Combining Transdiagnostic and Disorder-Level GWAS Enhances Precision of Psychiatric Genetic Risk Profiles in a Multi-Ancestry Sample

Yousef Khan[1]*, Christal N. Davis[1,2]*, Zeal Jinwala[1], Kyra L. Feuer[1], Sylvanus Toikumo[1,2], Emily E. Hartwell[1,2], Sandra Sanchez-Roige[3,4,5], Roseann E. Peterson[6], Alexander S. Hatoum[7], Henry R. Kranzler[1,2], Rachel L. Kember[1,2]

[1] Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA 19104
[2] Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia, PA 19104
[3] Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, United States
[4] Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37235, United States
[5] Genomic Medicine, University of California San Diego, La Jolla, CA, USA
[6] Institute for Department of Psychiatry and Behavioral Sciences, Institute for Genomics in Health, SUNY Downstate Health Sciences University, Brooklyn, NY 11203, United States
[7] Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO 63130, United States
* Authors made equal contributions.

Corresponding Author:
Rachel L. Kember, PhD, 3535 Market Street, Suite 500, Philadelphia, PA 19104. Email: rkember@pennmedicine.upenn.edu

Dr. Kranzler 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; 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; and a holder of U.S. patent 10,900,082 titled: "Genotype-guided dosing of opioid agonists," issued 26 January 2021.

Abstract

The etiology of substance use disorders (SUDs) and psychiatric disorders reflects a combination of both transdiagnostic (i.e., common) and disorder-level (i.e., independent) genetic risk factors. We applied genomic structural equation modeling to examine these genetic factors across SUDs, psychotic, mood, and anxiety disorders using genome-wide association studies (GWAS) of European- (EUR) and African-ancestry (AFR) individuals. In EUR individuals, transdiagnostic genetic factors represented SUDs (143 lead single nucleotide polymorphisms [SNPs]), psychotic (162 lead SNPs), and mood/anxiety disorders (112 lead SNPs). We identified two novel SNPs for mood/anxiety disorders that have probable regulatory roles on *FOXP1*, *NECTIN3*, and *BTLA* genes. In AFR individuals, genetic factors represented SUDs (1 lead SNP) and psychiatric disorders (no significant SNPs). The SUD factor lead SNP, although previously significant in EUR- and cross-ancestry GWAS, is a novel finding in AFR individuals. Shared genetic variance accounted for overlap between SUDs and their psychiatric comorbidities, with second-order GWAS identifying up to 12 SNPs not significantly associated with either first-order factor in EUR individuals. Finally, common and independent genetic effects showed different associations with psychiatric, sociodemographic, and medical phenotypes. For example, the independent components of schizophrenia and bipolar disorder had distinct associations with affective and risk-taking behaviors, and phenome-wide association studies identified medical conditions associated with tobacco use disorder independent of the broader SUDs factor. Thus, combining transdiagnostic and disorder-level genetic approaches can improve our understanding of co-occurring conditions and increase the specificity of genetic discovery, which is critical for psychiatric disorders that demonstrate considerable symptom and etiological overlap.

Introduction

Substance use disorders (SUDs) commonly co-occur with mood, anxiety, and psychotic disorders.[1-3] For example, about one-quarter of individuals with major depressive disorder (MDD) meet criteria for at least one SUD,[1] and 20-80% of individuals seeking SUD treatment have MDD.[4] Similarly, the prevalence of anxiety disorders among individuals with an illicit SUD is almost three times that of the general population.[2] Additionally, nearly half (42%) of individuals experiencing a first episode of psychosis[5] and a third of those with bipolar disorder[3] (BD) have a co-occurring SUD. Such comorbidity complicates the clinical course of affected individuals, resulting in greater healthcare and other costs.[6,7]

Large-scale genome-wide association studies (GWAS)[8-15] have shown these disorders to be highly polygenic, with individual variants exerting a small influence on risk. These studies also provide evidence of pleiotropy, whereby variants are associated with more than one psychiatric trait. These pleiotropic effects partially account for the co-occurrence of psychiatric traits and disorders.[8-15] Genomic structural equation modeling (gSEM) can leverage this shared liability to identify common genetic factors that underlie multiple disorders.[16] In combination with downstream analyses, these multivariate genetic approaches can improve our understanding of the etiology of commonly co-occurring conditions by identifying shared biological pathways and possible risk mechanisms.

Several gSEM studies have modeled pleiotropy across psychiatric disorders in European-ancestry (EUR) individuals. The first modeled shared genetic variance across schizophrenia (SCZ), BD, MDD, posttraumatic stress disorder (PTSD), and anxiety, identifying a single common factor.[16] Across eight psychiatric disorders—attention-deficit hyperactivity disorder, anorexia nervosa, autism spectrum disorder, BD, MDD, obsessive compulsive disorder, SCZ,

and Tourette syndrome[17]–three underlying genetic factors representing mood and psychotic disorders, early-onset neurodevelopmental disorders, and compulsive disorders—best represented the data. More recent work expanded this model by adding problematic alcohol use, anxiety disorders, and PTSD,[18] which resulted in the previously combined mood and psychotic disorders factor splitting into separate internalizing and psychotic factors.

Similarly, an addiction factor has been identified that underlies cannabis use disorder (CanUD), opioid use disorder (OUD), and measures of problematic alcohol and tobacco use.[19] Although this study included African-ancestry (AFR) in addition to EUR individuals, limited power precluded the application of gSEM in AFR individuals. Instead, an alternative approach, ASSET,[20] was taken to identify pleiotropic effects among AFR individuals. However, this approach failed to identify any variants having effects common to all four substance use behaviors in AFR individuals. Other research applying gSEM has derived a common factor underlying measures of substance use in EUR individuals,[21] reaffirming the shared genetic basis underlying multiple substance use behaviors or disorders.

Although research has demonstrated consistency in the genetic factor structure across models of substance use and psychiatric disorders, these studies have applied gSEM only in EUR individuals. The modeling of complex genetic relationships via structural equation modeling imposes greater demands for statistical power than simpler univariate GWAS analyses, which themselves are often underpowered in AFR and other non-EUR ancestry groups. Thus, the statistical power requirements of the approach, whose strength lies in its ability to enhance our understanding of the complexity of genetic relations, have hampered its application in non-EUR ancestries.

Previous studies have also typically focused on identifying transdiagnostic genetic risk, but gSEM can also enable more precise identification of disorder-specific genetic mechanisms than individual GWAS. In combination with transdiagnostic genetic approaches, GWAS-by-subtraction[22] can parse associations with single nucleotide polymorphisms (SNPs) into those that influence risk for a disorder through a common genetic factor from those that operate independently of the common factor. The two resulting genetic dimensions can be used to differentiate the associations of a common genetic factor with psychiatric, medical, and social phenotypes from the associations of genetic risk that operates independently of the common factor. Combining transdiagnostic and disorder-level approaches enhances statistical power to detect pleiotropic effects, while identifying patterns of genetic heterogeneity and increasing the specificity of SNP discovery.[23]

To extend previous study findings, we used gSEM to characterize the underlying genetic structure of SUDs, psychotic, and mood and anxiety disorders in EUR and AFR individuals. First, we examined the genetic factor structure of the disorders using exploratory and confirmatory factor analyses, and then we explored each factor's biological underpinnings by conducting GWAS. Next, we investigated the shared genetic comorbidity of SUDs with psychotic, mood, and anxiety disorders using a second-order gSEM approach, which involves examining the relationships between the lower-level factors to identify genetic risks that are shared across factors. Finally, we characterized the common and independent genetic variance for select disorders using GWAS-by-subtraction. Thus, we sought to address two critical gaps in previous research: first, by applying gSEM among AFR individuals, and second, by employing a hierarchical approach to explore both genetic specificity and transdiagnostic genetic risk.

Ultimately, these techniques were applied to obtain a more comprehensive understanding of the genetic underpinnings of SUDs, psychotic, mood, and anxiety disorders.

Methods

To explore the genetic relationships among SUDs, psychotic, mood, and anxiety disorders, we used gSEM, which is a statistical technique that integrates summary-level genetic data with structural equation modeling. By estimating the genetic covariance structure among traits, gSEM enables the explicit and flexible modeling of complex relationships.[16]

[Figure 1 here]

Summary Statistics

We used large discovery GWAS for SUDs, psychotic, mood, and anxiety disorders in EUR and AFR individuals (Supplementary Tables 1-4) as inputs for gSEM. Genetic ancestry was determined for each input GWAS by the researchers conducting the original study, and these inferences should not be considered proxies for either race or ethnicity. Instead, the groups represent statistical determinations of ancestral similarity, within which there remains heterogeneity. The traits included four SUDs: alcohol use disorder (AUD),[13,24] tobacco use disorder (TUD),[15] cannabis use disorder (CanUD),[12] and opioid use disorder (OUD);[11] two disorders that can include psychotic features: BD[10,25] and SCZ;[9,25] and two mood and anxiety traits: anxiety (ANX)[8,26,27] and MDD.[14,28] In EUR individuals, we used multi-trait analysis of GWAS (MTAG)[29] to enhance the power to detect effects associated with a broad spectrum of anxiety disorders, including generalized anxiety disorder, panic disorder, social phobia, agoraphobia, and specific phobias (see Supplementary Materials and Supplementary Table 5). The resulting ANX summary statistics produced by MTAG were then used as an input for gSEM.

[Table 1 here]

Prior to common factor modeling in accordance with the procedure for performing gSEM, we calculated genetic correlations between the disorders using linkage disequilibrium score regression (LDSC) implemented in GenomicSEM 0.0.5c.[16] To prevent downward bias in LDSC estimates, when SNP-level sample sizes were not available within each set of summary statistics, we calculated the effective sample size using the sum of effective sample sizes across the input GWAS cohorts (see Supplementary Materials).[30] For EUR analyses, SNPs were restricted to those contained within the EUR HapMap3 reference panel[31] with a minor allele frequency (MAF) > 0.01. We then performed LDSC using EUR 1000 Genomes Phase 3 linkage disequilibrium (LD) scores.[32] Given the statistical challenges associated with including non-EUR individuals in gSEM analyses due in part to differences in LD structure and admixture, we compared the performance of three sets of LD scores before performing LDSC in AFR individuals. We restricted each set of LD scores to well-imputed SNPs with MAF > 0.01 and then compared several parameters, including LD score distribution, LD block length, and the stability of SNP-based heritability and genetic correlation estimates to determine the optimal approach (see Supplementary Materials). This step ensured the selection of an appropriately matched reference panel to avoid biasing estimates. Ultimately, we selected the Pan-UKB AFR reference data,[33] which enabled the best performance of gSEM models among AFR individuals.

Genomic Structural Equation Modeling

Consistent with best approaches for gSEM, we performed exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) on independent data to evaluate the reliability of our results. We first performed EFA on the odd chromosomes using LDSC matrices derived from the summary statistics of the previously described SUDs and psychiatric disorders to evaluate the optimal number of factors and the loadings of each disorder in a hypothesis-free manner. We

used the *Matrix* R package to correct for the possibility of a non-positive definite matrix[34] within the LDSC output and then performed EFA for 1-4 latent factors using the *lavaan* R package and a promax rotation.[35] Following EFA, we examined model fit indices (i.e., chi-square value, Akaike information criterion (AIC), comparative fix index (CFI), and standardized root mean squared residual (SRMR)) and eigenvalues to determine the optimal model.[36] Traits with a loading ≥0.35 on each latent factor were retained for confirmatory factor analysis (CFA), which was performed on the even chromosomes to avoid overfitting the data. Because of the limited statistical power in AFR models, factor analyses were performed on all chromosomes.

Following CFA, we prepared the input summary statistics for GWAS by standardizing coefficients and standard error (SE) values, such that SNP effects were scaled similarly for binary and continuous phenotypes. For quality control, we included only SNPs with MAF>0.01 and imputation score>0.60. In GWAS, we regressed each SNP on each latent variable using diagonally weighted least squares (DWLS) estimation and standard genomic control. We calculated the effective sample size of each resulting common factor GWAS as described by Mallard et al. (2022).[37]

Next, we constructed second-order common factor models to capture genetic effects that account for co-occurrence among SUDs and their psychiatric comorbidities. For both the EUR and AFR analyses, to ensure identification of the second-order models, we set the loadings of each first-order factor onto its respective second-order factor equal to the square root of their genetic correlation.[38] We then ran GWAS on each second-order factor using the procedure described for the first-order common factor GWAS.

