Check for updates

ARTICLE   OPEN

Cross-ancestry genetic investigation of schizophrenia, cannabis use disorder, and tobacco smoking

Emma C. Johnson [1]✉, Isabelle Austin-Zimmerman[2], Hayley H. A. Thorpe[3], Daniel F. Levey[4,5], David A. A. Baranger [6], Sarah M. C. Colbert [7], Ditte Demontis [8,9], Jibran Y. Khokhar [3], Lea K. Davis [10,11,12], Howard J. Edenberg [13,14], Marta Di Forti[2], Sandra Sanchez-Roige[10,15], Joel Gelernter [4,5] and Arpana Agrawal [1]

Individuals with schizophrenia frequently experience co-occurring substance use, including tobacco smoking and heavy cannabis use, and substance use disorders. There is interest in understanding the extent to which these relationships are causal, and to what extent shared genetic factors play a role. We explored the relationships between schizophrenia (Scz; European ancestry $N = 161,405$; African ancestry $N = 15,846$), cannabis use disorder (CanUD; European ancestry $N = 886,025$; African ancestry $N = 120,208$), and ever-regular tobacco smoking (Smk; European ancestry $N = 805,431$; African ancestry $N = 24,278$) using the largest available genome-wide studies of these phenotypes in individuals of African and European ancestries. All three phenotypes were positively genetically correlated ($r_gs = 0.17$–0.62). Genetic instrumental variable analyses suggested the presence of shared heritable factors, but evidence for bidirectional causal relationships was also found between all three phenotypes even after correcting for these shared genetic factors. We identified 327 pleiotropic loci with 439 lead SNPs in the European ancestry data, 150 of which were novel (i.e., not genome-wide significant in the original studies). Of these pleiotropic loci, 202 had lead variants which showed convergent effects (i.e., same direction of effect) on Scz, CanUD, and Smk. Genetic variants convergent across all three phenotypes showed strong genetic correlations with risk-taking, executive function, and several mental health conditions. Our results suggest that both shared genetic factors and causal mechanisms may play a role in the relationship between CanUD, Smk, and Scz, but longitudinal, prospective studies are needed to confirm a causal relationship.

INTRODUCTION

Schizophrenia (Scz) is a psychiatric condition with an estimated twin-based heritability of around 80% [1, 2]. Substance use disorders (SUDs) are highly prevalent in individuals with Scz [3]. Of these co-occurring SUDs, the role of cannabis use as a risk factor for Scz and first episode psychosis onset remains a classical "chicken or egg" problem in psychiatry [4].

Some studies have suggested a causal, dose- and age-dependent effect of cannabis use on risk for onset of Scz and other forms of psychosis [5–7]. However, cannabis use and cannabis use disorder (CanUD) are heritable [8] (twin heritability ~50%), and an alternative hypothesis is that shared genetic pathways underlie liability to Scz and cannabis use phenotypes [9, 10]. Genetic correlations from genome-wide association studies (GWAS) have provided support for some genetic commonality (e.g., SNP-r_g (Scz, cannabis use) = 0.25 [11], SNP-r_g (Scz, CanUD) =

0.37 [12]). A recent study identified 27 and 21 genome-wide significant loci contributing to the shared genetic etiology between Scz and cannabis use and CanUD, respectively [13]. However, the identification of shared loci was largely driven by genome-wide significant loci in the Scz GWAS, due to the relative difference in discovery power between the Scz and cannabis GWASs. Furthermore, these prior studies have largely been performed in samples predominantly of European ancestry, limiting the generalizability of these findings.

Horizontal pleiotropy (i.e., genetic variants independently contributing to both CanUD and Scz) *and* vertical pleiotropy (i.e., genetic variants contributing to both traits via a causal path) are not mutually exclusive; both mechanisms may play a role in the co-occurrence of CanUD and Scz. Genetically informed causal inference studies of CanUD and Scz have reached mixed conclusions, with no single direction of causality receiving

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [2]Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [3]Department of Anatomy and Cell Biology, Western University, London, ON, Canada. [4]Division of Human Genetics, Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. [5]Department of Psychiatry, Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA. [6]Department of Psychological and Brain Sciences, Washington University in Saint Louis, St. Louis, MO, USA. [7]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Aarhus, Denmark. [9]Department of Biomedicine and Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark. [10]Department of Medicine, Division of Genetic Medicine, Vanderbilt University, Nashville, TN, USA. [11]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA. [12]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [13]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA. [14]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. [15]Department of Psychiatry, UC San Diego School of Medicine, La Jolla, CA, USA. ✉email: emma.c.johnson@wustl.edu

overwhelming support [14, 15]. Several Mendelian Randomization (MR) analyses have suggested greater support for Scz causing cannabis use and CanUD than the opposite direction [11, 16], while the most recent GWAS of CanUD found a bidirectional causal association between Scz and CanUD [17]. MR involves the use of genetic variants associated with an "exposure" (such as CanUD) as "instruments" to determine the likelihood of a causal relationship between the exposure and an "outcome" (e.g., Scz), relying on the fact that our genotypes are randomly "assigned" at birth, and thus confounders should in theory be randomly distributed across individuals with the effect allele (the genetic instrument) and those without. However, MR is dependent on the strength of the genetic instruments, and these analyses have thus far been limited by relatively under-powered CanUD GWASs with weak instruments. Furthermore, most MR methods assume the absence of horizontal pleiotropy, a strong assumption that is unlikely to be met for complex behavioral traits such as CanUD and Scz.

Few prior genetic studies have attempted to disentangle how nicotine/tobacco use genetics impacts the genetic relationship between Scz and CanUD. Approximately 72% of those with Scz report daily tobacco smoking (while this same report estimated 43% were regular cannabis users [18]), and there is evidence that individuals who smoke tobacco daily are at increased risk of psychosis [19], an earlier age of onset of first psychotic episode [19], and the development of schizophrenia [20]. The prevalence of tobacco use, whether as tobacco cigarettes or consumed with cannabis in certain preparations (e.g., blunts, where tobacco is removed from a cigar and replaced with cannabis, or spliffs, where cannabis and tobacco are rolled together), is also high in individuals with CanUD [21, 22]. Prior studies have reported genetic correlations of tobacco smoking with CanUD (SNP-$r_g = 0.61$ [17]) and Scz (SNP-$r_g = 0.14$ [23]). Despite this, few epidemiologic studies have taken potential genetic sharing into account when reporting evidence for causal relationships between tobacco, cannabis, and Scz [6, 7]. In turn, few genomic studies of cannabis and Scz have considered the role of tobacco [13], despite the frequent co-occurrence of tobacco and cannabis use, especially in Europe [24]. In a prior study, we found that genetic liability for CanUD was positively associated with genetic liability for Scz even when accounting for the genetic components of cannabis ever-use, tobacco smoking, and nicotine dependence [10]. Another study found a causal effect of genetic liability to cannabis use on risk for schizophrenia, and this association was unchanged when accounting for tobacco smoking [15]. Thus, the genetic association between cannabis and Scz appears to be independent of tobacco use genetics to some extent, although the relatively low power of prior CanUD GWAS meant limited conclusions could be drawn from these earlier studies.

Given the significant genetic correlations between CanUD, tobacco smoking, and Scz, the increasing pace of cannabis legalization with emerging increases in CanUD incidence [25], parallel increases in the popularity of nicotine vaping [26], and the consequent potential impact on the course of Scz in those with heavy cannabis and tobacco use [27–31], we investigated the evidence for causal relationships and horizontal pleiotropy between CanUD, tobacco smoking, and Scz. We used the largest genome-wide summary statistics available for Scz [32] (European ancestry $N = 161,405$; African ancestry $N = 15,846$), CanUD [17] (European ancestry $N = 886,025$; African ancestry $N = 120,208$), and ever-regularly smoking tobacco [33] (Smk; European ancestry $N = 805,431$; African ancestry $N = 24,278$) in samples whose genetic ancestry is most similar to those historically from Europe (henceforth referred to as "European ancestry") and samples whose genetic ancestry is most similar to those historically from Africa (henceforth referred to as "African ancestry"). We sought to identify and characterize pleiotropic signals, conduct genetic correlation and causal inference analyses, and explore the

relationships between subsets of pleiotropic loci and a range of mental and physical health traits. We focused on CanUD and Smk (as opposed to cannabis ever-use, or nicotine dependence) as CanUD was the cannabis phenotype with the largest genetic correlation with Scz, there was no available GWAS of cannabis consumption or heaviness of use, and current GWAS of nicotine dependence (relying on the Fagerström Test for Nicotine Dependence [34] (FTND)) have been relatively under-powered compared to Smk.

