

## ARTICLE OPEN



# Application of polygenic scores to a deeply phenotyped sample enriched for substance use disorders reveals extensive pleiotropy with psychiatric and somatic traits

Emily E. Hartwell <sup>1,2</sup>, Zeal Jinwala<sup>1,2</sup>, Jackson Milone<sup>2</sup>, Sarah Ramirez<sup>2</sup>, Joel Gelernter <sup>3,4</sup>, Henry R. Kranzler <sup>1,2</sup> and Rachel L. Kember <sup>1,2</sup> ✉

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024

Co-occurring psychiatric, medical, and substance use disorders (SUDs) are common, but the complex pathways leading to such comorbidities are poorly understood. A greater understanding of genetic influences on this phenomenon could inform precision medicine efforts. We used the Yale-Penn dataset, a cross-sectional sample enriched for individuals with SUDs, to examine pleiotropic effects of genetic liability for psychiatric and somatic traits. Participants completed an in-depth interview that provides information on demographics, environment, medical illnesses, and psychiatric and SUDs. Polygenic scores (PGS) for psychiatric disorders and somatic traits were calculated in European-ancestry (EUR;  $n = 5691$ ) participants and, when discovery datasets were available, for African-ancestry (AFR;  $n = 4918$ ) participants. Phenome-wide association studies (PheWAS) were then conducted. In AFR participants, the only PGS with significant associations was bipolar disorder (BD), all of which were with substance use phenotypes. In EUR participants, PGS for major depressive disorder (MDD), generalized anxiety disorder (GAD), post-traumatic stress disorder (PTSD), schizophrenia (SCZ), body mass index (BMI), coronary artery disease (CAD), and type 2 diabetes (T2D) all showed significant associations, the majority of which were with phenotypes in the substance use categories. For instance, PGS<sub>MDD</sub> was associated with over 200 phenotypes, 15 of which were depression-related (e.g., depression criterion count), 55 of which were other psychiatric phenotypes, and 126 of which were substance use phenotypes; and PGS<sub>BMI</sub> was associated with 138 phenotypes, 105 of which were substance related. Genetic liability for psychiatric and somatic traits is associated with numerous phenotypes across multiple categories, indicative of the broad genetic liability of these traits.

*Neuropsychopharmacology* (2024) 49:1958–1967; <https://doi.org/10.1038/s41386-024-01922-2>

## INTRODUCTION

Medical illness and psychiatric disorders, including substance use disorders (SUDs), frequently co-occur. Individuals with chronic medical conditions are more likely to have a co-occurring SUD or psychiatric diagnosis [1–5] and over 9 million U.S. adults have a psychiatric disorder that co-occurs with an SUD [6]. The development of a comorbid disorder can exacerbate pre-existing conditions [7, 8] and worsen an individual's prognosis [9, 10]. Moreover, co-occurring disorders can limit treatment options [11] and adversely affect treatment outcomes by reducing treatment adherence or decreasing its effectiveness [12–14]. Understanding the genetic underpinnings of comorbid disorders could improve their diagnosis, treatment, and ongoing management, thus informing precision medicine efforts.

Genetic liability for medical and psychiatric disorders has been discovered using genome-wide association studies (GWAS), which identify associations between common genetic variants and the trait of interest. These studies have identified pleiotropic variants, i.e., those associated with multiple conditions. GWAS findings have also demonstrated significant genetic correlations between SUDs and other psychiatric disorders [15, 16] and medical conditions

[17]. These findings contribute to a growing body of evidence that shared genetic risk loci or common biological pathways may underlie co-occurring conditions.

Polygenic scores (PGS) provide a measure of an individual's genetic risk for specific traits and as such are a complementary method to investigate genetic overlap. Previous studies have shown that PGS are associated with conditions such as cardiovascular disease [18], kidney disease [19], opioid use disorder [20], depression [21], and pain [22], among many others. PGS may also be used in phenome-wide association studies (PheWAS) [23] to provide insight into the pleiotropic nature of genetic liability for disorders [24, 25]. PheWAS, which have been commonly implemented using electronic health record (EHR) databases, measure the association between a PGS for a disorder by testing it against multiple phenotypes in a hypothesis-free paradigm.

Here, we used the Yale-Penn sample—which comprises a diverse sample of participants recruited for genetic studies of cocaine, opioid, and alcohol dependence—to conduct PheWAS of psychiatric and somatic PGS. Yale-Penn participants completed the Semi-Structured Assessment for Drug Dependence and

<sup>1</sup>Crescenz VA Medical Center, Philadelphia, PA, USA. <sup>2</sup>University of Pennsylvania, Philadelphia, PA, USA. <sup>3</sup>West Haven VA Medical Center, West Haven, CT, USA. <sup>4</sup>Yale University, New Haven, CT, USA. ✉email: [rkember@pennteam.upenn.edu](mailto:rkember@pennteam.upenn.edu)

Received: 22 January 2024 Revised: 7 June 2024 Accepted: 28 June 2024

Published online: 23 July 2024

Alcoholism (SSADDA) which queries medical, psychosocial, and substance use history and diagnoses, psychiatric diagnoses, and demographics [26, 27]. Previous studies have utilized the Yale-Penn sample to conduct gene x environment studies [28], linkage and association studies of substance use and dependence [29–36], and to examine phenotypic associations [37]. These studies have shown shared genetic liability across SUDs, psychiatric disorders, and environmental traits.

Using the Yale-Penn sample, we created a simplified PheWAS dataset for genetic analysis and calculated PGS to examine pleiotropy for four major substance-related traits: alcohol use disorder, opioid use disorder, smoking initiation, and lifetime cannabis use [38]. PheWAS analyses in European-ancestry participants identified significant associations between SUD PGS and substance and psychiatric diagnoses and demographic and environmental phenotypes. Here, we extend this work by examining the associations of PGS for a variety of psychiatric disorders and somatic traits in the Yale-Penn sample.

## METHODS

### Participants and procedures

The Yale-Penn sample ( $N = 14,040$ ) was recruited from five U.S. academic sites for studies of the genetics of cocaine, alcohol, and opioid use disorders. The institutional review boards at University of Connecticut Health, Medical University of South Carolina, McLean Hospital, University of Pennsylvania, and Yale University approved the study protocol and informed consent forms. After they gave informed consent, all participants were administered the SSADDA and provided a blood or saliva sample for genotyping. The SSADDA comprises 24 modules that assess demographic information, environmental variables, medical history, and psychiatric and substance use history and diagnoses [26]. Additional information on variable selection and cleaning has been published [38]. In brief, the SSADDA yields over 3700 variables, which we refined to 691 variables for use in PheWAS by selecting variables that were considered informative for genetic studies and non-duplicative [38]. These variables are grouped into 25 categories: Demographics, Medical History, Substance Use (Tobacco, Alcohol, Cocaine, Opiate, Marijuana, Sedatives, Stimulants, Other drugs), Psychiatric (Major Depression, Conduct Disorder, Antisocial Personality Disorder [ASPD], Attention Deficit Hyperactivity Disorder [ADHD], Suicidality, Post-Traumatic Stress Disorder [PTSD], Generalized Anxiety Disorder [GAD], Panic Disorder, Social Phobia, Mania, Agoraphobia, Obsessive Compulsive Disorder [OCD], Schizophrenia, and Gambling) and Environment.

### Case and control definitions

Participants who endorsed Diagnostic and Statistical Manual (DSM) criteria for a given lifetime disorder (DSM-IV for psychiatric disorders, DSM-IV and DSM-5 for SUDs) were coded as cases and those who met no diagnostic criteria were considered controls. Participants meeting a sub-threshold number of criteria (e.g., one criterion when multiple are required for diagnosis) were excluded from analyses for that disorder. For individual symptoms (e.g., suicide attempt), participants who responded affirmatively were considered cases and those who did not were considered controls. When an item was not answered, participants were coded as “NA” and not included as either a case or a control for that variable.

### Genotyping and imputation

In brief, Yale-Penn participants were genotyped in three batches using Illumina microarrays at Center for Inherited Disease Research (CIDR) or the Gelernter lab at Yale and imputed using the Michigan Imputation Server [39] with the 1000 Genomes phase 3 reference panel [40]. Details on genotyping, imputation, and quality control for the genetic data have previously been reported [36, 38, 41, 42].