To identify significant independent SNPs from GWAS, we performed LD clumping with PLINK 1.9[39] using an r^2 threshold of 0.1 and physical distance threshold of 3000 kb

(Supplementary Materials). For loci not previously significantly associated with a corresponding SUD, psychotic, mood, or anxiety disorder (hereafter referred to as novel), we performed PheWAS on the lead SNP to examine its associations across the phenomic spectrum using the GWAS Atlas (Supplementary Materials).[40] Additionally, for any novel lead SNPs identified, we used the LD-based Probabilistic Identification of Causal SNPs (PICS) v2.1.1 finemapping tool to assess their potential as the most likely causal variant to be responsible for the observed association in a given locus.

After performing each of the first- and second-order common factor GWAS, we calculated Q_{SNP}, a measure of heterogeneity that tests the null hypothesis that a SNP's effects operate entirely through a common factor. For example, a SNP that primarily influences SUDs through its effects on a single disorder, like TUD, should violate the null hypothesis. To identify SNPs with heterogeneous effects, we examined associations between each SNP and common factor via a common pathway model. Then, separately for each factor, we fit an independent pathway model in which the SNP predicted each of the factor's indicators. We performed a chi-square difference test on the two models (Supplementary Figure 1) and removed SNPs with $p < 5 \times 10^{-8}$ (Supplementary Tables 6-10) from the factor's summary statistics prior to conducting all post-GWAS analyses.

GWAS-by-Subtraction

To ensure that the GWAS-by-subtraction models were informative and adequately powered, we performed these analyses only on disorders with a standardized unexplained variance ≥ 0.30 in the first-order CFA. Following the paradigm of Demange et al. (2021),[22] we first specified two latent genetic factors. On the first factor, we loaded only the specific disorder of interest (i.e., depending on the model, TUD, BD, or SCZ), while on the other factor we loaded

the specific disorder *and* the other disorders included in its broader common factor. Then, we imposed two constraints by setting to 0 the covariance between: (1) the disorder of interest and all other disorders, and (2) the two latent genetic factors. Finally, to capture genetic effects on the disorder that were not correlated with the common factor, we regressed each SNP on the latent factor with a sole loading for the disorder of interest. To identify genetic effects on the disorder of interest that operated through the common factor, each SNP was regressed on the latent factor on which all common factor traits were loaded. Due to the limited statistical power of the AFR models (max N_{eff} = 6,421), we restricted GWAS-by-subtraction analyses to EUR models. The effective sample size calculations for each of the GWAS-by-subtraction models were adjusted to account for the fact that the GWAS modeled residual heritability.[22]

Biological Characterization

We used Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) version 1.6.0[41] to conduct post-GWAS analyses of each GWAS (i.e., first-order common factors in EUR and AFR individuals, second-order common factors in EUR and AFR individuals, and GWAS-by-subtraction models in EUR individuals). Gene-based tests, gene-set enrichment, and gene-tissue expression analyses were conducted using MAGMA version 1.08.[42] We examined gene expression in BrainSpan[43] and GTEx v8[44] tissue samples. SNP-to-gene associations and gene annotations were evaluated using: (1) expression quantitative trait loci (eQTLs) from PsychENCODE[45] and GTEx v8[44] brain tissue samples and (2) chromatin interactions via Hi-C data for the dorsolateral prefrontal cortex, hippocampus, ventricles, and neural progenitor cells.[46] To annotate the protein products and investigate protein-protein interactions of MAGMA-identified genes, we used the STRING database v12.0 and applied its default parameters.[47] Enrichment of protein-protein interactions was calculated as the observed

number of edges (i.e., interactions) divided by the expected number of edges in the protein network. Significant enrichment would suggest that the proteins that are encoded by genes associated with a factor participate in common pathways relevant to that factor.

Genetic Correlations with gSEM Factors

Genetic correlations between gSEM output summary statistics and other traits were calculated using LDSC[48,49] with 1000 Genomes Project phase 3[32] (for EUR) and PanUKB[33] (for AFR) data as LD references. For first- and second-order common factors in EUR individuals, we used the Complex-Traits Genetics Virtual Lab[50] to calculate batch genetic correlations with 1,437 traits across a wide variety of domains assessed via International Classification of Diseases (ICD) codes and self-report. In AFR individuals, genetic correlations were calculated for selected psychiatric and medical phenotypes as the Complex-Traits Genetics Virtual Lab does not currently facilitate LDSC in non-EUR ancestries. For GWAS-by-subtraction models in EUR individuals, we calculated genetic correlations with a selection of relevant psychiatric, social, and physical traits to facilitate comparison of the common and independent genetic effects associated with each disorder. We applied a Benjamini-Hochberg false discovery rate (FDR) correction to each set of genetic correlation analyses to account for multiple testing. Lastly, we calculated trans-ancestry genetic correlations using the regression fit method to compare gSEM common factors across EUR and AFR ancestry individuals with Popcorn v1.0[51] and ancestry-matched 1000 Genomes reference files. Reference files were prepared by excluding the MHC region, and Popcorn was used to compute LD scores for both populations.

Polygenic Score-Based PheWAS

Prior to calculating polygenic scores (PGS) in the Penn Medicine Biobank (PMBB), GWAS for the first-order, second-order, and TUD GWAS-by-subtraction models were re-run

263 excluding PMBB to ensure independence of the GWAS and target samples. We calculated PGS
264 from gSEM output summary statistics using PRS-CSx[52] and conducted a phenome-wide
265 association study (PheWAS) in the PMBB. PheWAS is a hypothesis-free approach to explore the
266 association between genetic variants and traits across the spectrum of human disease and health.
267 PMBB participants are recruited through the University of Pennsylvania Health System and
268 provide access to their electronic health record (EHR) and blood or tissue samples.[53] Genotyping
269 was performed using the Illumina Global Screening Array. Quality control procedures included
270 removing SNPs with marker call rates <95% and sample call rates <90%, as well as individuals
271 with sex discrepancies. Imputation was performed using Eagle2 (Reference-based phasing using
272 the Haplotype Reference Consortium panel) and Minimac4 on the TOPMed Imputation Server.
273 In the case of related individuals (pi-hat threshold ≥ 0.25), one from each pair was removed
274 from analyses. Genetic ancestry was determined using quantitative discriminant analysis of
275 principal components (PCs) using smartpca.[54,55] These procedures resulted in 10,383 AFR
276 individuals and 29,355 EUR individuals for inclusion in the PheWAS.
277 ICD-9 and ICD-10 codes were gathered from EHR and mapped to phecodes. Cases were
278 individuals with at least two instances of a given ICD code ("phecodes"). PGS were
279 standardized, and PheWAS was conducted by fitting a logistic regression predicting each
280 phecode from the PGS, with sex, age, and the top 10 PCs included as covariates using the
281 *PheWAS* package[56] in R. We used a Benjamini-Hochberg FDR corrected p-value to ascertain
282 significant associations.

283 **Results**

284 **Genetic Correlations among Input GWAS**

285 ***European Ancestry.*** Genetic correlations among SUDs ranged from 0.60 (SE = 0.06, $p <$
286 0.001; TUD and OUD) to 0.92 (SE = 0.05, $p < 0.001$; AUD and OUD). MDD and ANX were
287 strongly genetically correlated ($r_g = 0.91$, SE = 0.04, $p < 0.001$), as were BD and SCZ ($r_g = 0.68$,
288 SE = 0.03, $p < 0.001$). Furthermore, SUDs were significantly genetically correlated with the
289 other psychiatric disorders, ranging from 0.08 (SE = 0.02, $p < 0.001$; TUD and BD) to 0.51 (SE
290 = 0.03, $p < 0.001$; AUD and MDD; Supplementary Figure 2 and Supplementary Table 11).
291 ***African Ancestry.*** All four SUDs exhibited significant genetic correlations with one
292 another, ranging from 0.56 (SE = 0.10, $p < 0.001$; TUD and CanUD) to 0.89 (SE = 0.18, $p <$
293 0.001; OUD and CanUD). MDD and ANX were significantly correlated ($r_g = 0.89$, SE = 0.385, p
294 = 0.021), as were BD and SCZ ($r_g = 0.43$, SE = 0.17, $p = 0.011$). Across disorder classes, AUD
295 was correlated with all the psychiatric disorders except BD ($r_g = 0.19$, SE = 0.15, $p = 0.21$). On
296 the other hand, TUD was only significantly genetically correlated with BD ($r_g = 0.24$, SE = 0.12,
297 $p = 0.037$) and SCZ ($r_g = 0.29$, SE = 0.10, $p = 0.005$). CanUD was significantly genetically
298 correlated with all the psychiatric disorders except ANX ($r_g = 0.29$, SE = 0.20, $p = 0.145$), and
299 OUD was correlated with all (Supplementary Figure 3 and Supplementary Table 12).
300 **First-Order Common Factors: SUDs, Psychotic, and Mood Disorders**
301 ***European Ancestry.*** Of the EFA models examined, the 3-factor model fit the data best
302 (Supplementary Table 13). As all loadings were ≥ 0.35 with no significant cross-loadings, all
303 traits were carried forward on their respective factors for CFA, which fit the data well ($\chi^2(17) =$
304 130.04, $p = 1.84 \times 10^{-19}$, AIC = 168.04, CFI = 0.95, SRMR = 0.05). All SUDs loaded onto the first
305 factor, while BD and SCZ loaded onto the second, and MDD and ANX onto the third. The
306 factors were significantly intercorrelated, ranging from 0.39 (F1~F2) to 0.50 (F2~F3)
307 (Supplementary Table 14).

[Figure 2 here]

The SUD factor GWAS identified 143 lead SNPs (Supplementary Table 15), of which 47 were not genome-wide significant (GWS) or in LD with any GWS SNPs in the input GWAS, although they had all been significantly associated with SUDs or SUD-related traits in prior GWAS (Supplementary Table 15).[57] There were 17 independent Q-SNPs that demonstrated significant heterogeneity across SUDs (Supplementary Table 6). The psychotic disorder factor GWAS identified 9 independent Q-SNPs (Supplementary Table 7) and 162 lead SNPs (Supplementary Table 16), 27 of which had not been identified by the SCZ or BD GWAS, but all of which had previously been GWS or in LD with GWS SNPs in at least one psychotic trait GWAS (Supplementary Table 16). The mood factor GWAS identified 14 independent Q-SNPs (Supplementary Table 8) and 112 lead SNPs (Supplementary Table 17), 13 of which were not identified by the MDD or ANX GWAS, and 2 of which (rs75174029 and rs7652704) were not previously significantly associated with or in LD with GWS SNPs for any mood or anxiety disorder (Supplementary Table 17). PheWAS in the GWAS Atlas identified associations of rs75174029 with the number of non-cancer related illnesses, general risk tolerance, and having recent trouble relaxing; and rs7652704 with sensitivity/hurt feelings, neuroticism, and positive affect, among other related traits (Supplementary Figure 4).

rs75174029 is an intronic variant in *FOXP1*, which serves as a key regulatory gene in neural development.[58,59] Finemapping identified the SNP as the most likely causal variant accounting for the observed association within the locus (PICS probability = 0.307; Supplementary Table 18). Additionally, Hi-C chromatin interaction data (Supplementary Figure 5) revealed that rs75174029 contacts several regions of *FOXP1*, including its promoter. Taken together, this evidence suggests rs75174029's association with mood/anxiety disorders may be

due to its presence in an enhancer with cis-regulatory function on *FOXP1* during neural development. The second novel SNP, rs7652704 (Supplementary Figure 6), is an intronic variant in an uncharacterized non-coding RNA locus. rs7652704 and 20 other SNPs in strong LD all had PICS probabilities of 0.027, suggesting difficulty in determining the most likely causal variant accounting for the association in the locus (Supplementary Table 19). rs7652704 is an eQTL of *NECTIN3* (also known as *PVLR3)* in cultured fibroblasts and displays chromatin interaction with both *BTLA*, which is a gene involved in immune response,[60] and *NECTIN3*, which encodes a nectin adhesion molecule that regulates cell organization and modulates stress responses.[61-63] Thus, ample evidence associates both rs75174029 and rs7652704 with regulatory roles.