METHODS

Genome-wide summary statistics
We used summary statistics from the largest available GWAS of each trait: Scz, CanUD and tobacco smoking:

- *Schizophrenia (Scz):* We used data from the most recent Psychiatric Genomics Consortium (PGC) Schizophrenia genome-wide association study (GWAS) meta-analysis of individuals of European ancestry ($N = 161,405$; $N_{cases} = 67,390$) [32]. We also analyzed summary statistics from a GWAS meta-analysis of schizophrenia in African ancestry individuals ($N = 15,846$; $N_{cases} = 7509$), from the Cooperative Studies Program (CSP) #572 and the Genomic Psychiatry Cohort [35].
- *Cannabis use disorder (CanUD):* We used data from Levey et al.'s recent GWAS meta-analysis of cannabis use disorder [17], which combined data from the Million Veteran Program (MVP), the PGC, the Lundbeck Foundation Initiative for Integrative Psychiatric Research, and deCODE Genetics (European ancestry $N = 886,025$; $N_{cases} = 42,281$; African ancestry $N = 120,208$; $N_{cases} = 19,065$).
- *Ever-smoked tobacco regularly (Smk):* We used summary statistics from the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) GWAS of self-reported ever/never regular cigarette smoking (European ancestry $N = 805,431$; $N_{ever} = 393,707$; African ancestry $N = 24,278$; $N_{cases} = 9916$) [33]. We used the publicly available set of summary statistics, which does not include data from 23andMe; the sample sizes reported here reflect that exclusion. This phenotype was measured in a variety of ways in different cohorts (e.g., "Have you smoked over 100 cigarettes over the course of your life?", "Have you ever smoked every day for at least a month?", "Have you ever smoked regularly?").

We also used genome-wide summary statistics for attention deficit hyperactivity disorder (ADHD), bipolar disorder, depression, post-traumatic stress disorder (PTSD), educational attainment, executive function, risk-taking, and the Townsend Deprivation Index (TDI) for follow-up analyses; details are provided in the Supplementary Methods.

Genome-wide genetic correlation analyses
We used linkage disequilibrium score regression [36, 37] (LDSC) to estimate SNP-heritability and pairwise genome-wide genetic correlations (r_g) between Scz, Smk, and CanUD. For the European ancestry summary statistics, we used pre-computed LD scores from the 1000 Genomes Phase 3 European reference panel (available from the LDSC website). For the African ancestry summary statistics, we used pre-computed LD scores from the PanUKBB African ancestry sample (available from https://pan.ukbb.broadinstitute.org/downloads).

Causal inference analyses
We tested for causal relationships between Scz, CanUD, and Smk using CAUSE [38]. Compared to traditional Mendelian Randomization methods, CAUSE has the advantage of accounting for correlated horizontal pleiotropic effects (i.e., a genetic instrument is associated with a confounder which is related to both the exposure and the outcome) as well as uncorrelated horizontal pleiotropy. CAUSE uses a less stringent p-value threshold ($p < 1e-3$) to incorporate data from more variants across the genome. More details are provided in the Supplementary Methods.

We performed additional causal inference analyses, including Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO [39]) analyses to test for horizontal pleiotropy and causal relationships among Scz, CanUD, and Smk. We performed these analyses using the TwoSampleMR R package [40, 41]. More details are provided in the Supplementary Methods. We report the results from the MR-PRESSO global test for horizontal pleiotropy, MR-PRESSO test for causality after

removing outliers for horizontal pleiotropy, MR-Egger, weighted median, inverse variance weighted, simple mode, and weighted mode tests for causality, heterogeneity tests for the inverse variance weighted and MR-Egger tests, and the MR-Egger pleiotropy test.

We only performed these analyses using the European ancestry summary statistics, because the African ancestry summary statistics were relatively under-powered for a causal inference analysis, particularly the Scz summary statistics.

Cross-disorder genome-wide association study meta-analysis

We used 'Association analysis based on SubSETs' (ASSET [42]) to combine the GWAS summary data for CanUD, Smk and Scz (separately by ancestry), using the two-tailed meta-analysis approach to obtain a single cross-disorder association statistic. Unlike traditional meta-analysis approaches, ASSET takes into account SNPs with significant effects on multiple disorders even if the effects on the traits are in opposite directions. We used the LDSC genetic covariance intercept to approximate the degree of sample overlap amongst the studies and included it in the ASSET covariance matrix. Default parameters were applied using the 'h.traits' function. Summary statistics are available to download from https://github.com/WashU-BG/CanUD_Smk_Scz.

We then separated the ASSET results into subsets. Following Lam et al. [43], we use the following notation for each subset: ∩ represents variant subsets with the same directions of effect (+ or −), and | represents variant subsets whose effect sizes are in the opposite direction of those for one versus the other two traits. We defined four subsets: (1) Scz ∩ CanUD ∩ Smk (i.e., a subset with convergent effects across all 3 traits); (2) Scz ∩ CanUD | Smk (i.e., a subset of variants with convergent effects for Scz and CanUD, but divergent effects for Smk); (3) Scz ∩ Smk | CanUD; and (4) CanUD ∩ Smk | Scz.

For each subset, we used FUMA v1.6.1 [44] for annotation and identification of genome-wide significant risk loci and independent lead SNPs. We used the matching ancestry subset of the 1000 Genomes Project Phase 3 [45] reference panel for clumping and annotation of SNPs (e.g., the African ancestry reference panel for our African ancestry cross-disorder summary statistics). More details are provided in the Supplementary Methods.

To perform a cross-ancestry meta-analysis, we used the ancestry-specific one-sided meta-analysis results from ASSET. Unlike the two-tailed approach described above, the one-sided meta-analysis in ASSET is more akin to a traditional meta-analysis and results in one effect size per SNP, regardless of whether the SNP shows divergent directions of effect across traits. We combined the ancestry-specific ASSET results using a sample-size weighted meta-analysis scheme in METAL [46]. As before, we uploaded METAL results to FUMA for clumping and annotation, using the 1000 Genomes Project Phase 3 [45] all ancestries reference panel.

Identification of novel loci

To determine whether the ASSET meta-analysis revealed any novel loci in the European ancestry data that were not genome-wide significant in the original GWAS (CanUD, Smk, Scz), we used the LDLink package [47] in R to identify all LD proxy SNPs ($r^2 > 0.6$) for each of the 439 lead pleiotropic SNPs. We then merged these results with the summary statistics for the original CanUD, Smk, and Scz GWASs to determine whether the locus had been identified as genome-wide significant in any of the original GWASs.

Genetic correlations with other relevant phenotypes

After defining SNP subsets using ASSET, we used GeNetic cOVariance Analyzer (GNOVA [48]) to estimate genetic covariances (ρ_g) and correlations (r_g) between the SNP subsets and several psychiatric disorders and other relevant phenotypes in the European ancestry data. We included ADHD, bipolar disorder, depression, and PTSD, given previously reported genetic correlations with our three primary phenotypes (CanUD, Smk, and Scz). We also included educational attainment, executive function, risk-taking, and a regional measure of material deprivation (the Townsend deprivation index), which have also been correlated with CanUD, Smk, and Scz in previous studies. For all subsets, the effect estimate was aligned with the direction of effect for CanUD, for ease of interpretation. It was unclear how best to weight the estimate for each subset; following the example of Lam et al. [43], we used the largest absolute effect size from the three phenotypes as SNP weights in each subset (flipping the sign of the estimate as necessary, to align with the direction of effect for CanUD).

Polygenic scores of ASSET-derived SNP subsets and associations in BioVU

We sought to explore the relationships between the different SNP subsets and a range of health-related phenotypes, including mental health conditions, using a hypothesis-free approach. To accomplish this, we created polygenic scores for each ASSET-derived SNP subset in the European ancestry subset of the BioVU biobank ($N = 72,225$) [49, 50]. As described above for the genetic correlations, the effect estimate was aligned with the direction of effect for CanUD for all SNP subsets. We used PRS-CS to weight the SNP effect sizes, using the 'auto' function to allow the global shrinkage parameter to be learned from the data. We then used PLINK1.9's --score function to create the polygenic scores in BioVU based on the SNP weights from PRS-CS. We fitted a logistical regression model to each of 1338 case/control phenotypes ("phecodes") to estimate the odds of diagnosis given each PGS. Models were adjusted for sex, median age of the longitudinal electronic health records, and the first 10 PCs. More details are provided in the Supplementary Methods.