Ancestry-specific PGS were calculated using PRS-Continuous Shrinkage (PRS-CS) software [43] from GWAS in discovery samples for anorexia (AN) [44], autism spectrum disorder (ASD) [45], bipolar disorder (BD) [46], generalized anxiety disorder (GAD) [47], major depressive disorder (MDD) [48], obsessive compulsive disorder (OCD) [49], panic disorder (PD) [50], post-traumatic stress disorder (PTSD) [51], schizophrenia (SCZ) [52], Tourette syndrome (TS) [53], body mass index (BMI) [54], coronary artery

disease (CAD) [55] and type 2 diabetes (T2D) [56] (Supplementary Table 1). All GWAS were available for European-ancestry (EUR), but only BD [57], GAD [47], MDD [21], PTSD [51], and SCZ [52] were available for African-ancestry (AFR). Discovery GWAS were selected based on their public availability and excluded the Yale-Penn sample. The global shrinkage parameter  $\phi$  was learned from the data and default values were used for other parameters as described on the github page for the software (<https://github.com/getian107/PRScs>).

### Statistical analysis

For PGS with available primary phenotypes (diagnoses for AN, ASD and TS are not available in the Yale-Penn sample, and individuals with SCZ diagnoses were excluded from recruitment), we tested for association between the PGS and the primary phenotype using logistic regression models in R, with  $p < 0.05$  considered significant. We next conducted a series of PheWAS using logistic regression models for binary traits and linear regression models for continuous traits, adjusting for age, sex, and the top 10 principal genetic components within each ancestry. Phenotypes in which there were less than 100 cases or controls were excluded. For available phenotypes, a second PheWAS was run that covaried for the primary diagnostic phenotype. A Bonferroni correction was applied to each ancestry group to account for multiple comparisons (AFR phenotypes  $n = 574$ ,  $p = 8.7 \times 10^{-05}$ ; EUR phenotypes  $n = 620$ ,  $p = 8.1 \times 10^{-05}$ ). Nagelkerke  $R^2$  was calculated to quantify the variance explained by PGS only and PGS with covariates. Additionally, we also calculated a pseudo- $R^2$  metric developed by Lee et al. [58] measured on a liability scale to avoid bias in these estimates due to sample prevalence not being equal to population prevalence.

## RESULTS

### Sample

Genetic data were available for 10,275 of the 14,040 participants, the majority of whom (54.46%) were male. The sample included 4851 AFR participants (55.2% males) and 5424 EUR participants (51.1% males) whose mean ages were 41.47 ( $SD = 10.16$ ) and 39.79 ( $SD = 12.91$ ), respectively. Supplementary Table 2 shows demographic information by for available primary phenotypes.

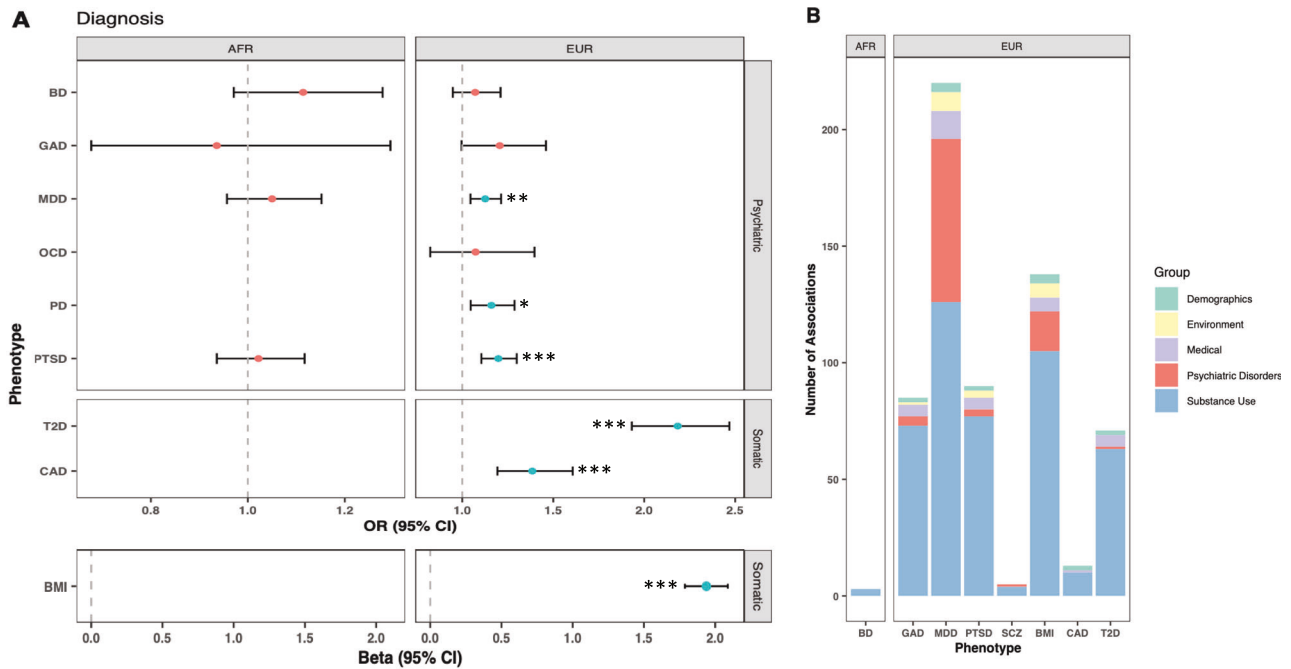
### Primary phenotypic associations of PGS

For the PGS with primary phenotypes available, we tested the association of PGS with each primary phenotype (Fig. 1A). In AFR participants, none of the PGS were associated with their primary phenotype. In EUR participants, PGS for three psychiatric disorders (PGS<sub>MDD</sub>, PGS<sub>PD</sub>, and PGS<sub>PTSD</sub>) and three somatic traits (PGS<sub>BMI</sub>, PGS<sub>CAD</sub> and PGS<sub>T2D</sub>) were associated with their primary phenotype at a  $p$ -value of  $< 0.05$ . The proportion of phenotypic variance explained by the PGS alone ranged from 0.26 to 10.10% (Nagelkerke’s pseudo- $R^2$ ) and 0.10 to 4.68% (liability scale  $R^2$ ), in line with previous estimates (Table 1).

We next examined phenotypic associations of each PGS other than the primary phenotype. In AFR participants, after Bonferroni correction, there were significant associations for PGS<sub>BD</sub> (Supplementary Table 3). No other associations were observed among AFR participants following Bonferroni correction (Supplementary Tables 4–7). In EUR participants, there were significant associations for five of the psychiatric disorders (PGS<sub>MDD</sub>, PGS<sub>GAD</sub>, PGS<sub>PTSD</sub>, PGS<sub>SCZ</sub>, and PGS<sub>TS</sub>) and somatic traits (PGS<sub>BMI</sub>, PGS<sub>CAD</sub> and PGS<sub>T2DM</sub>), whereas there were no significant associations for PGS<sub>BD</sub>, PGS<sub>AN</sub>, PGS<sub>ASD</sub>, PGS<sub>OCD</sub> or PGS<sub>PD</sub> (Supplementary Tables 8–20).

### Phenome-wide analysis of psychiatric PGS

**Bipolar disorder (BD).** In AFR participants, PGS<sub>BD</sub> was associated with three phenotypes in the substance use category, all related to cocaine (e.g., regularly use cocaine,  $OR = 1.14$ ,  $CI = 1.07–1.20$ ,  $p = 8.6 \times 10^{-5}$ ; Fig. 1B; Supplementary Tables 3 and 21). PGS<sub>BD</sub> was not associated with any phenotypes in EUR participants (Supplementary Table 8).



**Fig. 1 Primary and secondary associations of psychiatric and somatic PGS. A** Effect size and 95% confidence intervals for associations between PGS and their corresponding primary phenotype, if available. Asterisks indicate  $p$ -value for significant associations:  $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ . **B** Number of associations within each category for the PGS with significant associations. BD Bipolar Disorder, GAD generalized anxiety disorder, MDD major depressive disorder, OCD obsessive compulsive disorder, PD panic disorder, PTSD post-traumatic stress disorder, SCZ schizophrenia, T2D diabetes, CAD coronary artery disease, BMI body mass index.