Gene-tissue expression analyses showed a role for both SUD- and psychotic disorder-related genes during prenatal brain development (Supplementary Figures 7 and 8), but no developmental period was significant for the mood disorders factor (Supplementary Figure 9). Mapped genes from each of the factor GWAS were significantly enriched for protein-protein interactions, indicating shared biological functions among disorders of the same class (SUDs = 1.37x, psychotic = 1.40x, and mood/anxiety disorders = 1.43x enrichment; *p*s < 1.00×10^{-16}).

[Figure 3 here]

African Ancestry. Fit was generally poor for each of the EFA models (all CFIs < 0.40; Supplementary Table 20). Upon examining each model's factor loadings, the proportion of variance explained by each factor, and eigenvalues, we tested two CFA models: a 2-factor model representing SUDs and psychiatric disorders and a 3-factor CFA model replicating the factor structure identified in EUR individuals. Attempting to fit the 3-factor CFA led to a negative residual variance and a correlation >1 between the psychotic and mood disorder factors. Because negative variances and correlations that are out of bounds can indicate issues with model

misspecification or overfitting the data, we proceeded with the 2-factor CFA model, which had an adequate fit ($\chi^2(19) = 21.49$, $p = 0.31$, AIC = 55.49, CFI = 0.99, SRMR = 0.10) and required no constraints (Figure 2 and Supplementary Table 21).

The SUD factor GWAS (Supplementary Figure 10) identified 1 lead SNP, rs1944683, positioned within an intergenic region on chromosome 11. To our knowledge, this variant has not previously been GWS or in LD with any GWS SNPs in AFR GWAS of substance use traits. However, the locus has been associated with alcohol consumption, tobacco-related traits, opioid use disorder, and cannabis use disorder in EUR and cross-ancestry studies.[64-66] The SNP exhibited chromatin interaction with two genes: *BLID* and *C11orf63*. A PheWAS of the lead SNP in the GWAS Atlas identified significant associations with regular smoking, past-month stomach pain, and tobacco-related conditions, such as atrial fibrillation and respiratory function (i.e., forced vital capacity and peak expiratory flow). One lead Q_{SNP}, rs10489130, identified on chromosome 4, exhibited significant associations with AUD only and was GWS in a previous AUD GWAS in AFR individuals, suggesting that this SNP may display specificity for AUD rather than influencing SUDs broadly.[66]

For the SUD factor, gene expression was significantly enriched in brain tissues involved in emotion processing, reward signaling, and cognitive control, including the putamen, amygdala, caudate, and hippocampus (Supplementary Figure 11). No significant variants were identified for the psychiatric disorders factor (Supplementary Figure 12). Gene expression was also not significantly enriched for any developmental stage or tissue type, although the top associations were with brain tissues (Supplementary Figure 13).

Second-Order Common Factors

European Ancestry. The SUD factor was significantly genetically correlated with both the psychotic ($r_g = 0.38$, SE = 0.03, $p < 0.001$) and mood ($r_g = 0.44$, SE = 0.03, $p < 0.001$) disorder factors. We used a higher-order CFA model ($\chi^2(2) = 57.61$, $p = 3.09 \times 10^{-13}$, AIC = 65.61, CFI = 0.91, SRMR = 0.07; Supplementary Figure 14 and Supplementary Table 22) to examine this shared genetic structure. After accounting for shared genetic risk, there was less standardized residual variance in the SUD factor ($u_{SUD} = 0.19$, SE = 0.04) than the psychotic ($u_{Psychotic} = 0.63$, SE = 0.05) or mood disorder factors ($u_{Mood} = 0.56$, SE = 0.04). A GWAS of the SUD and psychotic disorders factor identified 4 independent Q-SNPs (Supplementary Table 9) and 76 lead SNPs (Supplementary Table 23 and Supplementary Figure 15), 12 of which were not significant in any input or the first-order GWAS. For the SUD and mood factor, 6 independent Q-SNPs (Supplementary Table 10) and 62 lead SNPs were identified (Supplementary Table 24 and Supplementary Figure 16), 5 of which were in loci that were not significant in any input or first-order GWAS. All lead SNPs at the second-order level were previously associated with related traits.

Genetic risk shared among SUD and psychotic disorders implicated several gene sets, including molecular functions such as transcription regulation and sequence-specific DNA binding, and biological processes such as neuron differentiation. Genetic risk shared between SUDs and mood disorders was associated with enrichment in two gene sets involved in the biological processes of mechanosensory behavior and axonal protein transport. For both second-order factors, gene expression was enriched in brain tissues (Supplementary Figures 17 and 18). Genes associated with both second-order factors were significantly enriched for protein-protein interactions (SUDs and psychotic disorder = 1.54x, SUDs and mood/anxiety disorders = 1.57x; ps $< 1.00 \times 10^{-16}$).

399 ***African Ancestry.*** The SUD and psychiatric disorder factors were highly genetically
400 correlated (r_g = 0.74, SE = 0.13, $p < 0.001$), and a second-order CFA model accounting for this
401 shared genetic risk fit the data well ($\chi^2(19) = 21.49$, $p = 0.31$, AIC = 55.49, CFI = 0.99, SRMR =
402 0.10; Supplementary Table 25 and Supplementary Figure 19). Although no significant SNPs
403 were identified by the second-order GWAS (Supplementary Figure 20), there was enriched gene
404 expression in several brain regions, including those associated with reward processing (putamen,
405 caudate, nucleus accumbens), emotion processing and memory (amygdala, hippocampus), and
406 executive functions and decision-making (anterior cingulate cortex; Supplementary Figure 21).

407 **GWAS-by-Subtraction**

408 Common and independent genetic effects were examined for TUD, SCZ, and BD, which
409 each had a residual variance ≥0.30 in the EUR first-order common factor models. We chose this
410 threshold because analyses performed on traits with less residual variance were underpowered
411 for parsing transdiagnostic from disorder-level effects. We identified 102 GWS lead SNPs for
412 TUD Common, which represented genetic effects on TUD that operate through the SUD factor.
413 There were 20 lead SNPs for TUD Independent, which represented genetic effects on TUD that
414 are not shared with the other three SUDs. 13 of these SNPs did not reach significance in the TUD
415 Common GWAS (Supplementary Table 26). Chromatin interaction mapping identified several of
416 the nicotinic acetylcholine receptor genes (*CHRNA2, CHRNA4,* and *CHRNA5*; Supplementary
417 Figure 22). Gene-set analyses for TUD Independent implicated genes involved in nicotinic
418 acetylcholine reception and behavioral responses to nicotine.

419 We identified 51 lead SNPs for SCZ Common and 18 for SCZ Independent, of which 15
420 were not GWS nor in LD with GWS SNPs in the SCZ Common GWAS (Supplementary Table
421 27). Chromatin interaction mapping identified several genes on chromosome 6, including the

ZSCAN and *HIST1H* gene families, as possible sources of functional effects related to SCZ risk independent of BD (Supplementary Figure 23). Although after Bonferroni correction no gene-sets were significantly enriched for SCZ Independent, the top set involved up-regulated genes in the prefrontal cortex in mouse models of 22q11.2 microdeletions, which, in humans, are associated with risk of developing SCZ.[67] Finally, there were 189 significant lead SNPs for BD Common and 13 for BD Independent, 10 of which did not reach significance in the BD Common GWAS (Supplementary Table 28). Although few SNPs were significant in the BD Independent GWAS, chromatin interactions identified potential functional effects of these variants on genes, including *MXI1*[68] and *ADD3*,[69] which have been previously associated with BD (Supplementary Figure 24). For BD Independent, one gene-set involved in the regulation of trans-synaptic signaling remained significant after Bonferroni correction. Protein-protein interactions were significantly enriched for TUD, SCZ, and BIP Independent (TUD = 3.31x, $p = 3.45 \times 10^{-13}$, SCZ = 5.48x, $p < 1.00 \times 10^{-16}$, BIP = 5.75x, $p = 2.19 \times 10^{-11}$), indicating potential molecular mechanisms with enhanced specificity (Supplementary Figures 25-27).

[Figure 4 here]

Genetic Correlations with gSEM Factors

European Ancestry. The SUD factor was, as expected, strongly genetically correlated with smoking and alcohol traits, as well as depression and socioeconomic factors, including reduced educational attainment, unemployment due to sickness/disability, and the Townsend deprivation index (Supplementary Figure 28). The psychotic factor correlated most strongly with related traits, such as SCZ, BD, depression, and anxiety. However, psychotic disorders also exhibited a positive genetic correlation with risk taking and a negative correlation with cognitive performance, which consists of fluid intelligence and the first principal component of scores on

445 neuropsychological tests (Supplementary Figure 29). Although the mood disorders factor
446 correlated most strongly with measures of depression and anxiety, the remaining correlations
447 were predominantly with somatic conditions, such as chronic pain, longstanding
448 illness/disability/infirmity, and prescription medication usage (Supplementary Figure 30).

449 For the second-order factors, the top genetic correlations were a mixture of SUD-related,
450 psychiatric, and medical traits. The SUD and psychotic disorders factor, for example, correlated
451 most strongly with SCZ and BD. Other significant genetic correlations included smoking-related
452 traits, cognitive measures, and risk taking, as with the first-order psychotic disorder factor. The
453 SUD and mood disorders factor correlated most strongly with mood and anxiety traits and illness
454 and medication use for pain or gastrointestinal problems, reflecting similar associations observed
455 for the first-order mood disorders factor (Figure 5).

456 Genetic correlations underscored differences between disorder-level and transdiagnostic
457 genetic effects (Figure 4). Although TUD Common was significantly positively genetically
458 correlated with SCZ (r_g = 0.35, SE = 0.02, $p < 0.001$), TUD Independent was not (r_g = -0.05, SE
459 = 0.03, $p = 0.08$), with a similar pattern observed for other thought/psychotic disorders. SCZ
460 Common had a nominally weaker negative genetic correlation with cognitive performance (r_g = -
461 0.09, SE = 0.02, $p < 0.001$) and was significantly positively correlated with educational
462 attainment (r_g = 0.11, SE = 0.02, $p < 0.001$), while SCZ Independent had consistently negative
463 associations with both (r_g = -0.22, SE = 0.03, $p < 0.001$ and r_g = -0.09, SE = 0.02, $p < 0.001$,
464 respectively). Among other traits, BD Common and Independent showed opposite patterns of
465 associations with automobile speeding propensity and cognitive performance, both of which
466 were negative for BD Common (r_g = -0.21, SE = 0.02, $p < 0.001$; and r_g = -0.24, SE = 0.02, $p <$
467 0.001, respectively) and positive for BD Independent (r_g = 0.13, SE = 0.03, $p < 0.001$; and r_g =

0.05, SE = 0.02, $p = 0.03$, respectively). SCZ and BIP Independent had significantly different associations with risk-taking ($r_g = 0.25$, SE = 0.03, $p < 0.001$ vs. $r_g = -0.04$, SE = 0.03, $p = 0.14$; $Z = 6.84$, $p = 8.18 \times 10^{-12}$) and MDD ($r_g = 0.35$, SE = 0.03, $p < 0.001$ vs. $r_g = -0.04$, SE = 0.03, $p = 0.22$; $Z = 9.19$, $p = 3.84 \times 10^{-20}$).

African Ancestry. The psychiatric disorders factor was genetically correlated with all 11 traits examined, and the SUD factor was significantly correlated with all except PTSD. There were minimal differences between the genetic correlations for the first-order factors, likely due to the large variance in estimates resulting from low statistical power (Supplementary Figure 31). The second-order SUD and psychiatric factor correlated significantly with all traits except PTSD, and the strongest correlations were with smoking trajectory, OUD, depression, and maximum alcohol consumption (Supplementary Figure 32).