Partitioned genetic covariance analyses

We used GNOVA [48] to partition the genetic covariance (ρ_g) between CanUD, Smk, and Scz into salient annotation categories. These included tissue-specific functionality (GenoSkyline-Plus annotations, which are tissue-specific functional regions defined by integrating high-throughput epigenetic annotations from the Roadmap Epigenomics Project) for 7 tissues: brain, cardiovascular, epithelium, gastrointestinal, immune, muscle, and "other" tissues. GNOVA is robust to potential sample overlap between summary statistics. We applied Bonferroni correction for multiple testing across all 3 trait pairs (CanUD ~ Smk, CanUD ~ Scz, and Smk ~ Scz) and 7 tissue types tested (e.g., we corrected for $3 \times 7 = 21$ tests, for an $\alpha = 0.002$.) Substantial enrichment in a particular tissue suggests that the genetic covariance shared between a pair of traits is enriched in the portion of the genome predicted to be functional in that tissue. We only performed these analyses using the European ancestry summary stats, as the annotation data was derived using European ancestry samples.

RESULTS

### SNP-heritability and genome-wide genetic correlations

Schizophrenia (Scz), Cannabis Use Disorder (CanUD), and ever-smoking tobacco regularly (Smk) all showed significant SNP-heritability (liability-scale $h^2_{SNP} = 0.21$, 0.09, and 0.11, respectively) and were significantly genetically correlated in the European ancestry data (Table S1). The magnitude of the genetic correlation between Scz and CanUD ($r_g = 0.37$, SE = 0.02, $p = 2.97e-60$) was statistically greater ($p_{diff} = 6.5e-18$) than the correlation between Scz and Smk ($r_g = 0.17$, SE = 0.02, $p = 6.88e-20$) or between Scz and a measure of nicotine dependence more similar to CanUD, the FTND ($r_g = 0.22$, SE = 0.04, $p = 1.56e-7$; $p_{diff} = 0.002$). This suggests that our choice of ever-regular smoking, rather than the FTND, as a measure of tobacco use was not the reason for the lower genetic correlation.

In the African ancestry data, the largest genetic correlation was between Scz and CanUD ($r_g = 0.61$, SE = 0.14, $p = 1.41e-5$; Table S1). While the genetic correlation between Scz and Smk ($r_g = 0.34$, SE = 0.15, $p = 0.03$) was of greater magnitude than in the European ancestry data, this estimate was not significantly different from zero after accounting for multiple testing, due to the much larger standard error.

Causal inference analyses

Using CAUSE [38], a method that accounts for both correlated and uncorrelated horizontal pleiotropic effects, we found evidence for bidirectional causal relationships between all three phenotypes in the European ancestry data (Fig. 1a, Table S2).

To explicitly test for the presence of horizontal pleiotropy, and to ensure our results were not isolated to a specific method of causal inference, we also performed Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO [39]) analyses in the European ancestry data. The MR-PRESSO global test for horizontal pleiotropy was significant for each pairwise test, and we
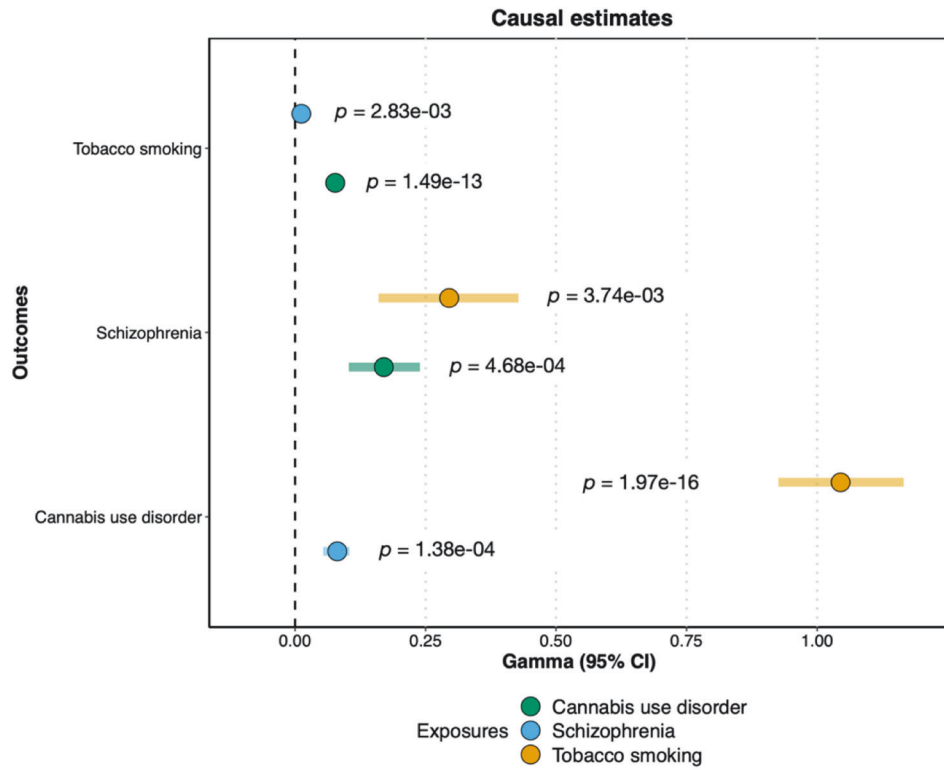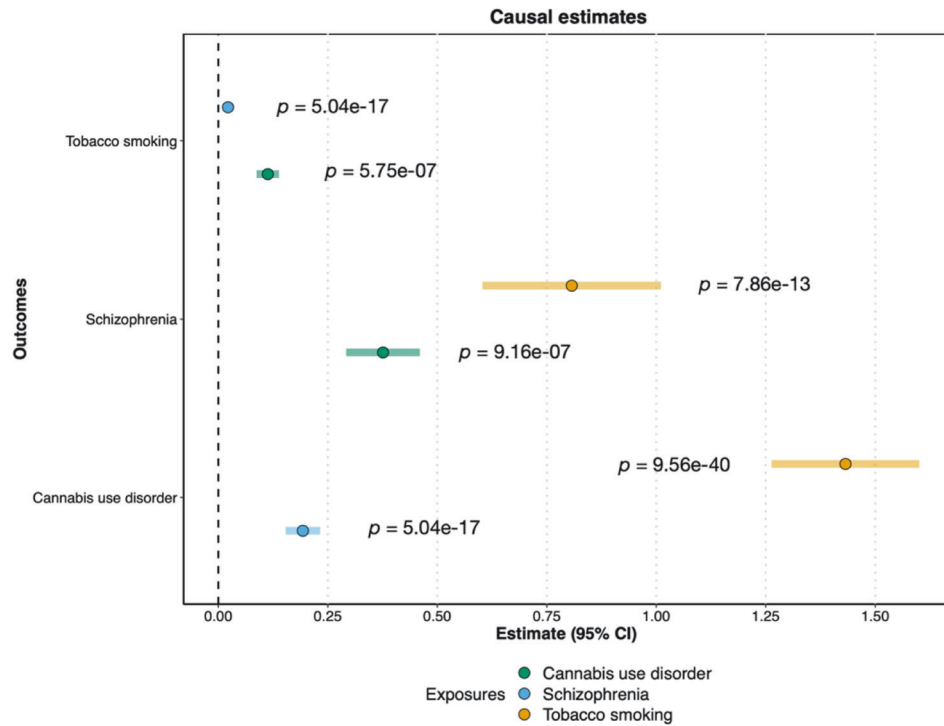
Fig. 1   Causal estimates from CAUSE and MR-PRESSO. A Causal estimates (gamma) and 95% confidence intervals from CAUSE. **B** Causal estimates (beta) and 95% confidence intervals from MR-PRESSO after removal of outliers. "Exposure" phenotypes are indicated by the color, while "Outcome" phenotypes are listed on the *y*-axis.
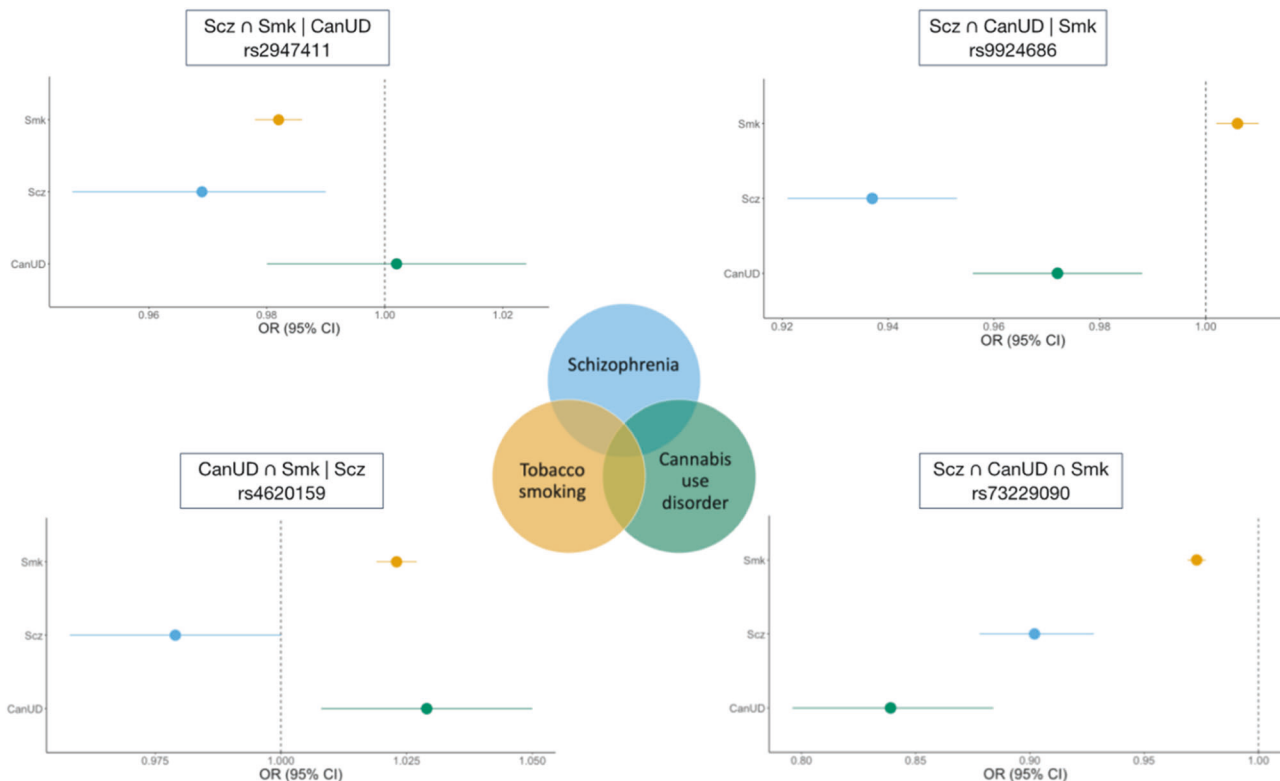
Fig. 2 Example forest plots from the ASSET European ancestry cross-disorder meta-analysis of CanUD, Smk, and Scz. The lower right panel shows lead SNP (rs73229090) in Scz ∩ CanUD ∩ Smk subset. The upper right panel shows SNP (rs9924686) in Scz ∩ CanUD | Smk subset. The upper left panel shows SNP (rs2947411) in Scz ∩ Smk | CanUD subset. The lower left panel shows SNP (rs4620159) in CanUD ∩ Smk | Scz subset.

