \textsuperscript{142}). 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\textsuperscript{142,143}. 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 or female participants who had available short-read whole genome sequencing data and applicable cannabis use data. We used 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, 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; Ge et al., 2019). 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 by Lee et al. 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 from 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 from hospital-based cohorts using data from the Vanderbilt University Medical Center (VUMC; IRB #160302, #172020, #190418). The BioVU cohort, a subset of VUMC biobank participants (N=72,821), provided genotyping data and EHR containing clinical data and laboratory-assessed biomarkers. 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 the PRS-CS v1.1.0. For 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 v0.12 R package. 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 TUD (phecode 318) and CUD (see Supplementary Table 12 for CUD ICD codes) as covariates to examine if SUD mediated associations with cannabis PGSs. We calculated the 5% FDR for all associations performed (N=1,405).

For LabWAS, we implemented the pipeline established by Dennis et al. 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).

All results are presented as the mean±standard error unless otherwise specified.


28. Scherrer, J.F., Grant, J.D., Duncan, A.E., Sartor, C.E., Haber, J.R., Jacob, T., and Bucholz, K.K. (2009). Subjective effects to cannabis are associated with use, abuse and
dependence after adjusting for genetic and environmental influences. Drug Alcohol Depend 105, 76-82. 10.1016/j.drugalcdep.2009.06.014.


10.1038/tp.2016.36.


10.1038/s41593-020-0609-7.


43. Tyler, R.E., Besheer, J., and Joffe, M.E. (2022). 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 219, 173450. 10.1016/j.pbb.2022.173450.


polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50, 1112-1121. 10.1038/s41588-018-0147-3.


through activation of group I metabotropic glutamate receptors. Mol Pharmacol 68, 1196-1202. 10.1124/mol.105.013961.


yields genome-wide association signals of low specificity for major depression. Nat Genet 52, 437-447. 10.1038/s41588-020-0594-5.


622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. Mol Psychiatry 25, 2392-2409. 10.1038/s41380-018-0313-0.


reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. Transl Psychiatry 8, 39. 10.1038/s41398-017-0079-1.


Vulnerability for Smoking and Cannabis Use: Associations With E-Cigarette and Water

986 100. Hodgson, K., Coleman, J.R.I., Hagenaars, S.P., Purves, K.L., Glanville, K., Choi, S.W.,
O'Reilly, P., Breen, G., Major Depressive Disorder Working Group of the Psychiatric
Genomics Consortium, and Lewis, C.M. (2020). Cannabis use, depression and self-
10.1111/add.14845.

Large potential but still limited clinical utility. Eur Neuropsychopharmacol 51, 68-70.
10.1016/j.euroneuro.2021.05.007.

Disorders in a Deeply Phenotyped Sample. Biol Psychiatry 93, 536-545.
10.1016/j.biopsych.2022.08.010.

998 103. Cheng, W., Parker, N., Karadag, N., Koch, E., Hindley, G., Icick, R., Shadrin, A.,
use, schizophrenia, and bipolar disorder: a genetically informed study. Lancet Psychiatry

and schizophrenia: a genetically informed perspective. Addiction 116, 3227-3234.
10.1111/add.15534.
53


Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51, 584-591. 10.1038/s41588-019-0379-x.


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 13, 910432. 10.3389/fpsyt.2022.910432.


1127  138. Minica, C.C., Verweij, K.J.H., van der Most, P.J., Mbarek, H., Bernard, M., van Eijk, K.R.,
1129  association meta-analysis of age at first cannabis use. Addiction 113, 2073-2086.
1130  10.1111/add.14368.
1131  139. Sey, N.Y.A., Hu, B., Mah, W., Fauni, H., McAfee, J.C., Rajarajan, P., Brennand, K.J.,
1132  Akbarian, S., and Won, H. (2020). A computational tool (H-MAGMA) for improved
1133  prediction of brain-disorder risk genes by incorporating brain chromatin interaction
1134  profiles. Nat Neurosci 23, 583-593. 10.1038/s41593-020-0603-0.
1137  phenotypic consequences of tissue specific gene expression variation inferred from
1139  141. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia
1140  Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J.,
1143  10.1038/ng.3211.
1144  142. All of Us Research Program Genomics Investigators (2024). Genomic data in the All of
1146  143. All of Us Research Program Investigators, Denny, J.C., Rutter, J.L., Goldstein, D.B.,
1147  Philippakis, A., Smoller, J.W., Jenkins, G., and Dishman, E. (2019). The "All of Us"
1150  Bayesian regression and continuous shrinkage priors. Nat Commun 10, 1776.
1151  10.1038/s41467-019-09718-5.