Trans-ancestry. The EUR SUD factor was significantly genetically correlated with the AFR SUD factor ($r_g = 0.730$, SE = 0.094, $p = 0.004$). Similarly, the AFR psychiatric disorders factor was genetically correlated with both the EUR psychotic factor ($r_g = 0.471$, SE = 0.216, $p = 0.014$) and the EUR mood disorders factor ($r_g = 0.571$, SE = 0.204, $p = 0.035$).

[Figure 5 here]

PheWAS in Penn Medicine BioBank

European Ancestry. Among participants in PMBB, the SUD factor was associated with TUD, tobacco-related illnesses (lung and other respiratory system cancers and chronic airway obstruction), and mood/anxiety disorders (Supplementary Figure 33 and Supplementary Table 29). The psychotic disorders factor was significantly associated only with psychiatric traits, including BD and mood/anxiety disorders (Supplementary Figure 34 and Supplementary Table 30). In contrast, although the mood disorders factor was most strongly associated with the

presence of various mood and anxiety disorders, it was also associated with physical health conditions like pain, sleep disorders, and obesity (Supplementary Figure 35 and Supplementary Table 31).

The second-order SUD and psychotic disorders factor was significantly associated with tobacco and alcohol-related disorders, with a nonsignificant association with BD (Supplementary Figure 36 and Supplementary Table 32). Although the PheWAS of the SUD and mood disorders factor revealed the strongest associations with substance use and psychiatric disorders, it also showed significant associations with pain, type 2 diabetes, ischemic heart disease, hypertension, and sleep disorders, amongst other physical health conditions (Supplementary Figure 37 and Supplementary Table 33).

PheWAS further highlighted the enhanced specificity of the GWAS-by-subtraction models. Although the TUD Common factor was associated with multiple SUDs, the TUD Independent factor demonstrated greater specificity, having the highest associations with TUD and related medical conditions, such as chronic airway obstruction (Supplementary Figure 38). PheWAS for SCZ Common factor showed broad associations with mood disorders, including BD, but there were no significant associations for SCZ Independent (Supplementary Figure 39). PheWAS of both the BD Common and Independent PRS showed significant associations with BD and other mood disorders, but the association with depression was only significant for BD Independent (Supplementary Figure 40).

African Ancestry. Although there were no statistically significant phenotypic associations among AFR individuals in PMBB, the top hits generally aligned with the factor being examined. For example, the SUD factor was most strongly associated with TUD, followed by SUDs broadly (Supplementary Figure 41 and Supplementary Table 34). Among the top associations for

514 the psychiatric factor were generalized anxiety disorder and the "other mental disorder"
515 phenotype (Supplementary Figure 42 and Supplementary Table 35). Similar results were seen for
516 the second-order SUD and psychiatric factor, which included TUD, alcohol-related disorders,
517 and mood disorders within the top associations (Supplementary Figure 43 and Supplementary
518 Table 36).

519 **Discussion**

520 Leveraging the largest available GWAS in European- and African-ancestry individuals,
521 we combined complementary multivariate methodologies to examine the shared genetic
522 architecture across SUDs, psychotic, mood, and anxiety disorders. We also examined genetic
523 effects that operate independent of the shared genetic risk to influence disorders. By integrating
524 transdiagnostic and disorder-level gSEM, we identified potential biological mechanisms that
525 contribute to comorbidity across disorder classes and those that distinguish commonly co-
526 occurring conditions. Our findings revealed both pervasive pleiotropy across SUDs and other
527 psychiatric disorders *and* trait-specific associations, while also highlighting the need to increase
528 representation of non-European ancestry individuals in genetic studies of mental health.

529 **Identification of SUDs, Psychotic, and Mood Disorder Factors**

530 Consistent with other psychiatric genetic findings,[19,70] we identified common genetic
531 factors underlying disorders that exhibit shared features. As expected, MDD and anxiety, which
532 are highly comorbid and share similar symptoms, loaded onto the same factor. Additionally,
533 across ancestries, common genetic risk partially accounted for the shared features of BD and
534 SCZ. Other gSEM studies have also showed that SCZ and BD load onto the same genetic
535 factor,[18,70-72] highlighting their strong shared etiology despite belonging to different diagnostic
536 classes. Although smaller samples and greater genetic diversity limited statistical power to

replicate the EUR factor structure in AFR individuals, there were significant commonalities across genetic ancestry groups. For example, in both AFR and EUR models, SUDs loaded onto a single factor that was highly genetically correlated with a previously identified genetic addiction factor.[19] Additionally, trans-ancestry genetic correlations highlighted the consistency of the genetic influences on the AFR psychiatric disorders factor with both the EUR psychotic and mood disorders factors.

In AFR individuals, we identified a lead SNP (rs1944683) associated with SUDs that had not been previously GWS or in LD with GWS SNPs in AFR substance-related GWAS. This locus has, however, been previously implicated in alcohol, tobacco, cannabis, and opioid-related traits in EUR and cross-ancestry GWAS.[64-66] Our ability to identify this lead SNP was facilitated by the use of gSEM, which allowed us to leverage power across multiple SUDs. In doing so, we were able to overcome some of the limitations posed by the relatively low statistical power of existing GWAS in AFR populations. However, despite statistical advancements, our study remained underpowered in AFR individuals.

In EUR individuals, we detected two significant loci for the mood and anxiety disorders factor that had not been previously significant in GWAS of mood or anxiety disorders. Performing chromatin interaction mapping on the lead SNPs for these loci implicated genes involved in immune and stress responses (*BTLA* and *NECTIN3*, respectively) and a gene related to hippocampal development (*FOXP1*).[58,60,62,73] The identification of *BTLA*, which encodes the B- and T-lymphocyte attenuator protein, is consistent with other research highlighting the role of immune dysregulation in psychiatric pathogenesis.[74-76] *NECTIN3* encodes a cell adhesion molecule that is involved in synaptic plasticity, enriched in hippocampal neurons, and implicated in stress-related disorders.[61,62,77] FOXP1, a transcription factor that plays a crucial role in

hippocampal development, has also been implicated in the regulation of synaptic plasticity.[58] Further highlighting the potential role of immune functioning, these loci have previously been implicated in GWAS of lymphocyte and leukocyte counts, as well as autoimmune conditions, such as inflammatory bowel disease. Thus, dysfunctions in synaptic plasticity and immune regulation may underlie a range of psychiatric conditions, including MDD and anxiety.[78,79] Targeting these common biological pathways could lead to the development of therapeutics with broader efficacy profiles.

In EUR individuals, both the SUD and psychotic disorder factors showed a significant association with genes expressed in brain tissue during prenatal development. Although no developmental stage reached significance for mood disorders or in AFR individuals, the top associations were consistently prenatal periods. This underscores the importance of early neurodevelopmental processes, including neurogenesis, synaptogenesis, and the formation of neurotransmitter systems, in shaping the susceptibility to SUDs and psychiatric disorders. Furthermore, prenatal development may represent a sensitive window during which genetic and environmental factors interact to influence long-term mental health outcomes.[80] Epigenetic mechanisms, such as those modulated by maternal stress and other environmental conditions,[81,82] may play a crucial role in mediating gene expression patterns and contributing to the developmental origins of psychiatric disorders. The specifics of these epigenetic mechanisms and other non-coding regulatory processes remain largely unclear. However, statistical and technological advances, such as next-generation sequencing and multi-omics analysis, show promise for furthering knowledge in this area.[83]

SUDs Share Genetic Liability with Psychotic and Mood Disorders

582 Evaluating shared genetic variance across the common factors revealed higher-order
583 dimensions of liability to psychopathology, including genetic risk shared between SUDs and
584 psychotic disorders and between SUDs and mood/anxiety disorders. A PheWAS in the PMBB
585 showed broad phenotypic manifestations of these dimensions of genetic liability. Notably, in
586 EUR individuals, SUD and mood/anxiety disorders exhibited significant associations with
587 various physical health conditions, including obesity, type 2 diabetes, chronic pain, chronic
588 airway obstruction, heart disease, hypertension, and sleep disorders. Because the PheWAS was
589 underpowered in AFR individuals, there were no significant associations, though the top
590 associations included relevant phenotypes, such as alcohol-related disorders, TUD, and mood
591 disorders.

592 Genetic correlations further underscored the pervasive impact of genetic risk for
593 psychiatric comorbidities and SUDs. The second-order factors (i.e., SUD and mood disorders
594 and SUD and psychotic disorders) were genetically correlated with adverse outcomes that
595 included lower cognitive performance, elevated HDL cholesterol, chronic pain, long-standing
596 illness, and miserableness. In AFR individuals, the second-order factor, representing shared
597 genetic effects across SUDs and psychiatric disorders, was significantly genetically correlated
598 with pain intensity and various substance use and psychiatric phenotypes. Thus, genetic risk for
599 co-occurring SUDs and psychiatric disorders has far-reaching implications for both mental and
600 physical health.

601 In addition to shared variance, in EUR individuals there was substantial unique genetic
602 variance for both psychotic (0.63, SE = 0.05) and mood/anxiety disorders (0.56, SE = 0.04). The
603 residual genetic variance in SUDs was also significant, albeit smaller, after accounting for
604 genetic variance shared with psychotic and mood/anxiety disorders (0.19, SE = 0.04). In EUR

individuals, more of the genetic variance in SUDs was shared with mood/anxiety than psychotic disorders. This suggests that there are different degrees of genetic convergence between SUDs and their common psychiatric comorbidities. Evaluating these patterns of genetic overlap and heterogeneity can help distinguish highly comorbid psychiatric disorders and pinpoint both shared and distinct etiological mechanisms.

Specificity of Genetic Effects for Tobacco Use Disorder, Schizophrenia, and Bipolar Disorder

To evaluate transdiagnostic and disorder-level liability, SNPs for TUD, SCZ, and BD were parsed into effects that: (1) operated through their respective common factors or (2) were uncorrelated with the common factor. Thus, we were able to distinguish biological mechanisms and genetic correlates that contribute to comorbidity across disorders from those that show greater specificity. This hierarchical approach can yield genetic knowledge across levels of psychopathology and identify patterns of convergence and divergence across disorders.

As an example of the utility of this approach, GWAS-by-subtraction findings provided insights into the underlying structure of psychotic disorders. SCZ and BD shared a common genetic core, consistent with their shared psychotic features and in line with empirical nosological models.[84] However, the genetic risk for each disorder that was independent of this common psychotic core showed different patterns of associations with other complex traits. For example, SCZ Independent was more negatively genetically correlated with measures of cognition and educational attainment than BD Independent, which was more strongly associated with risk-taking behaviors and affective disorders. Our results are highly consistent with the expanded psychosis continuum hypothesis, which proposes that although SCZ and BD share a psychotic core, cognitive and affective domains differentiate the disorders.[85] Specifically, SCZ is

628 characterized by greater cognitive impairments and BD by greater affective impairments. Our
629 findings lend genetic support to this hypothesis, previously investigated using only psychological
630 and neural evidence.[85]

631 PheWAS results further showcased the enhanced specificity of findings when
632 independent PGS for TUD, SCZ, and BD were compared with transdiagnostic liability. For
633 example, TUD Independent exhibited several associations not observed for the TUD Common
634 factor, including ischemic heart disease, atherosclerosis, obesity, and skin conditions. This may
635 reflect a combination of unhealthy lifestyle factors, direct physiological effects of tobacco use,
636 and shared biological pathways underlying these conditions. This finding is also consistent with
637 epidemiologic and genetic research supporting tobacco use as one of the strongest risk factors for
638 cardiovascular disease.[86-88] Differential patterns also emerged for SCZ and BD, with SCZ
639 Common and BD Independent showing broader associations with mood disorders than their
640 respective components. Recent research comparing transdiagnostic and disorder-specific genetic
641 effects across 11 psychiatric disorders similarly observed substantial differences in genetic
642 associations,[89] as has research parsing alcohol-specific risk from broader externalizing liability.[90]
643 Thus, hierarchical genetic approaches can facilitate a more nuanced understanding of patterns of
644 comorbidity and heterogeneity across co-occurring conditions.