found significant bidirectional causal effects between all three traits after the removal of outliers for horizontal pleiotropy (Fig. 1b), consistent with the results from CAUSE. Results from other MR methods were generally consistent, with the same direction of effect (Table S3), although the more conservative MR-Egger test [51, 52] only showed a statistically significant causal effect of Smk on CanUD.

Cross-trait loci: European ancestry

In consideration of the significant genetic correlations and evidence for horizontal pleiotropy from MR-PRESSO, we used 'Association analysis based on SubSETs' (ASSET [42]) to combine the GWAS summary data for CanUD, Smk and Scz (separately by ancestry), using the two-tailed meta-analysis approach. Unlike traditional meta-analysis approaches, ASSET accounts for SNPs with significant effects on multiple disorders even if the effects on the traits are in opposite directions. We defined four subsets of variants: (1) Scz ∩ CanUD ∩ Smk (i.e., a subset with convergent effects across all 3 traits); (2) Scz ∩ CanUD | Smk (i.e., a subset of variants with convergent effects for Scz and CanUD, but divergent effects for Smk); (3) Scz ∩ Smk | CanUD; and (4) CanUD ∩ Smk | Scz.

In total, we identified 327 pleiotropic genomic risk loci (i.e., loci where the lead SNP influences all three phenotypes) with 439 lead SNPs. Of these, 150 loci were novel (i.e., not genome-wide significant in any of the original GWAS; see Tables S4 and S5), with 127 of these loci having lead SNP $p \leq 1e-5$ in at least one of the original GWAS, and the remaining 23 having $p \leq 1.4e-4$.

For the subset of SNPs with convergent effects across all 3 traits (Scz ∩ CanUD ∩ Smk) in the European ancestry samples, we identified 202 genomic risk loci with 259 lead SNPs (Table S6). The strongest locus was on chromosome 8, with the top lead SNP being rs73229090 (chr8:27442127, $p = 1.5e-62$; Fig. 2), located in

an intron of the non-coding gene *GULOP*, replicating previous associations with each trait (e.g., [53–55]). This SNP is also an expression quantitative trait locus (eQTL) for *EPHX2* in B cells, tibial artery, esophagus, and cultured fibroblast cells, *CHRNA2* in the cerebellum, and *CCDC25* in the nucleus accumbens.

The Scz ∩ CanUD | Smk subset of SNPs revealed 37 genomic risk loci with 37 lead SNPs (Table S7). The top association was on chromosome 16, with lead SNP rs9924686 (chr16:30003076, $p = 3.3e-15$) within a locus previously implicated by Scz GWAS [53]. This SNP, located in the 3' untranslated region of the serine/threonine-protein kinase gene *TAOK2*, has a CADD score of 18.16, suggesting deleteriousness, and a RegulomeDB score of 1f (eQTL + transcription factor binding/DNase peak), suggesting that this SNP is likely to affect transcription factor binding and linked to expression of a gene target. Furthermore, rs9924686 is an eQTL for several genes, including genes associated with metabolic and immunological traits [56, 57] (*YPEL3* and *INO80E* in adipose tissue and several brain tissues) and alcohol intake [57, 58] (*PPP4C* and *MVP* in cultured cell fibroblasts).

We identified 46 genomic risk loci with 48 lead SNPs for the Scz ∩ Smk | CanUD subset (Table S8). Chromosome 2 had the strongest signal in this subset, with intergenic lead SNP rs2947411 (chr2:614168, $p = 3.6e-19$) that replicates previous associations with Smk [55]. This SNP was an eQTL for only one gene (*SH3YL1* in whole blood).