**Major depressive disorder (MDD).** In EUR participants,  $PGS_{MDD}$  was associated with 220 phenotypes across 17 categories (Figs. 1B and 2; Supplementary Tables 9 and 21). Although  $PGS_{MDD}$  was not significantly associated with the MDD diagnosis following Bonferroni correction (OR = 1.13, CI = 1.05–1.21,  $p = 3.8 \times 10^{-3}$ ; Fig. 1A), it was significantly associated with 15 phenotypes in the depression category, most significantly the MDD criterion count ( $\beta = 0.40$ , CI = 0.30–0.50,  $p = 3.0 \times 10^{-15}$ ). These phenotypes remained significantly associated when covarying for MDD diagnosis (Supplementary Fig. 1, Supplementary Table 9).

$PGS_{MDD}$  also showed 55 significant associations with other psychiatric disorders. Notably,  $PGS_{MDD}$  was the only PGS associated with any phenotypes in the depression, generalized anxiety (e.g., sum of physical reactions ( $\beta = 0.14$ , CI = 0.08–0.20,  $p = 5.9 \times 10^{-9}$ ), panic disorder (e.g. shortness of breath (OR = 1.29, CI = 1.18–1.41,  $p = 4.0 \times 10^{-9}$ )), agoraphobia (e.g., ever agoraphobia (OR = 1.16, CI = 0.08–0.23,  $p = 5.84 \times 10^{-5}$ )), and suicide categories (Supplementary Table 21). Five phenotypes related to suicide were significantly associated with  $PGS_{MDD}$ , the most significant being high suicidal intent (OR = 1.37, CI = 1.27–1.48,  $p = 6.5 \times 10^{-9}$ ). Covarying for the MDD diagnosis reduced the number of associations in the panic disorder category from 18 to 13, and the number of associations with GAD from 4 to 0.  $PGS_{MDD}$  was the only PGS associated with the number of inpatient psychiatric treatments ( $\beta = 0.50$ , CI = 0.32–0.68,  $p = 2.8 \times 10^{-8}$ ), emotional problems (OR = 1.28, CI = 1.21–1.35,  $p = 6.0 \times 10^{-12}$ ) and history of antidepressant use (OR = 1.24, CI = 1.18–1.30,  $p = 2.0 \times 10^{-12}$ ).

Other categories with significant associations included those with PTSD (e.g. criterion count ( $\beta = 0.26$ , CI = 0.20–0.31,  $p = 2.3 \times 10^{-18}$ )), conduct disorder (e.g., truancy, being suspended or expelled from school (OR = 1.31, CI = 1.25–1.38,  $p = 8.5 \times 10^{-16}$ ), ASPD (e.g., impulsivity (OR = 1.17, CI = 1.11–1.24,  $p = 8.6 \times 10^{-7}$ ), and ADHD (e.g., criterion count ( $\beta = 0.09$ , CI = 0.06–0.12,  $p = 9.6 \times 10^{-10}$ )). Additionally,  $PGS_{MDD}$  was significantly associated with demographic and environmental phenotypes, including negatively with education ( $\beta = -0.10$ , CI = -0.12–0.08,

$p = 8.7 \times 10^{-20}$ ) and positively with childhood adversity (OR = 1.29, CI = 1.22–1.35,  $p = 6.8 \times 10^{-13}$ ).

$PGS_{MDD}$  was also significantly associated with 126 substance use phenotypes, 122 of which remained significant after covarying for the MDD diagnosis. The substance use traits most significantly associated with  $PGS_{MDD}$  in each category were the Fagerström Test for Nicotine Dependence (FTND) score ( $\beta = 0.38$ , CI = 0.31–0.46,  $p = 8.7 \times 10^{-23}$ ), criterion count for DSM-5 cocaine use disorder (CocUD;  $\beta = 0.56$ , CI = 0.44–0.69,  $p = 1.3 \times 10^{-18}$ ), “ever used” opioids (OR = 1.31, CI = 1.25–1.37,  $p = 5.1 \times 10^{-17}$ ), and DSM-IV alcohol abuse (OR = 1.25, CI = 1.18–1.31,  $p = 2.3 \times 10^{-11}$ ). Notably,  $PGS_{MDD}$  had the most alcohol associations of the PGS tested.

**Generalized anxiety disorder (GAD).**  $PGS_{GAD}$  was associated with 85 phenotypes in EUR participants (Figs. 1B and 2; Supplementary Tables 10 and 21). Although it was not significantly associated with the primary diagnosis of GAD (OR = 1.21, CI = 1.01–1.40,  $p = 0.06$ ; Fig. 1), it was the only PGS to be associated with a history of anxiolytic treatment (OR = 1.19, CI = 1.11–1.26,  $p = 8.7 \times 10^{-6}$ ).

$PGS_{GAD}$  was significantly associated with four phenotypes related to other psychiatric disorders, which were hyperactivity-impulsivity ( $\beta = 0.18$ , CI = 0.12–0.25,  $p = 1.2 \times 10^{-7}$ ) and inattention ( $\beta = 0.17$ , CI = 0.09–0.25,  $p = 2.4 \times 10^{-5}$ ) for ADHD; truancy, being suspended or expelled from school (OR = 1.18, CI = 1.12–1.24,  $p = 2.9 \times 10^{-7}$ ) for conduct disorder; and seeking treatment for PTSD (OR = 1.21, CI = 1.12–1.31,  $p = 4.4 \times 10^{-5}$ ). Additionally,  $PGS_{GAD}$  was significantly associated with non-psychiatric phenotypes, such as health rating (higher value indicates poorer health;  $\beta = 0.07$ , CI = 0.05–0.10,  $p = 2.4 \times 10^{-7}$ ) and education ( $\beta = -0.07$ , CI = -0.09–0.05,  $p = 1.0 \times 10^{-11}$ ).

$PGS_{GAD}$  was significantly associated with 73 substance use phenotypes, particularly in the tobacco, cocaine, and opioid categories (e.g., FTND score ( $\beta = 0.21$ , CI = 0.14–0.28,  $p = 2.5 \times 10^{-8}$ ), using more cocaine than intended (OR = 1.18, CI = 1.07–1.19,  $p = 1.0 \times 10^{-8}$ ), DSM-IV opioid dependence (OR = 1.23, CI = 1.17–1.29,  $p = 3.3 \times 10^{-11}$ ), and DSM-5 alcohol use

**Table 1.** Phenotypic variance explained by the PGS that were significantly associated with their primary phenotype in EUR.

PGS	Sample prevalence	Population prevalence	Nagelkerke's R <sup>2</sup> (PGS + covariates)	Nagelkerke's R <sup>2</sup> (PGS only)	Liability scale R <sup>2</sup> (PGS + covariates)	Liability scale R <sup>2</sup> (PGS only)	Previously reported R <sup>2</sup>
MDD	14.48%	8.30% [68]	4.19%	0.27%	4.81%	0.17%	1.5–3.2% [69]
PD	7.06%	4.70% [70]	4.32%	0.26%	5.36%	0.10%	0.8–2.6% [50]
PTSD	13.18%	6.80% [71]	9.62%	0.76%	9.13%	1.11%	0.40% [51]
BMI	n/a	n/a	15%	10.10%	n/a	n/a	5.40% [72]
CAD	3.30%	7% [73]	12.09%	1.18%	23.20%	1.46%	6.1% [74]
T2D	4.96%	11.60% [75]	14.10%	5.26%	26.26%	4.68%	5.37% [56]

Previously reported R<sup>2</sup> have been calculated using a variety of methods, see references for details.

disorder (AUD, OR = 1.16, CI = 1.09–1.23,  $p = 2.1 \times 10^{-5}$ ). Unlike other psychiatric PGS, PGS<sub>GAD</sub> also had two significant associations with marijuana use.

Covarying for the primary phenotype, half of the phenotypes associated with PGS<sub>GAD</sub> were not significant, although the association with anxiolytic treatment remained significant (Supplementary Fig. 2, Supplementary Table 10). The tobacco category had the greatest reduction in number of associations, with 8 of 11 phenotypes no longer significant when GAD diagnosis was included as a covariate.