645 These findings also pave the way for developing more refined PGS than are currently
646 available. For example, depression PGS show little specificity, accounting for a similar amount
647 of variance in mood disorders, anxiety disorders, ADHD, and SUDs.[91] Similarly, an evaluation of
648 16 PGS for psychiatric phenotypes found that most were associated with general
649 psychopathology rather than the specific domain for which they purported to measure risk.[92]
650 Although PGS have shown clinical utility in predicting some non-psychiatric disorders,[93,94]

prediction performance for psychiatric phenotypes remains limited.[95] This may be due in part to a lack of specificity of currently available psychiatric PGS. Existing PGS are useful for providing a broad overview of genetic predispositions to psychiatric disorders. However, when the focus shifts to exploring associations of a given psychiatric disorder without the confounding influence of co-occurring conditions, a more granular PGS is needed. Efforts to develop more precise PGS, including via the combined application of transdiagnostic and disorder-level genetic methods, may ultimately enhance their clinical utility.

Conclusions

Using a combination of multivariate approaches across multiple ancestral groups, we examined the common and independent genetic effects on SUDs, psychotic, mood, and anxiety disorders. In doing so, we identified potential biological mechanisms contributing to comorbidity both within and across classes of disorders, while also identifying pathways that may distinguish commonly co-occurring conditions. Isolating transdiagnostic genetic risk factors from those exhibiting greater specificity may aid in enhancing the precision of genetic prediction, as we observed when comparing genetic correlations and phenotypic associations of genetic risk factors. Thus, integrating transdiagnostic and disorder-level genetic models both clarifies the biological underpinnings of psychiatric comorbidity and identifies distinct biological pathways that contribute to heterogeneity within classes of psychiatric conditions.

In addition to advancing our understanding of the genetic architecture of SUDs and other psychiatric disorders, our study also highlights the importance of inclusivity in genetic research. Previous gSEM studies have been limited to European-ancestry individuals,[16,96,97] but as gSEM has shown the potential to uncover novel genetic associations and provide greater insights into the etiology of complex traits and diseases, it is imperative that these advances be made available

to individuals of all ancestral backgrounds. Unfortunately, in combination with smaller samples, lower LD and higher genetic diversity among African-ancestry individuals present statistical challenges for many existing analyses.[98] This is further complicated by the high degree of admixture present among individuals of non-European ancestries within the United States.[99] Accurate consideration of population substructure via the use of large, representative reference panels is essential for advancing genetic discovery. We provide details (Supplementary Materials) on the efforts we took to ensure the inclusion of African-ancestry individuals despite present limitations in the hopes that this may aid researchers faced with similar decisions. Advancing genetic discovery across diverse populations will be key in shaping our understanding of psychiatric etiology.

Acknowledgements

This study is supported by NIH grants K01AA028292 (R.L.K.), R01MH125938 (R.E.P.), P50AA022537 (R.E.P.), T32 HG009495 (K.L.F.), the Brain & Behavior Research Foundation NARSAD grant 28632 PS Fund (R.E.P.), the California Tobacco-Related Disease Research Program (TRDRP; T32IR5226) (S.S-R), the Department of VA Office of Research and Development grants IK2 CX002336 (E.E.H.), I01 BX003341 (H.R.K.) and the Veterans Integrated Service Network 4 Mental Illness Research, Education and Clinical Center. This publication does not represent the views of the Department of VA or the United States Government. We acknowledge the Penn Medicine BioBank (PMBB) for providing data and thank the patient-participants of Penn Medicine who consented to participate in this research program. We would also like to thank the Penn Medicine BioBank team and Regeneron Genetics Center for providing genetic variant data for analysis. The PMBB is approved under IRB protocol #813913 and supported by the Perelman School of Medicine at the University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878.

Table 1. Genome-wide association studies from which summary statistics were obtained.

European Ancestry					
Study	**Phenotype**	**Type**	**Cases**	**Controls**	**Total**
Als et al., 2023, *Nature Medicine*	MDD	Case-control	387,429	976,554	1,363,983
Otowa et al., 2016, *Molecular Psychiatry**	ANX	Case-control	7,016	14,745	21,761
Levey et al., 2020, *American Journal of Psychiatry**	ANX	Continuous			175,163
Purves et al., 2020, *Molecular Psychiatry**	ANX	Case-Control	25,453	58,113	83,566
Trubetskoy et al., 2022, *Nature*	SCZ	Case-control	53,386	77,258	130,644
Mullins et al., 2021, *Nature Genetics*	BD	Case-control	41,917	371,549	413,466
Zhou et al., 2023, *Nature Medicine*	AUD	Case-control	113,325	639,923	753,248
Levey et al., 2023 *Nature Genetics*	CanUD	Case-control	42,281	843,744	886,025
Toikumo et al., 2024, *Nature Human Behaviour*	TUD	Case-control	163,734	331,271	495,005
Kember et al., 2022, *Nature Neuroscience*	OUD	Case-control	31,473	394,471	425,944
African-Ancestry					
Levey et al., 2021, *Nature Neuroscience*	MDD	Case-control	25,843	33,757	59,600
Levey et al., 2021, *American Journal of Psychiatry*	ANX	Continuous			24,448
Bigdeli et al., 2021, *Schizophrenia Bulletin*	SCZ	Case-control	7,509	8,612	16,121
Bigdeli et al., 2021, *Schizophrenia Bulletin*	BD	Case-control	3,027	8,097	11,124
Kember et al., 2023, *American Journal of Psychiatry*	AUD	Case-control	25,012	52,313	77,325
Levey et al., 2023, *Nature Genetics*	CanUD	Case-control	19,065	104,143	123,208
Toikumo et al., 2024, *Nature Human Behaviour*	TUD	Case-control	45,465	68,955	114,420
Kember et al., 2022, *Nature Neuroscience*	OUD	Case-control	8,968	79,530	88,498

Note: MDD = Major Depressive Disorder, ANX = Anxiety Disorders, SCZ = Schizophrenia, BD = Bipolar Disorder, AUD = Alcohol Use Disorder, CanUD = Cannabis Use Disorder, TUD = Tobacco Use Disorder, OUD = Opioid Use Disorder. See Supplementary Tables 1-4 for full description of cohorts.

*Summary statistics were jointly analyzed using MTAG to enhance the statistical power of the ANX phenotype prior to their being used in gSEM (see Supplementary Materials).

Figure Legends

Figure 1. Study schema. FUMA = Functional Mapping and Annotation of GWAS, CTG-VL = Complex Trait Genetics Virtual Lab, LDSC = linkage disequilibrium score regression, PheWAS = phenome-wide association study.

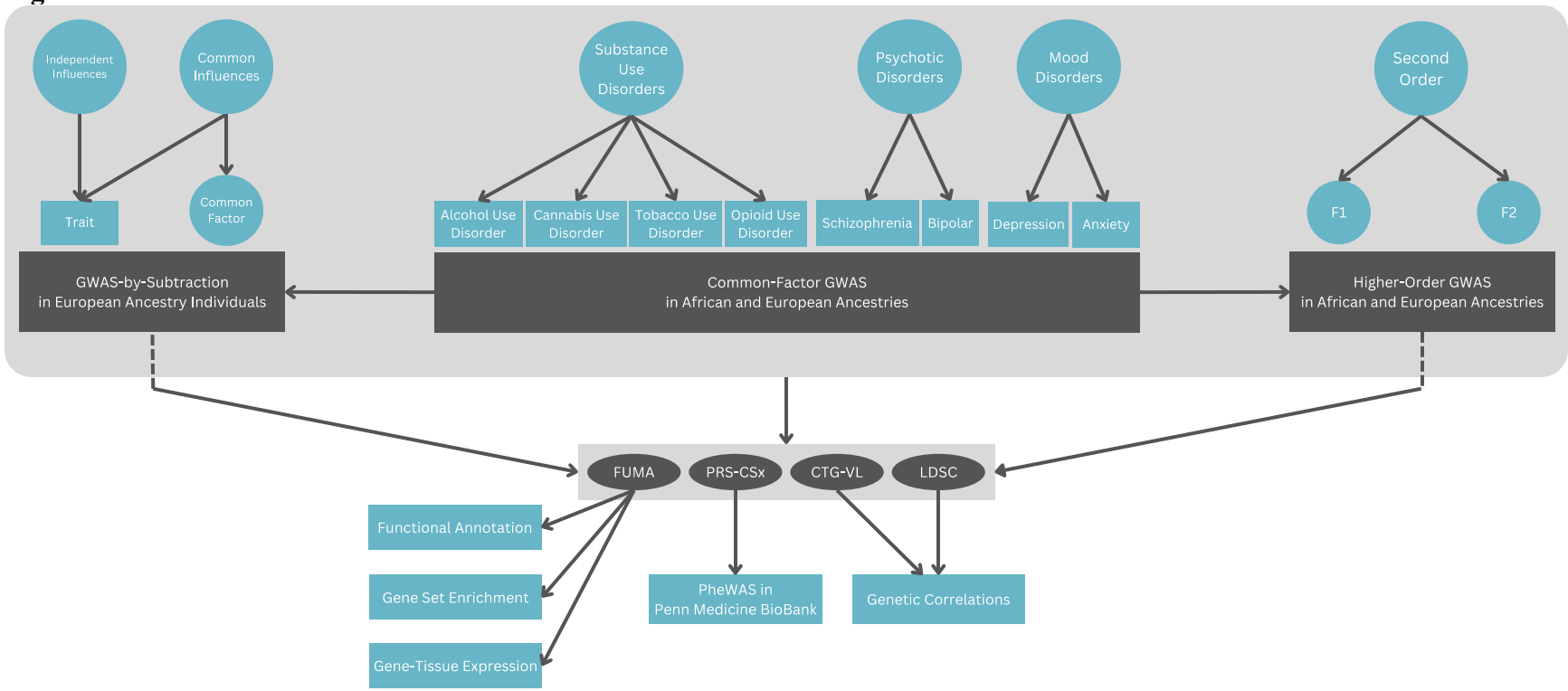
Figure 2. Genomic structural equation models. AUD = alcohol use disorder, CanUD = cannabis use disorder, TUD = tobacco use disorder, OUD = opioid use disorder, BD = bipolar disorder, SCZ = schizophrenia, MDD = major depressive disorder, ANX = anxiety, GAD-2 = 2-item GAD questionnaire.

Figure 3. Manhattan plots for common factors. The substance use disorders factor GWAS identified 143 lead SNPs, the psychotic disorders factor identified 162 lead SNPs, and the mood/anxiety disorders factor identified 112 lead SNPs. The lead SNPs for loci that were not significant in the input GWAS are annotated with yellow diamonds, and lead SNPs for loci not previously significantly associated with phenotypes related to the common factor (i.e., novel) are annotated with green diamonds.