There were 114 genomic risk loci and 143 lead SNPs for the CanUD ∩ Smk | Scz subset (Table S9). The strongest meta-analytic effect was at lead SNP rs4620159 on chromosome 6 (chr6:111744735, $p = 1.8e-28$); this locus was previously associated with Smk and CanUD [59, 60]. The lead SNP is an intronic variant in *REV3L*, a gene previously associated with smoking and several metabolic traits [23, 57].
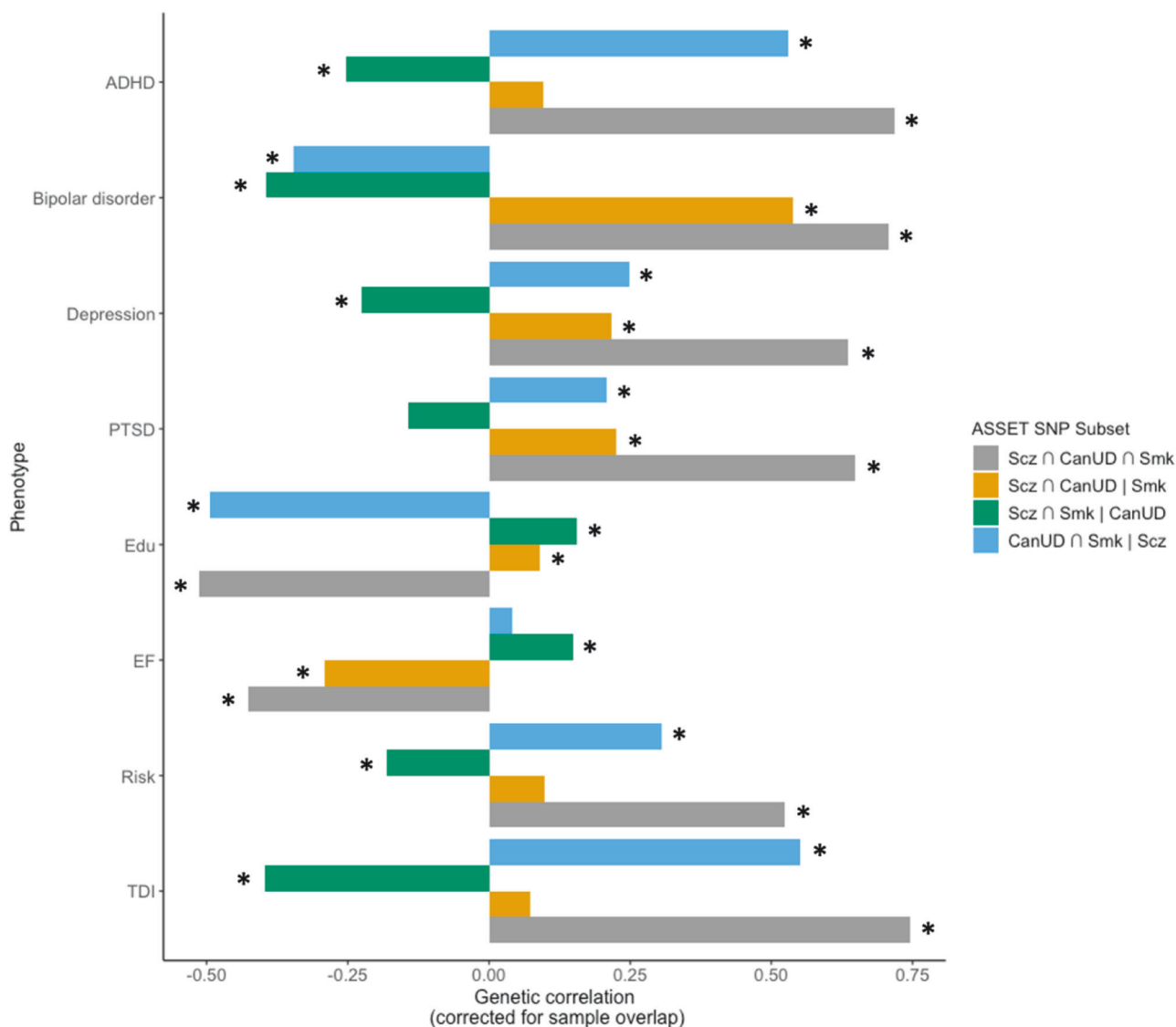
Fig. 3 **Estimated genetic correlations between SNP subsets from ASSET and other psychiatric disorders (attention deficit hyperactivity disorder (ADHD), bipolar disorder, depression, and post-traumatic stress disorder (PTSD)), educational attainment (Edu), executive function (EF), risk-taking (Risk), and Townsend deprivation index (TDI).** Asterisks (*) represent genetic correlations that are statistically significant after Bonferroni correction for 32 tests (*p* < 0.0016).

Cross-trait loci: African ancestry
No associations passed the genome-wide significance threshold (alpha = 5e-8) in the ASSET analysis of the African ancestry data. However, the 14,001 pleiotropic SNPs that were genome-wide significant in the European ancestry data showed smaller *p*-values than expected by chance in the African ancestry data (i.e., the distribution of p-values was significantly left-skewed, with a Kolmogorov–Smirnov goodness-of-fit test indicating significant (*p* < 2e-16) divergence from a distribution of 14,001 randomly sampled SNP p-values). This suggests that with larger sample sizes, future analyses might identify similar loci across both the European and African ancestry datasets.

Cross-ancestry meta-analysis
We performed a sample size-weighted cross-ancestry meta-analysis of the ancestry-specific *one-sided* meta-analysis results from ASSET. Unlike the ancestry-specific two-tailed meta-analyses described above, the one-sided meta-analysis in ASSET is more like a traditional meta-analysis, resulting in one effect size per SNP regardless of whether the SNP shows divergent directions of effect

across traits. The cross-ancestry meta-analysis of CanUD, Smk, and Scz resulted in 448 genome-wide significant risk loci (Table S10).

Genetic associations with other phenotypes
After defining SNP subsets using ASSET, we used GNOVA [48] to estimate genetic correlations between the SNP subsets and four psychiatric disorders (ADHD [61], bipolar disorder [62], depression [63], and PTSD [64]), educational attainment [65] (Edu), executive function [66], risk-taking [58], and the Townsend deprivation index (TDI; a regional measure of deprivation in the UK) in the European ancestry data (Fig. 3, Table S11). All four psychiatric disorders have previously been reported to be genetically correlated with CanUD, Smk, and/or Scz. Edu has previously been shown to be positively correlated with a subset of variants contributing to Scz risk [43], despite negative genetic correlations between Scz and cognitive function [67], and we expected that related socioeconomic status (i.e., TDI), executive function, and risk-taking phenotypes might be differentially associated with SNP subsets. For all subsets, the effect estimate was aligned with the direction of effect for CanUD.

For all phenotypes tested except for bipolar disorder and executive function, the Scz ∩ CanUD ∩ Smk and CanUD ∩ Smk | Scz subsets showed the same direction of genetic correlation, while the Scz ∩ Smk | CanUD subset showed correlations in the opposite direction. In other words, genetic variants with the same direction of effect on CanUD and Smk, regardless of the direction of effect on Scz, showed similar negative genetic correlations with Edu, and positive genetic correlations with psychiatric disorders (except bipolar disorder), risk-taking, and TDI, while genetic variants with the same direction of effect on Scz and Smk but *not* CanUD showed correlations in the opposite direction. Notably, the Scz ∩ CanUD ∩ Smk and Scz ∩ CanUD | Smk subsets were negatively genetically correlated with executive function, while the Scz ∩ Smk | CanUD subset was positively correlated, suggesting a pivotal role of the intersection of CanUD and Scz, regardless of Smk, on executive functioning. The direction of genetic correlation between bipolar disorder and the SNP subsets seemed to depend on Scz (unsurprising, given the high genetic correlation between these two disorders): when the effect estimate was aligned with Scz (the Scz ∩ CanUD ∩ Smk and Scz ∩ CanUD | Smk subsets), the genetic correlation was positive, and otherwise it was negative.

We also created polygenic scores (PGS) from each SNP subset in the European ancestry data and tested their associations with a range of health-related phenotypes in the BioVU biobank. We only present associations with mental health and metabolic outcomes in Fig. 4B, but results for all phenotypes with a significant association with at least one of the PGS are available in Table S12.

In line with the genetic correlations in GNOVA, the PGS for the convergent subset of SNPS (Scz ∩ CanUD ∩ Smk) showed the strongest associations overall with most subsets of traits (Fig. 4A), especially suicide attempt, psychosis, PTSD, conduct disorders, antisocial/borderline personality disorder, bipolar disorder, and alcohol-related disorders, among other psychiatric phenotypes (Fig. 4B). Exceptions to this pattern included metabolic and endocrine phenotypes (Fig. 4B), for which the PGS for the CanUD ∩ Smk | Scz subset had the greatest magnitude of associations with many of these traits, including acidosis, adult failure to thrive, type 2 diabetes, and hyperkalemia. In general, PGS for the Scz ∩ CanUD | Smk subset showed associations with metabolic phenotypes in the opposite direction of effect from the PGS for the CanUD ∩ Smk | Scz subset. For example, the PGS for the Scz ∩ CanUD | Smk subset was *negatively* associated with obesity, while the PGS for the CanUD ∩ Smk | Scz subset showed a *positive* association.

Partitioned genetic covariance analysis
When stratified by broad tissue type, the genetic covariance between CanUD and Scz was significantly enriched for brain tissues in the European ancestry data ($\rho = 0.029$, $p = 8.3e-4$), while the genetic covariance between Smk and Scz was not significantly enriched for any tissue category (Fig. S1, Table S13).

DISCUSSION
The nature of the relationship between cannabis use and schizophrenia is a compelling and fiercely debated question in psychiatry, one that is complicated by the possibility of shared genetic factors and the frequent co-occurrence with tobacco smoking. There are major public health implications associated with a causal effect of cannabis use on schizophrenia risk, so a resolution of this question is important. Here, we describe the largest genome-wide, cross-ancestry and cross-disorder analyses of cannabis use disorder (CanUD), tobacco smoking (Smk), and schizophrenia (Scz) to date.