**Post-traumatic stress disorder (PTSD).** PGS<sub>PTSD</sub> was associated with a total of 90 phenotypes in EUR participants (Figs. 1B and 2; Supplementary Tables 11 and 21). PGS<sub>PTSD</sub> showed significant associations with the diagnosis of PTSD (OR = 1.20, CI = 1.11–1.28,  $p = 3.8 \times 10^{-5}$ ) and with treatment-seeking for PTSD (OR = 1.21, CI = 1.11–1.30,  $p = 5.8 \times 10^{-5}$ ). The only other association in the psychiatric category was with truancy, being suspended, or expelled from school in the conduct disorder category (OR = 1.18, CI = 1.11–1.24,  $p = 3.9 \times 10^{-7}$ ). These associations were no longer significant when the PTSD diagnosis was used as a covariate in the analysis (Supplementary Fig. 3, Supplementary Table 11).

PGS<sub>PTSD</sub> was significantly associated with 77 cocaine, tobacco, and opioid use phenotypes, including DSM-IV dependence and withdrawal symptoms for all three substances. Almost half of these traits were not significant when the PTSD diagnosis was used as a covariate.

Similar to other PGS results, PGS<sub>PTSD</sub> was also significantly associated with demographic phenotypes (e.g., education ( $\beta = -0.07$ , CI =  $-0.09$ – $-0.05$ ,  $p = 5.8 \times 10^{-11}$ )), environment phenotypes (e.g., household members being cigarette smokers (OR = 1.18, CI = 1.11–1.24,  $p = 2.2 \times 10^{-7}$ )), and medical phenotypes (e.g., health rating ( $\beta = 0.08$ , CI = 0.05–0.10,  $p = 8.4 \times 10^{-8}$ )). However, the majority of these phenotypes became nonsignificant when the PTSD diagnosis was used as a covariate.

**Schizophrenia (SCZ).** PGS<sub>SCZ</sub> was associated with 14 phenotypes in EUR participants (Figs. 1B and 2; Supplementary Tables 12 and 21). The only associations with non-substance use phenotypes were with truancy, being suspended or expelled from school (OR = 1.21, CI = 1.13–1.28,  $p = 3.4 \times 10^{-7}$ ) in the conduct disorder category and a negative association with household income ( $\beta = -0.15$ , CI =  $-0.22$ – $-0.08$ ,  $p = 4.4 \times 10^{-5}$ ) in the demographics section.

Among substance use phenotypes, PGS<sub>SCZ</sub> was significantly associated with several alcohol use (e.g., reduction in other activities (OR = 1.22, CI = 1.16–1.29,  $p = 4.2 \times 10^{-9}$ )) and cocaine use phenotypes, such as failure to fulfill obligations (OR = 1.16, CI = 1.09–1.23,  $p = 1.4 \times 10^{-5}$ ).

**Tourette's syndrome (TS).** In EUR participants, PGS<sub>TS</sub> was associated with 1 environmental phenotype (Supplementary

Tables 13 and 21), frequency of moving/relocation as a child ( $\beta = 0.18$ , CI = 0.09–0.27,  $p = 7.8 \times 10^{-5}$ ).

### Phenome-wide analysis of somatic PGS

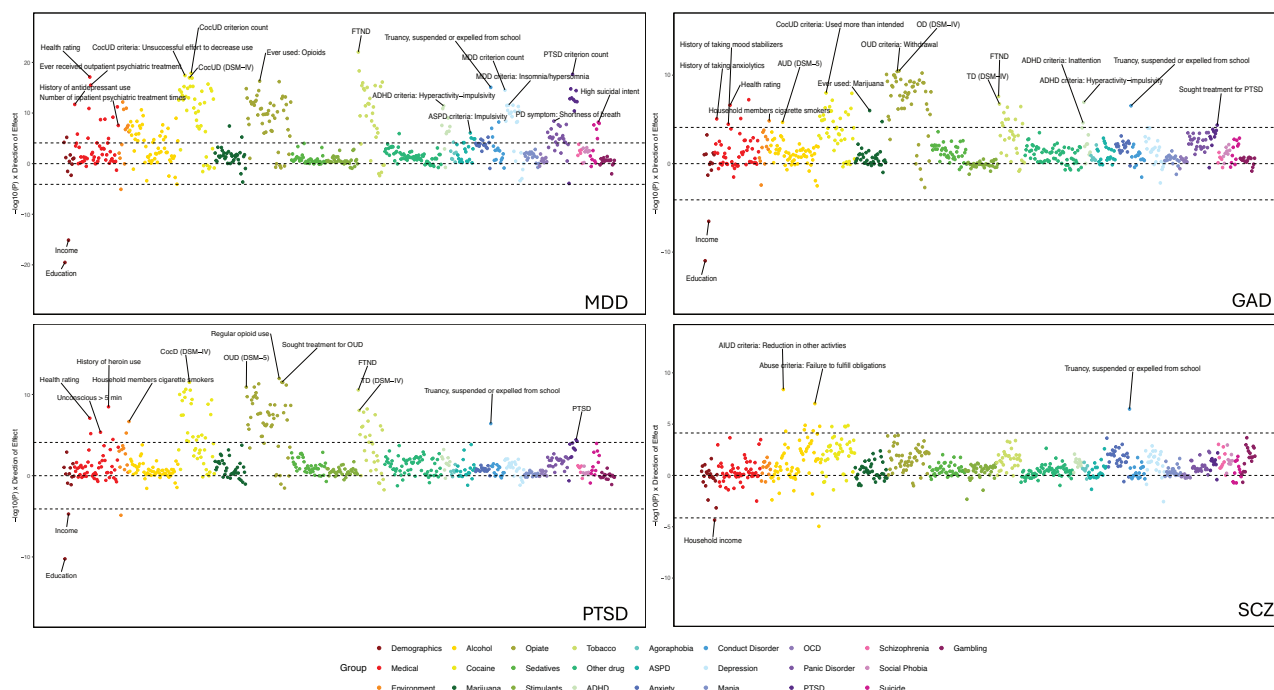
**Body-mass index (BMI).** In EUR participants, PGS<sub>BMI</sub> was associated with 138 phenotypes (Figs. 1B and 3; Supplementary Tables 18 and 21), which was most significant for the primary phenotype, BMI ( $\beta = 1.94$ , CI = 1.79–2.09,  $p = 4.1 \times 10^{-133}$ ). Three demographic variables were significant, including education ( $\beta = -0.11$ , CI =  $-0.14$ – $-0.09$ ,  $p = 1.2 \times 10^{-21}$ ). PGS<sub>BMI</sub> was associated with 6 medical phenotypes, including health rating ( $\beta = 0.13$ , CI = 0.10–0.16,  $p = 5.9 \times 10^{-17}$ ) and diabetes (OR = 1.67, CI = 1.53–1.81,  $p = 2.6 \times 10^{-12}$ ). The demographic phenotypes remained associated when BMI was used as a covariate, but half of the medical associations did not (Supplementary Fig. 4, Supplementary Table 18).

Seventeen psychiatric disorder phenotypes were associated with PGS<sub>BMI</sub>, including 10 associations with traits in the conduct disorder (e.g. truancy, being suspended or expelled from school (OR = 1.27, CI = 1.20–1.34,  $p = 1.5 \times 10^{-11}$ )) and ASPD (e.g. irritability/aggression (OR = 1.24, CI = 1.18–1.31,  $p = 6.8 \times 10^{-11}$ )) categories; hyperactivity-impulsivity ( $\beta = 0.19$ , CI = 0.11–0.26,  $p = 1.3 \times 10^{-6}$ ) in the ADHD section; and six phenotypes in the PTSD section, including PTSD diagnosis (OR = 1.24, CI = 1.14–1.34,  $p = 1.3 \times 10^{-5}$ ). When covarying for BMI, only 5 of the 17 associations remained significant.

PGS<sub>BMI</sub> was also significantly associated with 105 substance use phenotypes, including multiple SUD diagnoses. There were also numerous associations of PGS<sub>BMI</sub> with heaviness of use, withdrawal, and physiological symptoms for a variety of substances. Although PGS<sub>BMI</sub> was not associated with an AUD diagnosis, it was uniquely negatively associated with the ages of first alcohol use ( $\beta = -0.27$ , CI =  $-0.38$ – $-0.17$ ,  $p = 3.4 \times 10^{-7}$ ), regular use ( $\beta = -0.28$ , CI =  $-0.40$ – $-0.17$ ,  $p = 1.6 \times 10^{-6}$ ), and first intoxication ( $\beta = -0.25$ , CI =  $-0.35$ – $-0.14$ ,  $p = 4.6 \times 10^{-6}$ ). Although the alcohol phenotypes were no longer significant when BMI was included as a covariate, the majority of the other substance use phenotypes remained significantly associated.