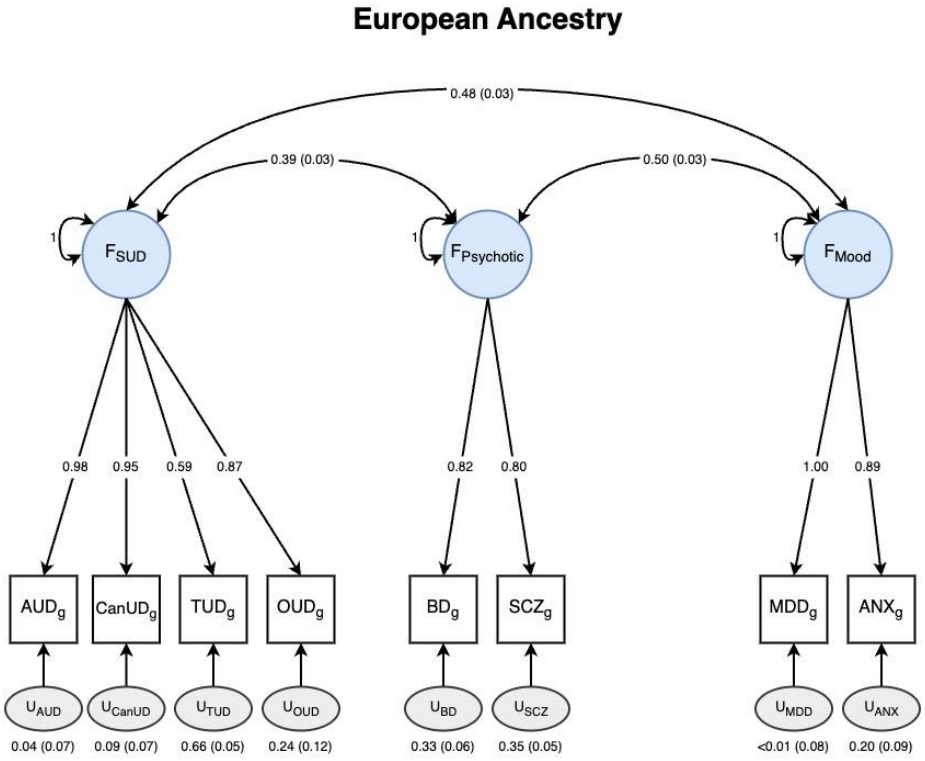
Figure 4. Hudson plots and genetic correlations of GWAS-by-subtraction models. The left panel presents Hudson plots of the GWAS-by-subtraction model results with the mapped gene for lead SNPs annotated. The right panel presents the genetic correlation results. Independent GWAS refers to the influences on a disorder that do not operate through the common factor, while the Common GWAS refers to influences on the disorder that do operate through the common factor.

Figure 5. Genetic correlations for second-order common factors using Complex Trait Genetics Virtual Lab. The dashed line represents the log-transformed Bonferroni-corrected p-value across the 1,437 traits included in the analysis.

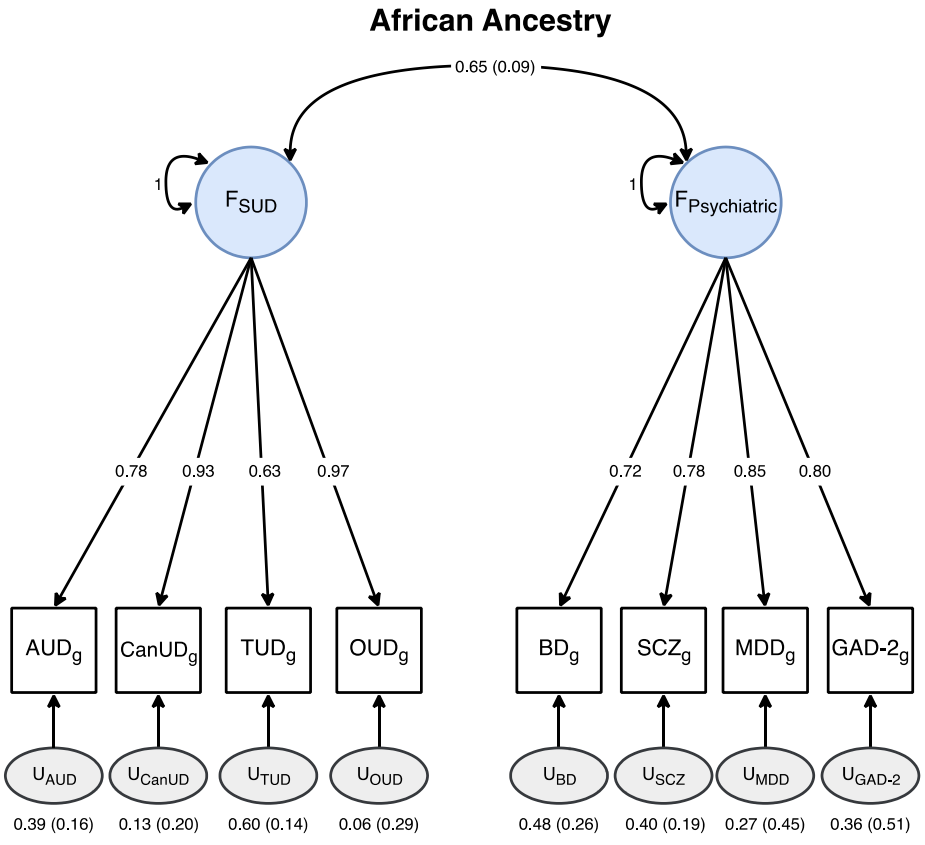
Figure 1.


Independent Influences
Common Influences
Trait
Common Factor
GWAS-by-Subtraction in European Ancestry Individuals
Substance Use Disorders
Alcohol Use Disorder
Cannabis Use Disorder
Tobacco Use Disorder
Opioid Use Disorder
Psychotic Disorders
Schizophrenia
Bipolar
Mood Disorders
Depression
Anxiety
Common-Factor GWAS in African and European Ancestries
Second Order
F1
F2
Higher-Order GWAS in African and European Ancestries
FUMA
PRS-CSx
CTG-VL
LDSC
Functional Annotation
Gene Set Enrichment
Gene-Tissue Expression
PheWAS in Penn Medicine BioBank
Genetic Correlations

736 **Figure 2.**

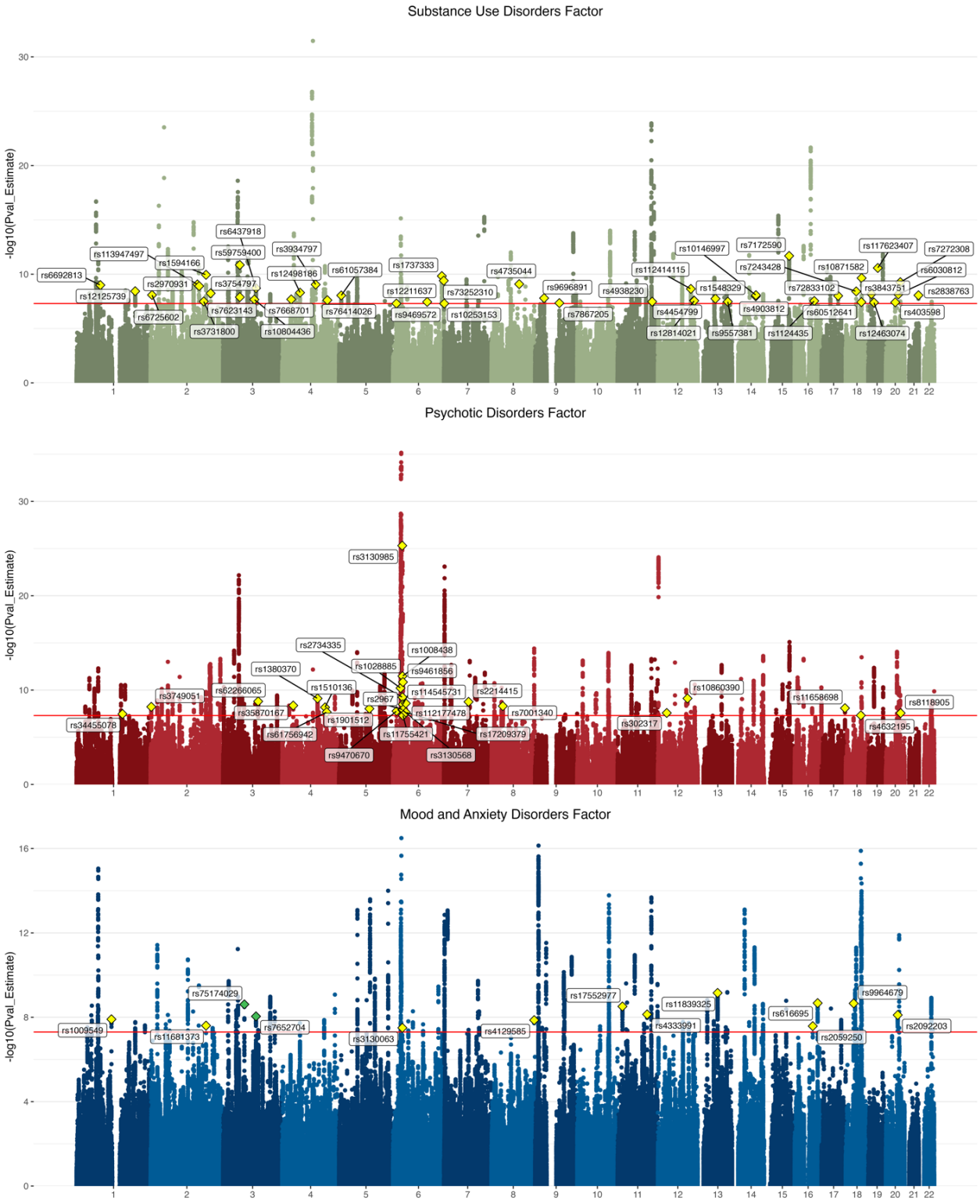


737



738

739 **Figure 3.**



740
741

Figure 4.