Our analyses revealed three key findings. First, CanUD and Smk are both genetically correlated with Scz, and this was consistent in both the European and African ancestry datasets. However,

CanUD and Scz showed a greater degree of genetic overlap than Smk and Scz. Second, causal inference analyses suggested evidence of bidirectional causality for genetic liability to Scz, CanUD, and Smk, albeit in the presence of horizontal pleiotropy. Third, genomic loci that comprise the intersection between CanUD and Scz are associated with other mental health conditions and executive functioning. Overall, our results support potential reciprocal causal links between schizophrenia, cannabis use disorder, and tobacco smoking, which may have implications for public health efforts. As cannabis use continues to rise alongside parallel increases in nicotine use through vaping, it is important that the public be informed of potential risks *and* that these risks are presented from a nuanced perspective that acknowledges other mechanisms contributing to CanUD, Smk, and Scz comorbidity (including potential shared genetic factors).

In causal inference analyses that accounted for both correlated and uncorrelated forms of horizontal pleiotropy, we saw evidence for bidirectional causal relationships between all three phenotypes. We found evidence of horizontal pleiotropy for all trait pairs through the MR-PRESSO global test, but again found significant bidirectional causal estimates even after removing outlier SNPs for horizontal pleiotropy. Collectively, these results support causal links between CanUD, Smk, and Scz, although it is worth noting that the MR-Egger test did not support any causal relationships except for genetic liability for Smk causing CanUD. Convergent evidence from additional sources (especially longitudinal, prospective cohort studies) are needed [68], especially in light of conflicting results from epidemiological studies [5, 6, 69] and the limitations (and assumptions) associated with genetic methods of causal inference [70].

Over 200 loci had convergent genome-wide significant effects on CanUD, Smk and Scz. The strongest convergent locus was on chromosome 8, with the lead SNP being a brain eQTL for *EPHX2*, *CHRNA2*, and *CCDC25*. While *CHRNA2*, a nicotinic cholinergic receptor (nAChR), seems an intuitive finding for Smk, this locus was most strongly associated with Scz and CanUD (Fig. 2), and the top lead variant in all recent CanUD GWASs has mapped to this locus [12, 17, 54]. The role of cholinergic disturbance in positive [71] symptoms and cognitive symptoms [72] of Scz raise the potential for use of nChR agonists for treatment of comorbid Scz, CanUD and Smk [73]. *EPHX2* encodes soluble epoxide hydrolase (sEH), the overexpression of which has been implicated in Scz [74] and other diseases with a neuroinflammatory component (e.g., Alzheimer's Disease). There is evidence for synergy between sEH and fatty acid amide hydrolase (FAAH [75]), which metabolizes endogenous cannabinoids and the inhibition of which is being evaluated for the treatment of pain. Given the emerging and paradoxical role of CanUD and a proinflammatory state [76], the role of *EPHX2* at the intersection of these disorders is intriguing.

Variants previously implicated in metabolic phenotypes emerged from the Scz ∩ CanUD | Smk subset. For instance, the lead SNP rs9924686 in *TAOK2* was negatively associated with CanUD and Scz but not Smk and has been implicated in numerous prior GWAS of metabolic traits [77, 78]. Further, while PGS derived from this subset were associated with both psychiatric and metabolic traits, the direction of association differed between this subset and the polygenic score of the fully convergent subset for metabolic but not psychiatric phenotypes (Fig. 4).

Genetic predisposition for executive functioning was negatively correlated with the subset of fully convergent variants (Scz ∩ CanUD ∩ Smk), as well as those in the Scz ∩ CanUD | Smk subset, but positively associated with the other two subsets (i.e., where effects diverged for either CanUD or Scz), suggesting that only variants with risk-increasing effects on both CanUD *and* Scz related to lower executive functioning. Executive functioning deficits are a defining feature of Scz [67] and a broad range of substance use disorders [66, 79, 80], consistent with our findings. Executive functioning deficits have also been implicated in a
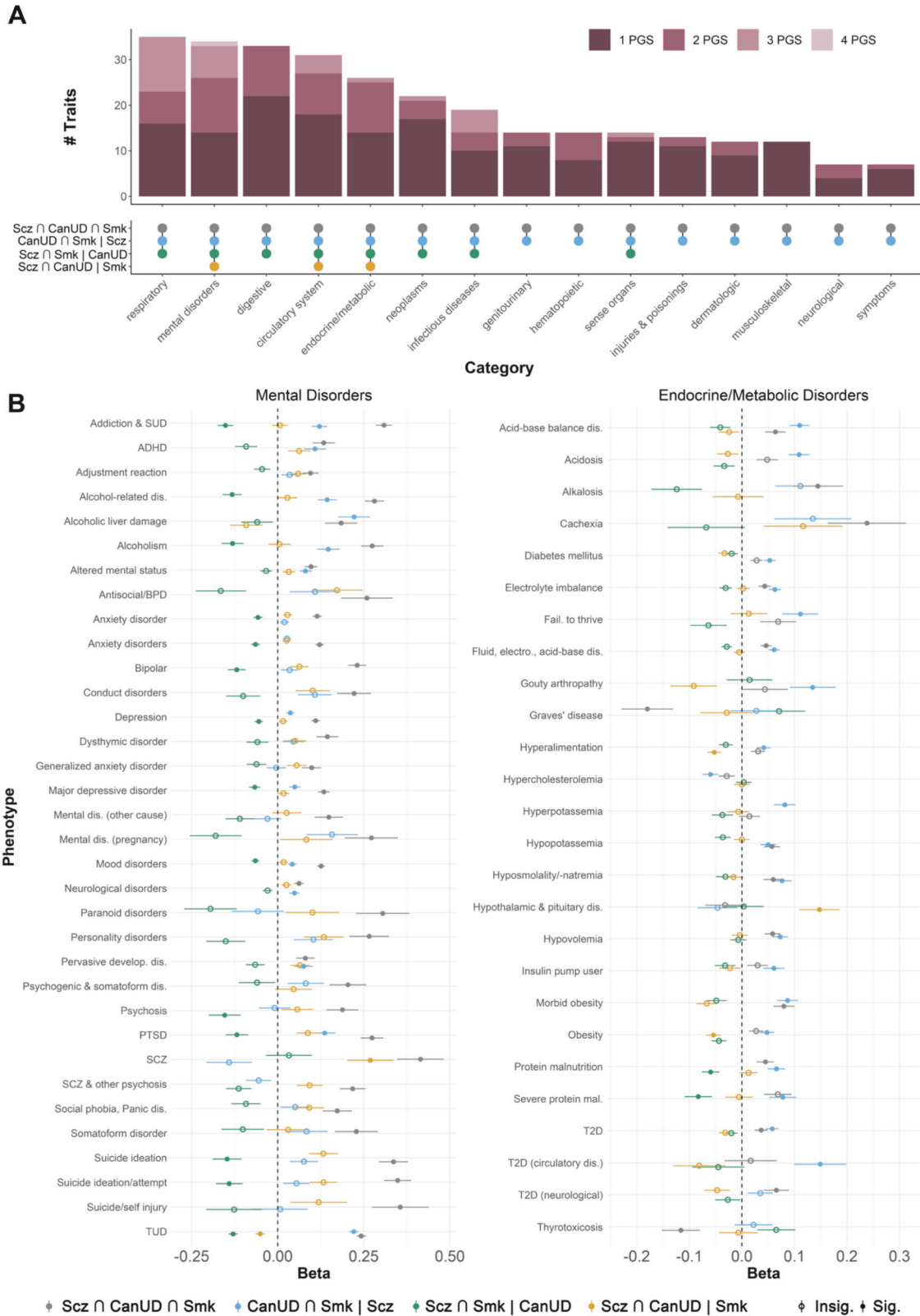
Fig. 4 Associations between polygenic scores for SNP subsets from ASSET and health-related phenotypes in the BioVU biobank. A Upset plot showing the number of phenotypes within different categories associated with one or more PGS. **B** Forest plots showing associations (represented by the regression coefficient, i.e., the change in log(odds)) between the four ASSET SNP subset PGSs and mental disorders (left panel) and endocrine/metabolic traits (right panel) in BioVU.

broader range of mental health conditions [66], which also aligns with our observation that variants influencing CanUD and ScZ, regardless of their effects of Smk, appear to index serious psychiatric comorbidity. Thus, our study implicates the genetic liability to lower executive functioning as a common mechanism undergirding CanUD and Scz, which may prove useful for future studies seeking to develop improved treatment and early intervention efforts. Notably, while executive functioning is related to educational attainment [66], the pattern of associations between subsets of variants and educational attainment appeared to be quite different—for instance, subsets where effects for CanUD and Smk diverged (Scz ∩ CanUD | Smk and Scz ∩ Smk | CanUD) were associated with greater educational attainment, while subsets with convergent effects on both CanUD and Smk were associated with lower educational attainment. Interestingly, the subset of variants with a risk-increasing effect on CanUD but *protective* effects on Scz and Smk (i.e., variants which were divergent for CanUD; Scz ∩ Smk | CanUD) were generally correlated with lower risk for psychopathology (ADHD, bipolar disorder, and depression), risk-taking, and material deprivation, and greater educational attainment and executive functioning.