PGS<sub>BMI</sub> was also associated with several environmental variables. These included exposures to substance use in childhood, such as household members being cigarette smokers (OR = 1.25, CI = 1.18–1.32,  $p = 7.8 \times 10^{-10}$ ) and frequent drug/alcohol use in the household (OR = 1.23, CI = 1.16–1.29,  $p = 1.0 \times 10^{-9}$ ). Lifetime trauma assessment (OR = 1.20, CI = 1.14–1.24,  $p = 7.9 \times 10^{-9}$ ) and childhood adversity (OR = 1.23, CI = 1.15–1.30,  $p = 3.6 \times 10^{-8}$ ) were also significantly associated with PGS<sub>BMI</sub>.

**Coronary artery disease (CAD).** In EUR participants, PGS<sub>CAD</sub> was significantly associated with 13 phenotypes (Figs. 1B and 3; Supplementary Tables 19 and 21), including the primary phenotype



**Fig. 2 Phenome-wide association results for MDD, GAD and PTSD.** Phenotype categories are plotted along the x-axis, and  $-\log_{10} p$ -value  $\times$  direction of effect is plotted on the y-axis. Selected phenotypes passing Bonferroni correction are labeled.

of heart disease ( $OR = 1.38$ ,  $CI = 1.22\text{--}1.54$ ,  $p = 4.7 \times 10^{-5}$ ), BMI ( $\beta = 0.30$ ,  $CI = 0.16\text{--}0.45$ ,  $p = 4.7 \times 10^{-5}$ ) and a negative association with education ( $\beta = -0.07$ ,  $CI = -0.09\text{--}0.04$ ,  $p = 1.3 \times 10^{-9}$ ). The remaining significant associations were with 10 tobacco use phenotypes, (e.g. FTND score ( $\beta = 0.25$ ,  $CI = 0.18\text{--}0.33$ ,  $p = 3.4 \times 10^{-11}$ )). The majority of these associations remained significant when the primary phenotype was included as a covariate (Supplementary Fig. 5; Supplementary Table 19).

**Type 2 diabetes (T2D).**  $PGS_{T2D}$  in EUR participants was significantly associated with 71 phenotypes, including diabetes ( $OR = 2.18$ ,  $CI = 2.02\text{--}2.34$ ,  $p = 7.2 \times 10^{-22}$ ) (Figs. 1B and 3; Supplementary Tables 20 and 21).  $PGS_{T2D}$  was associated with seven medical and demographic phenotypes, including BMI ( $\beta = 0.58$ ,  $CI = 0.41\text{--}0.76$ ,  $p = 3.3 \times 10^{-11}$ ) and health rating ( $\beta = 0.10$ ,  $CI = 0.07\text{--}0.13$ ,  $p = 9.3 \times 10^{-9}$ ), five of which remained significant when the primary phenotype was included as a covariate.

Truancy, being suspended or expelled from school in the conduct disorder group was the only psychiatric phenotype associated with  $PGS_{T2D}$  ( $OR = 1.17$ ,  $CI = 1.09\text{--}1.25$ ,  $p = 4.2 \times 10^{-5}$ ).  $PGS_{T2D}$  was significantly associated with 63 substance use phenotypes, including the FTND score ( $\beta = 0.29$ ,  $CI = 0.20\text{--}0.38$ ,  $p = 3.5 \times 10^{-10}$ ), DSM-5 OUD ( $OR = 1.25$ ,  $CI = 1.18\text{--}1.32$ ,  $p = 7.6 \times 10^{-10}$ ), DSM-5 CocUD ( $OR = 1.20$ ,  $CI = 1.12\text{--}1.27$ ,  $p = 8.9 \times 10^{-7}$ ), and DSM-5 CanUD ( $OR = 1.17$ ,  $CI = 1.09\text{--}1.24$ ,  $p = 5.9 \times 10^{-5}$ ). The majority of these remained significant when covarying for the primary phenotype (Supplementary Fig. 6, Supplementary Table 20).

### Comparison between PGS phenotypic associations

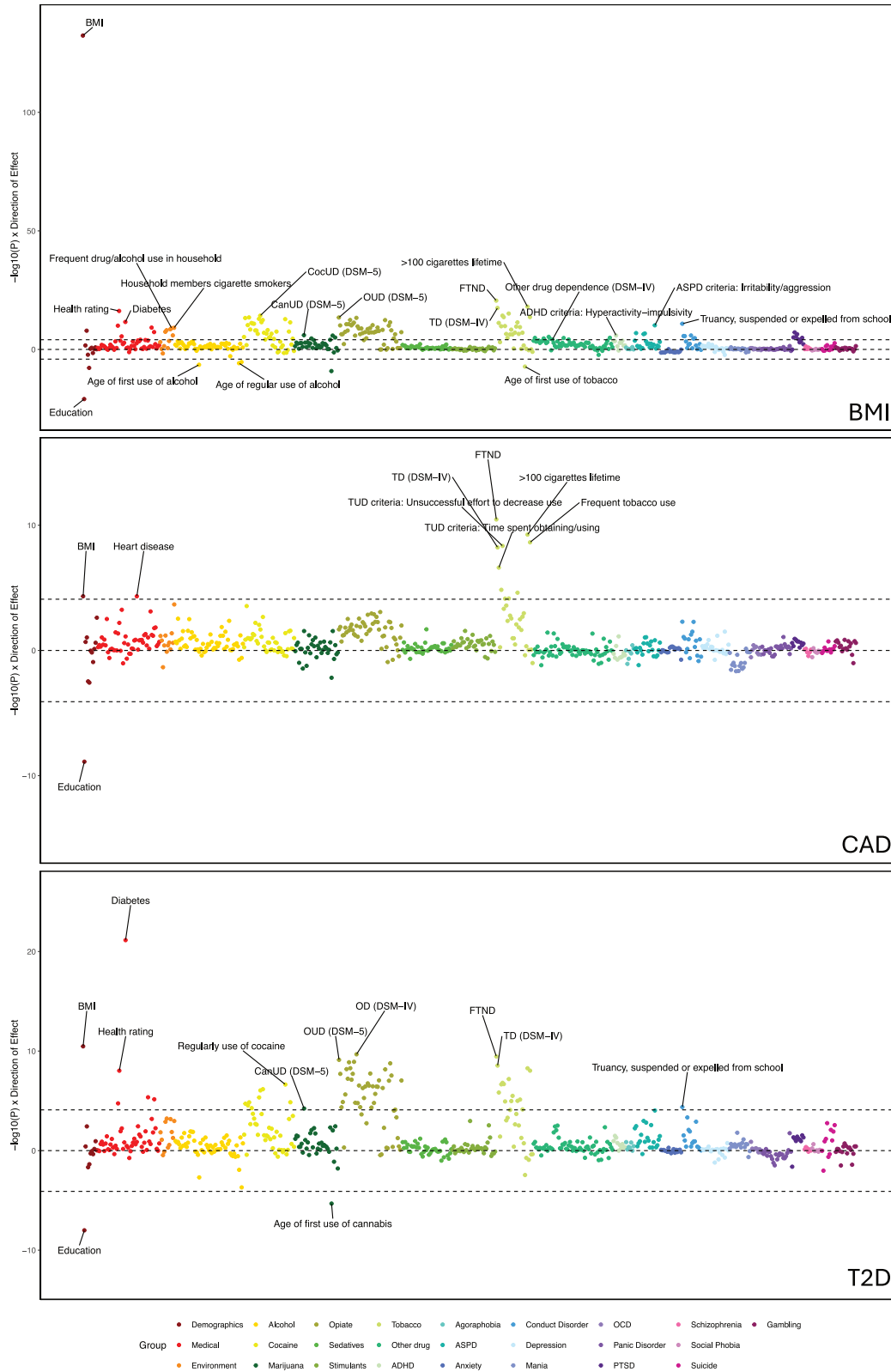
Ninety-eight phenotypes were significantly associated with two or more PGS (Fig. 4). Common demographic variables across PGS included negative associations with education and income, and positive associations with BMI. Of the medical phenotypes, health rating and number of medical problems were the most common significant associations. Several environmental variables were also associated with multiple PGS, most commonly household members using cigarettes. Interestingly,  $PGS_{MDD}$  and  $PGS_{BMI}$  share

many associated phenotypes in the environment category and in psychiatric categories, where common associations were observed for ADHD, antisocial personality disorder, conduct disorder, and PTSD phenotypes. Truancy, suspended, and expelled from school from the conduct disorder section was significant across all PGS except  $PGS_{CAD}$ . Substance use categories (alcohol, cocaine, marijuana, opiate, and tobacco use) exhibited widespread commonalities across both psychiatric and somatic PGS. Notably, DSM-5 criterion count for cocaine was associated with four PGS; whereas sum of withdrawal problems and DSM-5 criterion count for opiate use were both associated with five of the seven PGS. Numerous tobacco use phenotypes were associated with all PGS save for  $PGS_{SCZ}$ ; whereas the majority of common significant phenotypes observed for alcohol phenotypes were between  $PGS_{MDD}$  and  $PGS_{SCZ}$ .