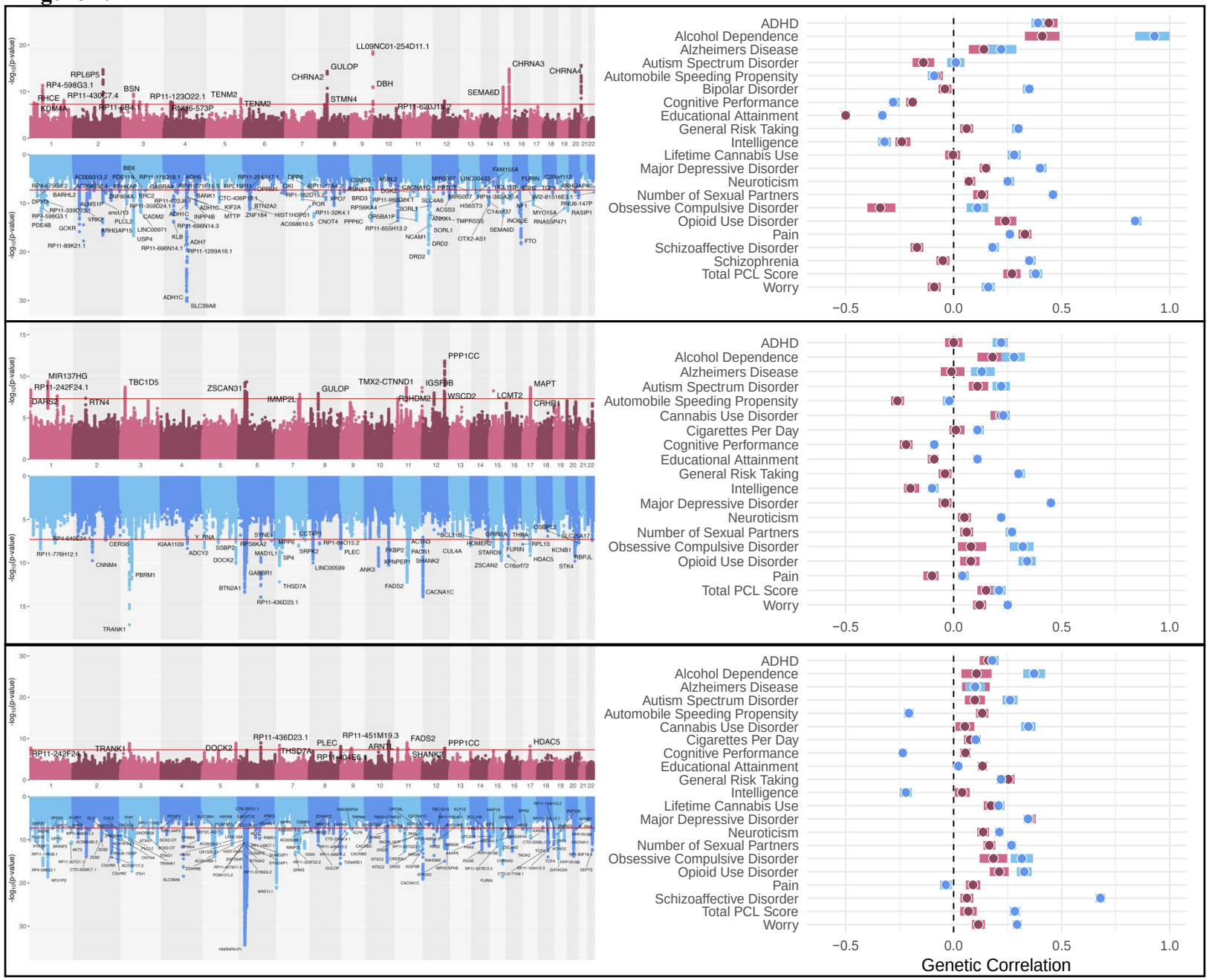
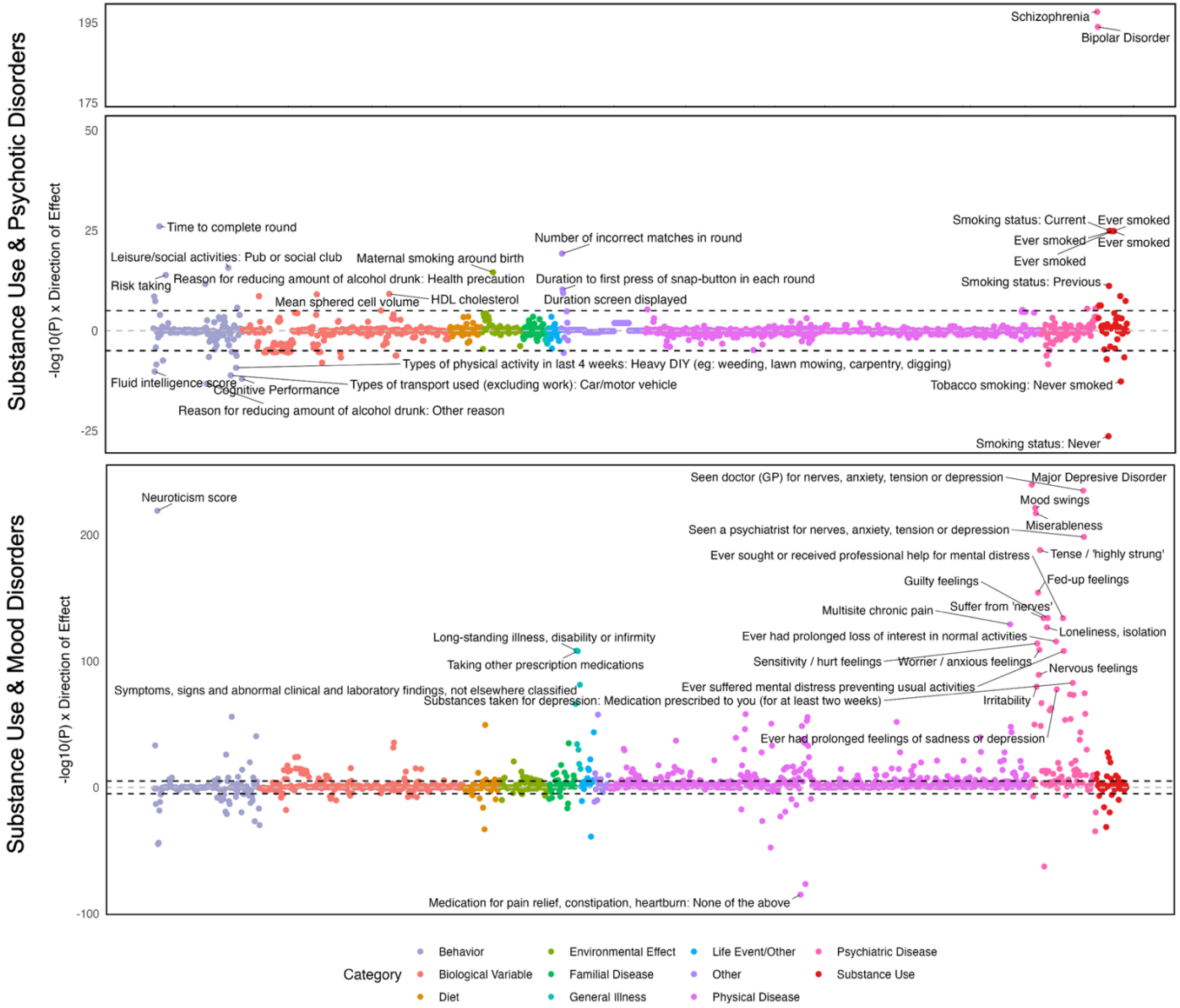

Tobacco Use Disorder
Schizophrenia
Bipolar Disorder
-log10(p-value)
ADHD
Alcohol Dependence
Alzheimers Disease
Autism Spectrum Disorder
Automobile Speeding Propensity
Bipolar Disorder
Cannabis Use Disorder
Cigarettes Per Day
Cognitive Performance
Educational Attainment
General Risk Taking
Intelligence
Lifetime Cannabis Use
Major Depressive Disorder
Neuroticism
Number of Sexual Partners
Obsessive Compulsive Disorder
Opioid Use Disorder
Pain
Schizoaffective Disorder
Schizophrenia
Total PCL Score
Worry
−0.5
0.0
0.5
1.0
Genetic Correlation
GWAS
Independent
Common

Figure 5.


Substance Use & Psychotic Disorders
-log10(P) x Direction of Effect
195
175
50
25
0
-25
Schizophrenia
Bipolar Disorder
Time to complete round
Leisure/social activities: Pub or social club
Risk taking
Reason for reducing amount of alcohol drunk: Health precaution
Maternal smoking around birth
Number of incorrect matches in round
Duration to first press of snap-button in each round
Mean sphered cell volume
HDL cholesterol
Duration screen displayed
Smoking status: Current
Ever smoked
Ever smoked
Ever smoked
Ever smoked
Smoking status: Previous
Types of physical activity in last 4 weeks: Heavy DIY (eg: weeding, lawn mowing, carpentry, digging)
Fluid intelligence score
Cognitive Performance
Types of transport used (excluding work): Car/motor vehicle
Reason for reducing amount of alcohol drunk: Other reason
Tobacco smoking: Never smoked
Smoking status: Never
Substance Use & Mood Disorders
-log10(P) x Direction of Effect
200
100
0
-100
Neuroticism score
Seen doctor (GP) for nerves, anxiety, tension or depression
Major Depresive Disorder
Mood swings
Seen a psychiatrist for nerves, anxiety, tension or depression
Miserableness
Ever sought or received professional help for mental distress
Tense / 'highly strung'
Guilty feelings
Fed-up feelings
Multisite chronic pain
Suffer from 'nerves'
Loneliness, isolation
Long-standing illness, disability or infirmity
Ever had prolonged loss of interest in normal activities
Taking other prescription medications
Sensitivity / hurt feelings
Worrier / anxious feelings
Nervous feelings
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
Ever suffered mental distress preventing usual activities
Substances taken for depression: Medication prescribed to you (for at least two weeks)
Irritability
Ever had prolonged feelings of sadness or depression
Medication for pain relief, constipation, heartburn: None of the above
Category
Behavior
Biological Variable
Diet
Environmental Effect
Familial Disease
General Illness
Life Event/Other
Other
Physical Disease
Psychiatric Disease
Substance Use

References

1. Hunt, G.E., Malhi, G.S., Lai, H.M.X. & Cleary, M. Prevalence of comorbid substance use in major depressive disorder in community and clinical settings, 1990–2019: Systematic review and meta-analysis. *Journal of Affective Disorders* **266**, 288-304 (2020).

2. Lai, H.M.X., Cleary, M., Sitharthan, T. & Hunt, G.E. Prevalence of comorbid substance use, anxiety and mood disorders in epidemiological surveys, 1990–2014: A systematic review and meta-analysis. *Drug and Alcohol Dependence* **154**, 1-13 (2015).

3. Hunt, G.E., Malhi, G.S., Cleary, M., Lai, H.M.X. & Sitharthan, T. Comorbidity of bipolar and substance use disorders in national surveys of general populations, 1990–2015: Systematic review and meta-analysis. *Journal of Affective Disorders* **206**, 321-330 (2016).

4. European Monitoring Centre for Drugs and Drug Addiction, Domingo-Salvany, A., Torrens, M. & Mestre-Pintó, J. *Comorbidity of substance use and mental disorders in Europe*, (Publications Office, 2015).

5. Hunt, G.E., Large, M.M., Cleary, M., Lai, H.M.X. & Saunders, J.B. Prevalence of comorbid substance use in schizophrenia spectrum disorders in community and clinical settings, 1990–2017: Systematic review and meta-analysis. *Drug and Alcohol Dependence* **191**, 234-258 (2018).

6. Horsfall, J., Cleary, M., Hunt, G.E. & Walter, G. Psychosocial treatments for people with co-occurring severe mental illnesses and substance use disorders (dual diagnosis): A review of empirical evidence. *Harvard Review of Psychiatry* **17**, 24-34 (2009).

7. McGovern, M.P., Xie, H., Segal, S.R., Siembab, L. & Drake, R.E. Addiction treatment services and co-occurring disorders: Prevalence estimates, treatment practices, and barriers. *Journal of Substance Abuse Treatment* **31**, 267-275 (2006).

8. Levey, D.F. *et al.* Reproducible genetic risk loci for anxiety: Results from ~200,000 participants in the Million Veteran Program. *American Journal of Psychiatry* **177**, 223-232 (2020).

9. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022).

10. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics* **53**, 817-829 (2021).

11. Kember, R.L. *et al.* Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. *Nature Neuroscience* **25**, 1279-1287 (2022).

12. Levey, D.F. *et al.* Multi-ancestry genome-wide association study of cannabis use disorder yields insight into disease biology and public health implications. *Nature Genetics* (2023).

13. Zhou, H. *et al.* Multi-ancestry study of the genetics of problematic alcohol use in over 1 million individuals. *Nature Medicine* (2023).

14. Als, T.D. *et al.* Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nature Medicine* **29**, 1832-1844 (2023).

15. Toikumo, S. *et al.* Multi-ancestry meta-analysis of tobacco use disorder identifies 461 potential risk genes and reveals associations with multiple health outcomes. *Nature Human Behaviour* (2024).

16. Grotzinger, A.D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour* **3**, 513-525 (2019).

17. Lee, P.H. *et al.* Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469-1482.e11 (2019).

18. Grotzinger, A.D. *et al.* Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics* **54**, 548-559 (2022).

19. Hatoum, A.S. *et al.* Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nature Mental Health* **1**, 210-223 (2023).

20. Bhattacharjee, S. *et al.* A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *American Journal of Human Genetics* **90**, 821-35 (2012).

21. Schoeler, T. *et al.* Novel biological insights into the common heritable liability to substance involvement: A multivariate genome-wide association study. *Biological Psychiatry* **93**, 524-535 (2023).

22. Demange, P.A. *et al.* Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nature Genetics* **53**, 35-44 (2021).

812 23. Waszczuk, M.A. *et al.* Dimensional and transdiagnostic phenotypes in psychiatric
813 genome-wide association studies. *Molecular Psychiatry* (2023).
814 24. Kember, R.L. *et al.* Genetic underpinnings of the transition from alcohol consumption to
815 alcohol use disorder: Shared and unique genetic architectures in a cross-ancestry sample.
816 *American Journal of Psychiatry* **180**, 584-593 (2023).
817 25. Bigdeli, T.B. *et al.* Genome-wide association studies of schizophrenia and bipolar
818 disorder in a diverse cohort of US veterans. *Schizophrenia Bulletin* **47**, 517-529 (2020).
819 26. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders.
820 *Molecular Psychiatry* **21**, 1391-1399 (2016).
821 27. Purves, K.L. *et al.* A major role for common genetic variation in anxiety disorders.
822 *Molecular Psychiatry* **25**, 3292-3303 (2020).
823 28. Levey, D.F. *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and
824 meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nature*
825 *Neuroscience* **24**, 954-963 (2021).
826 29. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using
827 MTAG. *Nature Genetics* **50**, 229-237 (2018).
828 30. Grotzinger, A.D., Fuente, J.d.l., Privé, F., Nivard, M.G. & Tucker-Drob, E.M. Pervasive
829 downward bias in estimates of liability-scale heritability in genome-wide association
830 study meta-analysis: A simple solution. *Biological Psychiatry* **93**, 29-36 (2023).
831 31. The International HapMap 3 Consortium. Integrating common and rare genetic variation
832 in diverse human populations. *Nature* **467**, 52-8 (2010).
833 32. The 1000 Genomes Project Consortium. A global reference for human genetic variation.
834 *Nature* **526**, 68-74 (2015).