The genetic correlation between CanUD and Scz ($r_g = 0.37$, SE = 0.02) was significantly greater ($p_{diff} = 6.5e{-}18$) than that between Smk and Scz ($r_g = 0.17$, SE = 0.02) in the European ancestry data (with a similar but non-significant pattern in the African ancestry data: r_g(CanUD, Scz) = 0.61, SE = 0.14 vs. r_g(Smk, Scz) = 0.34, SE = 0.15). This suggests a greater proportion of shared genetic effects for CanUD and Scz than for Smk and Scz. When we partitioned the genetic covariance between phenotype pairs (Scz and CanUD, and Scz and Smk) into broad tissue types, the genetic covariance of CanUD and Scz was significantly enriched for genes functional in brain tissue, while the genetic covariance between Smk and Scz was not significantly enriched in any tissue category. While it is important to note that the size of the annotation set is linked to statistical power, and therefore *p*-values do not necessarily indicate the relative *importance* of different tissue types (e.g., genes functional in immune tissues might also be important for CanUD ~ Scz), the greater statistical power of the Smk GWAS compared to the CanUD GWAS and the substantially larger genetic covariance observed for CanUD~Scz relative to Smk ~ Scz (e.g., $\rho_{Brain} = 0.029$ vs. $\rho_{Brain} = 0.009$, respectively) suggests a meaningful difference in the degree of functional genomic overlap between CanUD and Scz compared to Smk and Scz that is not attributable to statistical power alone. These results are consistent with an overall pattern of findings in our study: the degree of genetic overlap, and the extent to which the genetic covariance is enriched in meaningful biological categories, is greater for CanUD and Scz than for Smk and Scz.

Our analyses of African ancestry data increase the generalizability of our findings. However, the smaller sample size of the individual African ancestry GWASs and limited available data for follow-up analyses (e.g., annotation files for partitioned genetic covariance analyses) constrained the extent to which we were able to accomplish our goal of equitable analyses. The genetic correlation between CanUD and Scz was substantially larger in the African ancestry data ($r_g = 0.610$, SE = 0.140, $p = 1.41e{-}5$) than in the European ancestry data ($r_g = 0.373$, SE = 0.023, $p = 2.97e{-}60$), albeit with a much larger standard error, suggesting that with increasing sample size, there could be considerable opportunity to identify pleiotropic loci.

Several other limitations applied to our study. First, while early age of cannabis initiation and use of high-potency cannabis have been suggested as risk factors for Scz, we did not have data available on potency and did not include age at first use in our analyses, as the only available GWAS for this phenotype was relatively underpowered and had non-significant SNP-heritability [81]. Similarly, we were unaware of any GWAS of cannabis consumption (i.e., heaviness or frequency of use). We also acknowledge the potential for collider bias to affect the causal inference analyses; for example, selection bias (which can induce collider bias) in the UK biobank has been shown to result in over- and under-estimated genetic correlations and MR causal estimates, including for substance use-related traits like drinking frequency and smoking status [82]. It is difficult to definitively diagnose the presence of collider biases, but future studies should explore this further. Another limitation is that the individual GWAS likely contain comorbid cases (e.g., a SCZ case with co-occurring CanUD), and this could artificially inflate our estimates of genetic correlations. Similarly, cannabis is often mixed with tobacco in Europe as well as in certain preparations in the US (e.g., blunts), but existing CanUD GWASs do not account for this potential confounder. Furthermore, cross-trait assortative mating has been shown to bias genetic correlations (e.g., between alcohol use disorders and schizophrenia) [83], although the extent to which this could be affecting estimates of correlation between Scz, CanUD, and Smk specifically has not yet been quantified. We were unable to assess the presence of sex-specific effects related to the relationship between Scz, CanUD, and Smk because sex-stratified versions of the primary GWASs were unavailable. As sample sizes increase, such analyses should be prioritized. Finally, there may be some sample overlap among the different GWAS (especially for CanUD and Smk), which could have inflated our MR results, but not results from CAUSE, which accounts for sample overlap.

Overall, our results add to the body of literature suggesting that both Smk and CanUD may be important predisposing factors as well as sequela of Scz. We demonstrate that the relationship between Smk, CanUD, and Scz may be due to both correlated genetic and reciprocal causal effects. While cigarette use is generally decreasing [84], nicotine exposure through vaping is increasing [26, 85] and cannabis legalization and use are becoming more widespread worldwide [86]. As substance use policies and modes of use continue to change, it is important to carefully monitor epidemiologic trends in mental health conditions, especially schizophrenia and other psychotic disorders, and consider targeted interventions that may benefit individuals with heavy cannabis and tobacco use.

REFERENCES

1. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015;47:702–9.
2. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 2003;60:1187–92.
3. Volkow ND. Substance use disorders in schizophrenia—clinical implications of comorbidity. Schizophr Bullet. 2009;35:469–72.
4. Gage SH, Hickman M, Zammit S. Association between cannabis and psychosis: epidemiologic evidence. Biol Psychiatry. 2016;79:549–56.
5. Di Forti M, Marconi A, Carra E, Fraietta S, Trotta A, Bonomo M, et al. Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study. Lancet Psychiatry. 2015;2:233–8.
6. Hjorthøj C, Compton W, Starzer M, Nordholm D, Einstein E, Erlangsen A. et al. Association between cannabis use disorder and schizophrenia stronger in young males than in females. Psychol Med. 2023:1–7. https://doi.org/10.1017/S0033291723000880.
7. Andréasson S, Engström A, Allebeck P, Rydberg U. CANNABIS AND SCHIZO-PHRENIA A Longitudinal Study of Swedish Conscripts. The Lancet. 1987;330:1483–6.
8. Deak JD, Johnson EC. Genetics of substance use disorders: a review. Psychol Med. 2021:1–12. https://doi.org/10.1017/S0033291721000969.
9. 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. Addiction Biol. 2017;53:7322–8. https://doi.org/10.1111/adb.12496.
10. 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.
11. Pasman JA, Verweij K, 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.

12. 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. https://doi.org/10.1016/S2215-0366(20)30339-4.

13. 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.

14. Gillespie NA, Kendler KS. Use of genetically informed methods to clarify the nature of the association between cannabis use and risk for schizophrenia. JAMA Psychiatry. 2021;78:467–8.

15. Vaucher J, Keating BJ, Lasserre AM, Gan W, Lyall DM, Ward J, et al. Cannabis use and risk of schizophrenia: a Mendelian randomization study. Mol Psychiatry. 2017;23:1287–92.

16. Gage SH, Jones HJ, Burgess S, Bowden J, Davey Smith G, Zammit S, et al. Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. Psychol Med. 2017;47:971–80.

17. 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.

18. Hartz SM, Pato CN, Medeiros H, Cavazos-Rehg P, Sobell JL, Knowles JA, et al. Comorbidity of severe psychotic disorders with measures of substance use. JAMA Psychiatry. 2014;71:248–54.

19. Gurillo P, Jauhar S, Murray RM, MacCabe JH. Does tobacco use cause psychosis? Systematic review and meta-analysis. Lancet Psychiatry. 2015;2:718–25.

20. Kendler KS, Lönn SL, Sundquist J, Sundquist K. Smoking and schizophrenia in population cohorts of Swedish women and men: a prospective co-relative control study. Am J Psychiatry. 2015;172:1092–1100.

21. Agrawal A, Budney AJ, Lynskey MT. The co-occurring use and misuse of cannabis and tobacco: a review. Addiction. 2012;107:1221–33.

22. Agrawal A, Lynskey MT. Tobacco and cannabis co-occurrence: does route of administration matter? Drug Alcohol Depend. 2009;99:240–7.

23. 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.

24. Hindocha C, Freeman TP, Ferris JA, Lynskey MT, Winstock AR. No smoke without tobacco: a global overview of cannabis and tobacco routes of administration and their association with intention to quit. Front Psychiatry. 2016;7:104.

25. Cerdá M, Mauro C, Hamilton A, Levy NS, Santaella-Tenorio J, Hasin D, et al. Association between recreational Marijuana legalization in the United States and changes in Marijuana use and Cannabis use disorder from 2008 to 2016. JAMA Psychiatry. 2020;77:165–71.