### DISCUSSION

This study examined the performance of psychiatric and somatic PGS in the deeply-phenotyped Yale-Penn sample, in which most participants were ascertained based on having one or more lifetime SUDs. The SSADDA yields a wealth of phenotypic data not typically available in EHR-based biobanks traditionally used for this type of analysis, therefore we were able to both replicate previous findings and identify several novel cross-trait associations. For all PGS, the largest number of associations were with phenotypes in the substance use categories. This is consistent with the high prevalence of SUDs and the large number of individual traits ascertained for each substance in this sample, and highlights the high degree of pleiotropy of SUDs with both psychiatric and medical phenotypes. Also, as might be expected, compared to the somatic PGS, psychiatric disorder PGS showed more associations with phenotypes in psychiatric categories, both within and cross disorder.

Several PGS were significantly associated with their primary phenotypes. The three somatic PGS were associated with their respective primary phenotypes, indicative of the power of the PGS.  $PGS_{PTSD}$  in EUR participants was associated with a PTSD



**Fig. 3 Phenome-wide association results for BMI, CAD and T2D.** Phenotype categories are plotted along the x-axis, and  $-1\log_{10} p\text{-value} \times \text{direction of effect}$  is plotted on the y-axis. Selected phenotypes passing Bonferroni correction are labeled.

diagnosis. However, PGS for MDD and GAD in EUR participants were not associated with their respective primary diagnosis following Bonferroni correction, though both were associated with related phenotypes, such as DSM criterion count for MDD

and the use of medications to treat anxiety. The lack of association of PGS with their primary phenotypes could be due to the sample’s ascertainment strategy, which focused on the presence of one or more SUDs.



**Fig. 4** Heatmap of selected phenotypes that were common across two or more PGS.

Some PGS did not yield any significant associations. In AFR participants, only the PGS<sub>BD</sub> showed any significant associations and none were BD-related phenotypes. Although no other associations with PGS were significant, some of the AFR PGS showed nominal associations (i.e.,  $p < 0.05$ ) that may become significant with a better powered PGS derived from a larger originating GWAS (e.g., the association of PGS<sub>MDD</sub> with “ever

depressed”). Because the SSADDA interview does not assess autism, Tourette’s Syndrome, or eating disorders, primary associations for these PGS could not be tested.

PGS<sub>MDD</sub> showed the most associations of any PGS tested. Notably, it was also the only PGS to yield significant associations with depression- and suicide-related phenotypes. While other EHR-based PheWAS have demonstrated strong associations of

MDD with the primary diagnosis [59, 60], our strongest association among depression phenotypes was for the MDD criterion count. Moreover, each of the individual nine MDD diagnostic criteria were also significantly associated with  $PGS_{MDD}$ , suggesting a genetic contribution to each. Most of the associations with psychiatric phenotypes remained significant when the depression diagnosis was covaried, indicating that the associations are not due to co-occurring MDD. As with previous findings in an EHR-based PheWAS, we observed associations of  $PGS_{MDD}$  with alcohol and tobacco use phenotypes, GAD, PTSD, and agoraphobia [59]. The association between SUDs and MDD was also found in our previous analysis in this sample, which demonstrated associations between PGS for SUDs and a number of depression phenotypes [38]. Interestingly, numerous withdrawal-related phenotypes for cocaine, tobacco, opioids, and alcohol were also significantly associated with  $PGS_{MDD}$ , as were treatment-seeking for depression and other psychiatric disorders.

Few studies have examined the performance of anxiety-related PGS. One study in which a  $PGS_{PTSD}$  was tested in four EHR-based biobanks [61] showed significant associations with a PTSD diagnosis, a SUD diagnosis, and tobacco dependence, as well as numerous associations with medical conditions, including circulatory and respiratory diseases. In contrast to our findings, that study showed associations with various anxiety disorders and depression, which may have been due to the large size of the included biobanks and the higher number of cases for anxiety disorders. Our PheWAS results for both  $PGS_{PTSD}$  and  $PGS_{GAD}$  were associated with cocaine, opioid, and tobacco diagnoses, criterion counts, and treatment seeking for use of those substances, which is suggestive of an association of greater genetic risk for anxiety with greater SUD severity.

Participants who, during screening for study participation, self-reported having a schizophrenia or bipolar disorder diagnosis were excluded from the Yale-Penn sample. Thus, the lack of associations of  $PGS_{BD}$  with the primary diagnoses and related phenotypes was not unexpected; and we did not test for association between  $PGS_{SCZ}$  and schizophrenia due to low sample size ( $n = 6$ ). As with previous PheWAS,  $PGS_{SCZ}$  was associated with substance use and personality disorder phenotypes [62]. Given the high rate at which SCZ and tobacco use co-occur, and previously observed association of PGS for SCZ with tobacco use [62], the lack of associations here were unexpected and may also be attributable to the exclusion of participants with psychotic disorders from the Yale-Penn sample.

In addition to all three somatic PGS being strongly associated with their primary phenotype, they were associated with BMI and several tobacco-related phenotypes. Previous studies conducted using data from the UK Biobank and Penn Medicine BioBank showed associations of  $PGS_{BMI}$  with T2D, circulatory system disorders, and sleep problems [25, 63]. We also found associations of the  $PGS_{BMI}$  with numerous substance-related phenotypes and environmental factors. Lifetime trauma assessment, childhood adversity, and childhood exposure to substance use were also significantly associated with  $PGS_{BMI}$ , experiences that have been shown to predict higher BMI [64].  $PGS_{T2D}$ , as expected, was associated with measures of poor health and numerous substance use phenotypes, the majority of which persisted after controlling for a diabetes diagnosis. Higher rates of SUDs have been observed in individuals with T2D [65] and individuals with a SUD and T2D experience poorer medical outcomes and higher mortality than those with T2D alone [66], though little is currently known about pleiotropy of these traits. Akin to previous work [67],  $PGS_{CAD}$  was associated with tobacco use phenotypes but no other substance use, such as alcohol phenotypes, or medical disorders, such as T2D.  $PGS_{CAD}$  did not yield any associations in the psychiatric category and  $PGS_{T2D}$  only had one, which was no longer significant covarying for diabetes diagnosis, suggesting that

genetic liability for these medical disorders is not associated with psychiatric phenotypes in this sample.

This study should be interpreted in light of the strengths and limitations. The Yale-Penn dataset used as a target sample is comparatively small and cross-sectional, without longitudinal data and medical records data available in large, EHR-based genetic studies. However, the in-depth SSADDA interview provides granular psychiatric and substance use data not available in EHR-based biobanks, which provide the possibility of novel insights into the pleiotropy of co-occurring traits. The Yale-Penn sample excluded individuals with certain psychiatric illnesses, including self-reported diagnosis of schizophrenia or bipolar disorder at the time of telephone screening, thus limiting our ability to observe some associations. For the primary phenotypes that we did test, PGS for psychiatric disorders explains only a small proportion of phenotypic variance (<1.4%), although PGS for somatic traits explains a higher proportion (up to 10% for BMI). Available discovery GWAS varied in size and those that included individuals of AFR ancestry GWAS were not available for all the phenotypes of interest. Moreover, the number of participants in the originating AFR GWAS were consistently much smaller than those available for EUR. Because the Yale-Penn sample includes similar numbers of AFR and EUR participants, we believe that larger discovery GWAS in AFR participants and the accompanying increase in statistical power will be more informative of pleiotropy in non-EUR populations.