33. Pan-UKB Team. https://pan.ukbb.broadinstitute.org. (2020).

34. Bates, D. & Maechler, M. Matrix: Sparse and dense matrix classes and methods. *R package*, http://cran.r-project.org/package=Matrix (2010).

35. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* **48**, 1 - 36 (2012).

36. West, S.G., Taylor, A.B. & Wu, W. Model fit and model selection in structural equation modeling. in *Handbook of structural equation modeling.* 209-231 (The Guilford Press, New York, NY, US, 2012).

37. Mallard, T.T. *et al.* Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Cell Genomics* **2**, 100140 (2022).

38. Loehlin, J.C. The Cholesky approach: A cautionary note. *Behavior Genetics* **26**, 65-69 (1996).

39. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575 (2007).

40. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339-1348 (2019).

41. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).

42. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Computational Biology* **11**, e1004219 (2015).

43. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).

858 44. The GTEx Consortium *et al.* The GTEx Consortium atlas of genetic regulatory effects
859 across human tissues. *Science* **369**, 1318-1330 (2020).
860 45. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for
861 the human brain. *Science* **362**(2018).
862 46. Schmitt, Anthony D. *et al.* A compendium of chromatin contact maps reveals spatially
863 active regions in the human genome. *Cell Reports* **17**, 2042-2059 (2016).
864 47. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks
865 and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids*
866 *Research* **51**, D638-d646 (2023).
867 48. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from
868 polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295 (2015).
869 49. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.
870 *Nature Genetics* **47**, 1236-1241 (2015).
871 50. Cuéllar-Partida, G. *et al.* Complex-Traits Genetics Virtual Lab: A community-driven web
872 platform for post-GWAS analyses. *bioRxiv*, 518027 (2019).
873 51. Brown, Brielin C., Ye, Chun J., Price, Alkes L. & Zaitlen, N. Transethnic genetic-
874 correlation estimates from summary statistics. *The American Journal of Human Genetics*
875 **99**, 76-88 (2016).
876 52. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nature*
877 *Genetics* **54**, 573-580 (2022).
878 53. Verma, A. *et al.* The Penn Medicine BioBank: Towards a genomics-enabled learning
879 healthcare system to accelerate precision medicine in a diverse population. *Journal of*
880 *Personalized Medicine* **12**, 1974 (2022).

881 54. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide
882 association studies. *Nature Genetics* **38**, 904-909 (2006).
883 55. Galinsky, Kevin J. *et al.* Fast principal-component analysis reveals convergent evolution
884 of ADH1B in Europe and East Asia. *The American Journal of Human Genetics* **98**, 456-
885 472 (2016).
886 56. Denny, J.C. *et al.* Systematic comparison of phenome-wide association study of
887 electronic medical record data and genome-wide association study data. *Nature*
888 *Biotechnology* **31**, 1102-1111 (2013).
889 57. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition
890 resource. *Nucleic Acids Research* **51**, D977-D985 (2023).
891 58. Araujo, D.J. *et al.* FOXP1 in forebrain pyramidal neurons controls gene expression
892 required for spatial learning and synaptic plasticity. *The Journal of Neuroscience* **37**,
893 10917-10931 (2017).
894 59. Braccioli, L. *et al.* FOXP1 promotes embryonic neural stem cell differentiation by
895 repressing jagged1 expression. *Stem Cell Reports* **9**, 1530-1545 (2017).
896 60. Ning, Z., Liu, K. & Xiong, H. Roles of BTLA in immunity and immune disorders.
897 *Frontiers in Immunology* **12**(2021).
898 61. Wang, X.D. *et al.* Nectin-3 links CRHR1 signaling to stress-induced memory deficits and
899 spine loss. *Nature Neuroscience* **16**, 706-13 (2013).
900 62. Wang, H.L. *et al.* Prefrontal nectin3 reduction mediates adolescent stress-induced deficits
901 of social memory, spatial working memory, and dendritic structure in mice. *Neuroscience*
902 *Bulletin* **36**, 860-874 (2020).

903 63. van der Kooij, M.A. *et al.* Role for MMP-9 in stress-induced downregulation of nectin-3
904 in hippocampal CA1 and associated behavioural alterations. *Nature Communications* **5**,
905 4995 (2014).
906 64. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into
907 the genetic etiology of tobacco and alcohol use. *Nature Genetics* **51**, 237-244 (2019).
908 65. Zhou, H. *et al.* Genome-wide association study identifies glutamate ionotropic receptor
909 GRIA4 as a risk gene for comorbid nicotine dependence and major depression.
910 *Translational Psychiatry* **8**, 208 (2018).
911 66. Xu, H. *et al.* Identifying genetic loci and phenomic associations of substance use traits: A
912 multi-trait analysis of GWAS (MTAG) study. *Addiction* **118**, 1942-1952 (2023).
913 67. Stark, K.L. *et al.* Altered brain microRNA biogenesis contributes to phenotypic deficits in
914 a 22q11-deletion mouse model. *Nature Genetics* **40**, 751-60 (2008).
915 68. Wu, Y. *et al.* Multi-trait analysis for genome-wide association study of five psychiatric
916 disorders. *Translational Psychiatry* **10**, 209 (2020).
917 69. Charney, A.W. *et al.* Evidence for genetic heterogeneity between clinical subtypes of
918 bipolar disorder. *Translational Psychiatry* **7**, e993-e993 (2017).
919 70. Romero, C. *et al.* Exploring the genetic overlap between twelve psychiatric disorders.
920 *Nature Genetics* **54**, 1795-1802 (2022).
921 71. Martin, E., Schoeler, T., Pingault, J.-B. & Barkhuizen, W. Understanding the relationship
922 between loneliness, substance use traits and psychiatric disorders: A genetically informed
923 approach. *Psychiatry Research* **325**, 115218 (2023).

924 72. Paul, S.E. *et al.* Phenome-wide investigation of behavioral, environmental, and neural
925 associations with cross-disorder genetic liability in youth of European ancestry. *medRxiv*,
926 2023.02.10.23285783 (2023).
927 73. Trelles, M.P. *et al.* Individuals with FOXP1 syndrome present with a complex
928 neurobehavioral profile with high rates of ADHD, anxiety, repetitive behaviors, and
929 sensory symptoms. *Molecular Autism* **12**, 61 (2021).
930 74. Drevets, W.C., Wittenberg, G.M., Bullmore, E.T. & Manji, H.K. Immune targets for
931 therapeutic development in depression: Towards precision medicine. *Nature Reviews*
932 *Drug Discovery* **21**, 224-244 (2022).
933 75. Tubbs, J.D., Ding, J., Baum, L. & Sham, P.C. Immune dysregulation in depression:
934 Evidence from genome-wide association. *Brain, Behavior, & Immunity - Health* **7**,
935 100108 (2020).
936 76. Katrinli, S., Oliveira, N.C.S., Felger, J.C., Michopoulos, V. & Smith, A.K. The role of the
937 immune system in posttraumatic stress disorder. *Translational Psychiatry* **12**, 313 (2022).
938 77. Wang, X.X. *et al.* Nectin-3 modulates the structural plasticity of dentate granule cells and
939 long-term memory. *Translational Psychiatry* **7**, e1228-e1228 (2017).
940 78. Parekh, P.K., Johnson, S.B. & Liston, C. Synaptic mechanisms regulating mood state
941 transitions in depression. *Annual Review of Neuroscience* **45**, 581-601 (2022).
942 79. Appelbaum, L.G., Shenasa, M.A., Stolz, L. & Daskalakis, Z. Synaptic plasticity and
943 mental health: Methods, challenges and opportunities. *Neuropsychopharmacology* **48**,
944 113-120 (2023).

945 80. Marco, E.M., Macrì, S. & Laviola, G. Critical age windows for neurodevelopmental
946 psychiatric disorders: Evidence from animal models. *Neurotoxicity Research* **19**, 286-307
947 (2011).
948 81. Lund, R.J. *et al.* Placental DNA methylation marks are associated with maternal
949 depressive symptoms during early pregnancy. *Neurobiology of Stress* **15**, 100374 (2021).
950 82. McGill, M.G. *et al.* Maternal prenatal anxiety and the fetal origins of epigenetic aging.
951 *Biological Psychiatry* **91**, 303-312 (2022).
952 83. Jourdon, A., Scuderi, S., Capauto, D., Abyzov, A. & Vaccarino, F.M. PsychENCODE and
953 beyond: Transcriptomics and epigenomics of brain development and organoids.
954 *Neuropsychopharmacology* **46**, 70-85 (2021).
955 84. Kotov, R. *et al.* Validity and utility of Hierarchical Taxonomy of Psychopathology
956 (HiTOP): I. Psychosis superspectrum. *World Psychiatry* **19**, 151-172 (2020).
957 85. Sorella, S. *et al.* Testing the expanded continuum hypothesis of schizophrenia and bipolar
958 disorder. Neural and psychological evidence for shared and distinct mechanisms.
959 *NeuroImage: Clinical* **23**, 101854 (2019).
960 86. Gallucci, G., Tartarone, A., Lerose, R., Lalinga, A.V. & Capobianco, A.M. Cardiovascular
961 risk of smoking and benefits of smoking cessation. *Journal of Thoracic Disease* **12**,
962 3866-3876 (2020).
963 87. Rosoff, D.B., Davey Smith, G., Mehta, N., Clarke, T.-K. & Lohoff, F.W. Evaluating the
964 relationship between alcohol consumption, tobacco use, and cardiovascular disease: A
965 multivariable Mendelian randomization study. *PLOS Medicine* **17**, e1003410 (2020).
966 88. Rostron, B.L. *et al.* Smokeless tobacco use and circulatory disease risk: a systematic
967 review and meta-analysis. *Open Heart* **5**, e000846 (2018).

89. Keser, E. *et al.* Isolating transdiagnostic effects reveals specific genetic profiles in psychiatric disorders. *medRxiv*, 2023.12.20.23300292 (2023).

90. Barr, P.B. *et al.* Parsing genetically influenced risk pathways: genetic loci impact problematic alcohol use via externalizing and specific risk. *Translational Psychiatry* **12**, 420 (2022).

91. Shi, Y. *et al.* Multi-polygenic scores in psychiatry: From disorder specific to transdiagnostic perspectives. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **195**, e32951 (2024).

92. Neumann, A. *et al.* Combined polygenic risk scores of different psychiatric traits predict general and specific psychopathology in childhood. *Journal of Child Psychology and Psychiatry* **63**, 636-645 (2022).

93. Khera, A.V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219-1224 (2018).

94. Inouye, M. *et al.* Genomic risk prediction of coronary artery disease in 480,000 adults. *Journal of the American College of Cardiology* **72**, 1883-1893 (2018).

95. Lewis, C.M. & Vassos, E. Polygenic scores in psychiatry: On the road From discovery to implementation. *American Journal of Psychiatry* **179**, 800-806 (2022).

96. Grotzinger, A.D. *et al.* Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature Genetics* **54**, 548-559 (2022).

97. Grotzinger, A.D. *et al.* Multivariate genomic architecture of cortical thickness and surface area at multiple levels of analysis. *Nature Communications* **14**, 946 (2023).

991 98. Bentley, A.R., Callier, S.L. & Rotimi, C.N. Evaluating the promise of inclusion of
992 African ancestry populations in genomics. *npj Genomic Medicine* **5**, 5 (2020).
993 99. Bick, A.G. *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340-346
994 (2024).
995