26. Miech R, Leventhal A, Johnston L, O'Malley PM, Patrick ME, Barrington-Trimis J. Trends in use and perceptions of nicotine vaping among US youth from 2017 to 2020. JAMA Pediatr. 2021;175:185–90.

27. 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.

28. Iasevoli F, Balletta R, Gilardi V, Giordano S, De Bartolomeis A. Tobacco smoking in treatment-resistant schizophrenia patients is associated with impaired cognitive functioning, more severe negative symptoms, and poorer social adjustment. Neuropsychiatr Dis Treat. 2013;9:1113–20.

29. Wang C-Y, Xiang YT, Weng YZ, Bo QJ, Chiu HF, Chan SS, et al. Cigarette smoking in patients with schizophrenia in China: prospective, multicentre study. Aust N Z J Psychiatry. 2010;44:456–62.

30. Dickerson F, Stallings CR, Origoni AE, Vaughan C, Khushalani S, Schroeder J, et al. Cigarette smoking among persons with schizophrenia or bipolar disorder in routine clinical settings, 1999–2011. Psychiatr Serv. 2013;64:44–50.

31. Foti DJ, Kotov R, Guey LT, Bromet EJ. Cannabis use and the course of schizophrenia: 10-year follow-up after first hospitalization. Am J Psychiatry. 2010;167:987–93.

32. Trubetskoy V, Pardiñas 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.

33. Saunders G, 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.

34. Quach BC, Bray MJ, Gaddis NC, Liu M, Palviainen T, Minica CC, et al. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. Nat Commun. 2020;11:5562.

35. Bigdeli TB, Genovese G, Georgakopoulos P, Meyers JL, Peterson RE, Iyegbe CO et al. Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. Mol Psychiatry. 2019;10:1–13.

36. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.

37. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.

38. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet. 2020;52:740–7.

39. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693–8.

40. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.

41. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13:e1007081.

42. Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, et al. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am J Hum Genet. 2012;90:821–35.

43. Lam M, Hill WD, Trampush JW, Yu J, Knowles E, Davies G, et al. Pleiotropic meta-analysis of cognition, education, and schizophrenia differentiates roles of early neurodevelopmental and adult synaptic pathways. Am J Hum Genet. 2019;105:334–50.

44. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.

45. Consortium T. 1000 G. P. A global reference for human genetic variation. Nature. 2015;526:68–74.

46. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. Bioinformatics. 2010;26:2190–1.

47. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R package for rapidly calculating linkage disequilibrium statistics in diverse populations. Front Genet. 2020;11.

48. Lu Q, Li B, Ou D, Erlendsdottir M, Powles RL, Jiang T, et al. A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics. Am J Hum Genet. 2017;101:939–64.

49. 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. 2019;1–11.

50. Dennis JK, Sealock JM, Straub P, Lee YH, Hucks D, Actkins K, et al. Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. Genome Med. 2021;13:6.

51. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.

52. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I 2 statistic. Int J Epidemiol. 2016;45:1961–74.

53. Consortium. S. W. G. of the P. G. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.

54. Demontis D, Rajagopal VM, Thorgeirsson TE, Als TD, Grove J, Leppälä K, et al. Genome-wide association study implicates CHRNA2 in cannabis use disorder. Nat Neurosci. 2019;1.

55. 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.

56. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 2016;167:1415–29.

57. Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019. https://doi.org/10.1038/s41588-019-0481-0.

58. Karlsson Linnér 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.

59. 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.

60. Pasman JA, Demange PA, Guloksuz S, Willemsen A, Abdellaoui A, Ten Have M, et al. Genetic risk for smoking: disentangling interplay between genes and socioeconomic status. Behav Genet. 2022;52:92–107.

61. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.

62. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman J, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021;53:817–29.

63. Levey DF, Stein MB, Wendt FR, Pathak GA, Zhou H, Aslan M, et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat Neurosci. 2021;24:954–63.

64. Nievergelt CM, Maihofer AX, Klengel T, Atkinson EG, Chen CY, Choi KW, et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. Nat Commun. 2019;10:4558.

65. 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.

66. Hatoum AS, Morrison CL, Mitchell EC, Lam M, Benca-Bachman CE, Reineberg AE, et al. Genome-wide association study shows that executive functioning is influenced by GABAergic processes and is a neurocognitive genetic correlate of psychiatric disorders. Biol Psychiatry. 2023;93:59–70.

67. Song J, Yao S, Kowalec K, Lu Y, Sariaslan A, Szatkiewicz JP, et al. The impact of educational attainment, intelligence and intellectual disability on schizophrenia: a Swedish population-based register and genetic study. Mol Psychiatry. 2022;27:2439–47.

68. Munafò MR, Davey Smith G. Robust research needs many lines of evidence. Nature. 2018;553:399–401.

69. Elser H, Humphreys K, Kiang MV, Mehta S, Yoon JH, Faustman WO, et al. State cannabis legalization and psychosis-related health care utilization. JAMA Netw Open. 2023;6:e2252689–e2252689.

70. Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. bmj. 2018;362.

71. Caton M, Ochoa ELM, Barrantes FJ. The role of nicotinic cholinergic neurotransmission in delusional thinking. NPJ Schizophr. 2020;6:16.

72. D'Souza MS, Markou A. Schizophrenia and tobacco smoking comorbidity: nAChR agonists in the treatment of schizophrenia-associated cognitive deficits. Neuropharmacology. 2012;62:1564–73.

73. Lieberman JA, Javitch JA, Moore H. Cholinergic agonists as novel treatments for schizophrenia: the promise of rational drug development for psychiatry. Am J Psychiatry. 2008;165:931–6.

74. Ma M, Ren Q, Yang J, Zhang K, Xiong Z, Ishima T, et al. Key role of soluble epoxide hydrolase in the neurodevelopmental disorders of offspring after maternal immune activation. Proc Natl Acad Sci. 2019;116:7083–8.

75. Kodani SD, Wan D, Wagner KM, Hwang SH, Morisseau C, Hammock BD. Design and potency of dual soluble epoxide hydrolase/fatty acid amide hydrolase inhibitors. ACS Omega. 2018;3:14076–86.

76. Wei T-T, Chandy M, Nishiga M, Zhang A, Kumar KK, Thomas D, et al. Cannabinoid receptor 1 antagonist genistein attenuates marijuana-induced vascular inflammation. Cell. 2022;185:1676–1693.e23.

77. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet. 2019;28:166–74.

78. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet. 2018;27:3641–9.

79. Koob GF & Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. The Lancet Psychiatry. 2016;3:760–73.

80. Hatoum AS, Johnson EC, Colbert S, Polimanti R, Zhou H, Walters RK, et al. The addiction risk factor: a unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. Neuropsychopharmacology. 2022;47:1739–45.

81. Minică CC, Verweij K, 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.

82. 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.

83. Border R, Athanasiadis G, Buil A, Schork AJ, Cai N, Young AI, et al. Cross-trait assortative mating is widespread and inflates genetic correlation estimates. Science (1979). 2022;378:754–61.

84. Meza R, Jimenez-Mendoza E, Levy DT. Trends in tobacco use among adolescents by grade, sex, and race, 1991-2019. JAMA Netw Open. 2020;3:e2027465–e2027465.

85. Tattan-Birch H, Brown J, Shahab L, Jackson SE. Trends in use of e-cigarette device types and heated tobacco products from 2016 to 2020 in England. Sci Rep. 2021;11:13203.

86. Hall W, Stjepanović D, Caulkins J, Lynskey M, Leung J, Campbell G, et al. Public health implications of legalising the production and sale of cannabis for medicinal and recreational use. Lancet. 2019;394:1580–90.

AUTHOR CONTRIBUTIONS

ECJ and AA contributed to conceptualization of the study. ECJ, IAZ, HHAT, SSR, and DAAB contributed to analysis of data. AA contributed to supervision of analyses. ECJ, IAZ, HHAT, and AA contributed to the initial draft of the manuscript. All authors contributed to critical revisions of the manuscript and approved of the final draft for submission.

FUNDING

COMPETING INTERESTS

JG is a holder of US patent 10,900,082 titled: 'Genotype-guided dosing of opioid agonists', issued 26 January 2021. JG is paid for their editorial work on the journal Complex Psychiatry. HHAT is on the *Neuropsychopharmacology* Special Projects Team. The remaining authors have nothing to disclose.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41386-024-01886-3.

Correspondence and requests for materials should be addressed to Emma C. Johnson.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.