Despite these limitations, our findings demonstrate the pleiotropic nature of genetic liability for psychiatric disorders and somatic traits. Both psychiatric and somatic PGS were broadly associated with substance use phenotypes in a sample enriched for individuals with SUDs. Despite the extensive pleiotropy found, we also identified associations that were unique to specific PGS. Furthermore, psychiatric PGS were more likely to be associated with psychiatric disorders compared to somatic PGS, suggesting some level of specificity of genetic architecture within categories. Many phenotypes remained associated when covarying for the primary phenotype on which the PGS was based, suggesting that the genetic liability for the disorders in question is the primary driver of the associations. Overall, we find evidence that genetic liability for psychiatric disorders and somatic traits partially underlies the common co-occurrence of these traits with SUDs.

## DATA AVAILABILITY

Yale-Penn data are available through dbGAP (study accession numbers phs000425.v1.p1, phs000952.v1.p1, phs000277.v2.p1). Summary statistics used to calculate PRS are available from the originating studies (summarized in Supplementary Table 1).

## REFERENCES

1. Nakash O, Levav I, Aguilar-Gaxiola S, Alonso J, Andrade LH, Angermeyer MC, et al. Comorbidity of common mental disorders with cancer and their treatment gap: findings from the World Mental Health Surveys. *Psychooncology*. 2014;23:40–51.
2. Wells KB, Golding JM, Burnam MA. Psychiatric disorder in a sample of the general population with and without chronic medical conditions. *Am J Psychiatry*. 1988;145:976–81.
3. Danna SM, Graham E, Burns RJ, Deschênes SS, Schmitz N. Association between Depressive Symptoms and Cognitive Function in Persons with Diabetes Mellitus: A Systematic Review. *PLOS One*. 2016;11:e0160809.
4. Zhang MWB, Ho RCM, Cheung MWL, Fu E, Mak A. Prevalence of depressive symptoms in patients with chronic obstructive pulmonary disease: a systematic review, meta-analysis and meta-regression. *Gen Hosp Psychiatry*. 2011;33:217–23.
5. Esposito M, Gallai B, Roccella M, Marotta R, Lavano F, Lavano SM, et al. Anxiety and depression levels in prepubertal obese children: a case-control study. *Neuropsychiatr Dis Treat*. 2014;10:1897–902.
6. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2020 National



- Survey on Drug Use and Health (HHS Publication No. PEP21-07-01-003, NSDUH Series H-56), p.A-48. 2021.
7. Davis L, Uezato A, Newell JM, Frazier E. Major depression and comorbid substance use disorders. *Curr Opin Psychiatry*. 2008;21:14–18.
  8. Drake RE, Mueser KT. Co-Occurring Alcohol Use Disorder and Schizophrenia. *Alcohol Res Health*. 2002;26:99–102.
  9. Momen NC, Plana-Ripoll O, Agerbo E, Christensen MK, Iburg KM, Laursen TM, et al. Mortality Associated With Mental Disorders and Comorbid General Medical Conditions. *JAMA Psychiatry*. 2022;79:444–53.
  10. Solmi M, Soardo L, Kaur S, Azis M, Cabras A, Censori M, et al. Meta-analytic prevalence of comorbid mental disorders in individuals at clinical high risk of psychosis: the case for transdiagnostic assessment. *Mol Psychiatry*. 2023;28:2291–2300.
  11. Lindsey WT, Stewart D, Childress D. Drug interactions between common illicit drugs and prescription therapies. *Am J Drug Alcohol Abus*. 2012;38:334–43.
  12. Dual Disorders: Counseling Clients with Chemical Dependency and Mental Illness, Daley DC, Moss HB - Google Books. 2024. [https://books.google.com/books?hl=en&lr=&id=nzLXDQAAQBAJ&oi=fnd&pg=PT16&ots=3W00uBu13g&sig=TYg\\_nHqhIndV-4U8uT3xlHX36y4](https://books.google.com/books?hl=en&lr=&id=nzLXDQAAQBAJ&oi=fnd&pg=PT16&ots=3W00uBu13g&sig=TYg_nHqhIndV-4U8uT3xlHX36y4). Accessed 9 January 2024.
  13. Grella CE, Hser YI, Joshi V, Rounds-Bryant J. Drug treatment outcomes for adolescents with comorbid mental and substance use disorders. *J Nerv Ment Dis*. 2001;189:384–92.
  14. Magura S, Rosenblum A, Fong C. Factors associated with medication adherence among psychiatric outpatients at substance abuse risk. *Open Addict J*. 2011;4:58–64.
  15. Greco LA, Reay WR, Dayas CV, Cairns MJ. Pairwise genetic meta-analyses between schizophrenia and substance dependence phenotypes reveals novel association signals with pharmacological significance. *Transl Psychiatry*. 2022;12:403.
  16. Abdellaoui A, Smit DJA, van den Brink W, Denys D, Verweij KJH. Genomic relationships across psychiatric disorders including substance use disorders. *Drug Alcohol Depend*. 2021;220:108535.
  17. Sanchez-Roige S, Kember RL, Agrawal A. Substance use and common contributors to morbidity: A genetics perspective. *EBioMedicine*. 2022;83:104212.
  18. Vassy JL, Posner DC, Ho Y-L, Gagnon DR, Galloway A, Tanukonda V, et al. Cardiovascular Disease Risk Assessment Using Traditional Risk Factors and Polygenic Risk Scores in the Million Veteran Program. *JAMA Cardiol*. 2023;8:564–74.
  19. Khan A, Turchin MC, Patki A, Srinivasasainagendra V, Shang N, Nadukuru R, et al. Genome-wide polygenic score to predict chronic kidney disease across ancestries. *Nat Med*. 2022;28:1412–20.
  20. Kember RL, Vickers-Smith R, Xu H, Toikumo S, Niarchou M, Zhou H, et al. Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. *Nat Neurosci*. 2022;25:1279–87.
  21. 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.
  22. Toikumo S, Vickers-Smith R, Jinwala Z, Xu H, Saini D, Hartwell E, et al. The genetic architecture of pain intensity in a sample of 598,339 U.S. veterans. *Nat Med*. 2024;30:1075–84.
  23. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinforma Oxf Engl*. 2010;26:1205–10.
  24. Hartwell EE, Merikangas AK, Verma SS, Ritchie MD, Regeneron Genetics Center, Kranzler HR, et al. Genetic liability for substance use associated with medical comorbidities in electronic health records of African- and European-ancestry individuals. *Addict Biol*. 2022;27:e13099.
  25. Kember RL, Verma SS, Xiao B, Lucas A, Kripke CM, et al. Polygenic risk scores for cardiometabolic traits demonstrate importance of ancestry for predictive precision medicine. *Pac Symp Biocomput Pac Symp Biocomput*. 2024;29:611–26.
  26. Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, et al. Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug Alcohol Depend*. 2005;80:303–12.
  27. Pierucci-Lagha A, Gelernter J, Chan G, Arias A, Cubells JF, Farrer L, et al. Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend*. 2007;91:85–90.
  28. Sartor CE, Wang Z, Xu K, Kranzler HR, Gelernter J. The joint effects of ADH1B variants and childhood adversity on alcohol related phenotypes in African-American and European-American women and men. *Alcohol Clin Exp Res*. 2014;38:2907–14.
  29. 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. *Addict Abingdon Engl*. 2023;118:1942–52.
  30. Zhou H, Cheng Z, Bass N, Krystal JH, Farrer LA, Kranzler HR, et al. Genome-wide association study identifies glutamate ionotropic receptor GRIA4 as a risk gene for comorbid nicotine dependence and major depression. *Transl Psychiatry*. 2018;8:208.
  31. Gelernter J, Panhuysen C, Wilcox M, Hesselbrock V, Rounsaville B, Poling J, et al. Genomewide linkage scan for opioid dependence and related traits. *Am J Hum Genet*. 2006;78:759–69.
  32. Gelernter J, Kranzler HR, Sherva R, Koesterer R, Almasy L, Zhao H, et al. Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. *Biol Psychiatry*. 2014;76:66–74.
  33. Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, et al. Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol Psychiatry*. 2014;19:717–23.
  34. Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, et al. Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. *Am J Med Genet Part B Neuropsychiatr Genet Publ Int Soc Psychiatr Genet*. 2005;136B:45–52.
  35. Gelernter J, Kranzler HR, Panhuysen C, Weiss RD, Brady K, Poling J, et al. Dense genomewide linkage scan for alcohol dependence in African Americans: significant linkage on chromosome 10. *Biol Psychiatry*. 2009;65:111–5.
  36. Gelernter J, Kranzler H, Sherva R, Almasy L, Koesterer R, Smith A, et al. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry*. 2014;19:41–49.
  37. Peer K, Rennert L, Lynch KG, Farrer L, Gelernter J, Kranzler HR. Prevalence of DSM-IV and DSM-5 alcohol, cocaine, opioid, and cannabis use disorders in a largely substance dependent sample. *Drug Alcohol Depend*. 2013;127:215–9.
  38. Kember RL, Hartwell EE, Xu H, Rotenberg J, Almasy L, Zhou H, et al. Phenome-wide Association Analysis of Substance Use Disorders in a Deeply Phenotyped Sample. *Biol Psychiatry*. 2023;93:536–45.
  39. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–7.
  40. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
  41. Gelernter J, Kranzler HR, Sherva R, Almasy L, Herman AI, Koesterer R, et al. Genome-Wide Association Study of Nicotine Dependence in American Populations: Identification of Novel Risk Loci in Both African-Americans and European-Americans. *Biol Psychiatry*. 2015;77:493–503.
  42. Sherva R, Wang Q, Kranzler H, Zhao H, Koesterer R, Herman A, et al. Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry*. 2016;73:472–80.
  43. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10:1776.
  44. Watson HJ, Yilmaz Z, Thornton LM, Hübel C, Coleman JRI, Gaspar HA, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet*. 2019;51:1207–14.
  45. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51:431–44.
  46. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, 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.
  47. Levey DF, Gelernter J, Polimanti R, Zhou H, Cheng Z, Aslan M, et al. Reproducible Genetic Risk Loci for Anxiety: Results From ~200,000 Participants in the Million Veteran Program. *Am J Psychiatry*. 2020;177:223–32.
  48. Als TD, Kurki MI, Grove J, Voloudakis G, Therrien K, Tasanko E, et al. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat Genet*. 2023;29:1832–44.
  49. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OC GAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry*. 2018;23:1181–8.
  50. Forstner AJ, Awasthi S, Wolf C, Maron E, Erhardt A, Czamara D, et al. Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Mol Psychiatry*. 2021;26:4179–90.
  51. Stein MB, Levey DF, Cheng Z, Wendt FR, Harrington K, Pathak GA, et al. Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program. *Nat Genet*. 2021;53:174–84.
  52. Trubetskoy V, Pardinas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604:502–8.
  53. Yu D, Sul JH, Tsetsos F, Nawaz MS, Huang AY, Zelaya I, et al. Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry*. 2019;176:217–27.

54. 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 ~700,000 individuals of European ancestry. *Hum Mol Genet.* 2018;27:3641–9.
55. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011;43:333–8.
56. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet.* 2022;54:560–72.
57. Bigdeli TB, Fanous AH, Li Y, Rajeevan N, Sayward F, Genovese G, et al. Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. *Schizophr Bull.* 2021;47:517–29.
58. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating Missing Heritability for Disease from Genome-wide Association Studies. *Am J Hum Genet.* 2011;88:294–305.
59. Fang Y, Fritsche LG, Mukherjee B, Sen S, Richmond-Rakerd LS. Polygenic Liability to Depression Is Associated With Multiple Medical Conditions in the Electronic Health Record: Phenome-wide Association Study of 46,782 Individuals. *Biol Psychiatry.* 2022;92:923–31.
60. Mulugeta A, Zhou A, King C, Hyppönen E. Association between major depressive disorder and multiple disease outcomes: a phenome-wide Mendelian randomisation study in the UK Biobank. *Mol Psychiatry.* 2020;25:1469–76.
61. Pathak GA, Singh K, Choi KW, Fang Y, Kouakou MR, Lee YH, et al. Genetic Liability to Posttraumatic Stress Disorder Symptoms and Its Association With Cardiometabolic and Respiratory Outcomes. *JAMA Psychiatry.* 2024;81:34–44.
62. Zheutlin AB, Dennis J, Karlsson Linnér R, Moscatti A, Restrepo N, Straub P, et al. Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *Am J Psychiatry.* 2019;176:846–55.
63. Hyppönen E, Mulugeta A, Zhou A, Santhanakrishnan VK. A data-driven approach for studying the role of body mass in multiple diseases: a phenome-wide registry-based case-control study in the UK Biobank. *Lancet Digit Health.* 2019;1:e116–e126.
64. Fogelman N, Magin Z, Hart R, Sinha R. A Longitudinal Study of Life Trauma, Chronic Stress and Body Mass Index on Weight Gain over a 2-Year Period. *Behav Med Wash DC.* 2022;48:162–70.
65. Wu L-T, Ghitza UE, Batch BC, Pencina MJ, Rojas LF, Goldstein BA, et al. Substance use and mental diagnoses among adults with and without type 2 diabetes: Results from electronic health records data. *Drug Alcohol Depend.* 2015;156:162–9.
66. Winhusen T, Theobald J, Kaelber DC, Lewis D. Medical complications associated with substance use disorders in patients with type 2 diabetes and hypertension: electronic health record findings. *Addict Abingdon Engl.* 2019;114:1462–70.
67. Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med.* 2022;28:1679–92.
68. Major Depression - National Institute of Mental Health (NIMH). <https://www.nimh.nih.gov/health/statistics/major-depression>. Accessed 6 June 2024.
69. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shiralil M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci.* 2019;22:343–52.
70. Panic Disorder - National Institute of Mental Health (NIMH). <https://www.nimh.nih.gov/health/statistics/panic-disorder>. Accessed 6 June 2024.
71. Post-Traumatic Stress Disorder (PTSD) - National Institute of Mental Health (NIMH). <https://www.nimh.nih.gov/health/statistics/post-traumatic-stress-disorder-ptsd>. Accessed 6 June 2024.
72. Huang J, Huffman JE, Huang Y, Do Valle Í, Assimes TL, Raghavan S, et al. Genomics and phenomics of body mass index reveals a complex disease network. *Nat Commun.* 2022;13:7973.
73. The National Academies, Committee on a National Surveillance System for Cardiovascular and Select Chronic Diseases. Cardiovascular Disease. In: A Nationwide Framework for Surveillance of Cardiovascular and Chronic Lung Diseases. Washington, DC: National Academies Press (US); 2011.
74. Patel AP, Wang M, Ruan Y, Koyama S, Clarke SL, Yang X, et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat Med.* 2023;29:1793–803.
75. CDC. National Diabetes Statistics Report. Diabetes. 2024. <https://www.cdc.gov/diabetes/php/data-research/index.html>. Accessed 6 June 2024.

## AUTHOR CONTRIBUTIONS

EEH and RLK conceived and designed the study; ZJ, JM, SR analyzed the data; JG and HRK acquired the data; EEH, ZJ, JM, SR, RLK interpreted the data and drafted the manuscript; all authors revised the manuscript for intellectual content and provided final approval for submission.

## FUNDING

This study was funded by Department of Veterans Affairs grants IK2 CX002336 (EEH), I01 BX004820 (HRK), VISN1 MIRECC (JG), and the VISN4 MIRECC (HRK, RLK, EEH), and NIH grants R01DA037974 (JG), R01DA058862 (JG), K01AA028292 (RLK).

## COMPETING INTERESTS

HRK is a member of advisory boards for Dicerna Pharmaceuticals, Sophrosyne Pharmaceuticals, Enthion Pharmaceuticals, and Clearmind Medicine; a consultant to Sobrera Pharmaceuticals; the recipient of research funding and medication supplies for an investigator-initiated study from Alkermes; and a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last three years by Alkermes, Dicerna, Ethypharm, Lundbeck, Mitsubishi, Otsuka, and Pear Therapeutics. JG and HRK hold U.S. patent 10,900,082 titled: "Genotype-guided dosing of opioid agonists," issued 26 January 2021. Other authors have nothing to disclose.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41386-024-01922-2>.

**Correspondence** and requests for materials should be addressed to Rachel L. Kember.

**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.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024