# Genetic Insights into Externalizing and Internalizing Traits through Integration of the Research Domain Criteria and Hierarchical Taxonomy of Psychopathology

Christal N. Davis[1,2], Yousef Khan[2], Sylvanus Toikumo[1,2], Zeal Jinwala[2], Dorret I. Boomsma[3], Daniel F. Levey[4,5], Joel Gelernter[5,6], Rachel L. Kember[1,2], & Henry R. Kranzler[1,2]

[1] Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia, PA, USA

[2] Center for Studies of Addiction, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

[3] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, The Netherlands and Amsterdam Reproduction and Development Research Institute, Amsterdam, The Netherlands

[4] Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

[5] VA Connecticut Healthcare Center, West Haven, CT, USA

[6] Departments of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine, New Haven, CT, USA

**Corresponding Author:** Christal N. Davis, 3535 Market Street Suite 500, Philadelphia, PA 19104, (215) 746-0465, christal.davis@pennmedicine.upenn.edu

**Abstract**

**Background:** There is considerable comorbidity between externalizing (EXT) and internalizing (INT) psychopathology. Understanding the shared genetic underpinnings of these spectra is crucial for advancing knowledge of their biological bases and potential health impacts, and for informing empirical models like the Research Domain Criteria (RDoC) and Hierarchical Taxonomy of Psychopathology (HiTOP).

**Methods:** We conducted a multivariate genome-wide association study (GWAS) of EXT and INT psychopathology by applying genomic structural equation modeling to summary statistics from 16 EXT and INT traits in European-ancestry individuals (n = 16,400 to 1,074,629). Downstream analyses explored associations across RDoC units of analysis (i.e., genes, molecules, cells, circuits, physiology, and behaviors).

**Results:** The GWAS identified 409 lead single nucleotide polymorphisms (SNPs) for EXT, 85 for INT, and 256 for EXT+INT (i.e., shared) traits. Bivariate causal mixture models estimated that nearly all EXT and INT causal variants overlapped, despite a genetic correlation of 0.37 (SE = 0.02). Drug repurposing analyses identified potential therapeutic targets, including perturbagens affecting dopamine and serotonin pathways. EXT genes had enriched expression in GABAergic, cortical, and hippocampal neurons, while INT genes were more narrowly linked to GABAergic neurons. EXT+INT liability was associated with reduced grey matter volumes in the amygdala and subcallosal cortex.

**Conclusions:** These findings reveal both genetic overlap and distinct molecular and neurobiological pathways underlying EXT and INT psychopathology. By integrating genomic insights with the RDoC and HiTOP frameworks, this study advances our understanding of the mechanisms driving these dimensions of psychopathology.

## Introduction

Traditional categorical classifications of psychopathology suffer from significant limitations. Psychiatric disorders frequently co-occur more often than would be expected by chance alone [1, 2], suggesting that shared genetic underpinnings transcend diagnostic categories [3]. Additionally, current diagnostic classifications rely on arbitrary thresholds and polythetic criterion sets, resulting in thousands of unique symptom combinations that yield the same diagnosis [4]. These limitations complicate clinical care and constrain progress in psychiatric genetics and neuroscience research [5, 6].

To address these issues, dimensional approaches such as the Hierarchical Taxonomy of Psychopathology (HiTOP) and the Research Domain Criteria (RDoC) initiative [5, 7-9] have been proposed. HiTOP conceptualizes psychopathology along a continuum from specific symptoms to broad spectra, culminating in a general psychopathology (*p*) factor. In contrast, RDoC focuses on domains of functioning across multiple levels of analysis (e.g., molecular, cellular, behavioral) to identify potential causal mechanisms. Although HiTOP and RDoC differ in their approaches, both aim to move beyond traditional diagnostic categories to capture the complexity of psychopathology [10]. However, the integration of these frameworks, particularly through genetic investigations, remains largely unexplored.

Genetic research has advanced our understanding of psychopathology dimensions. Twin and family studies have shown that both externalizing (EXT; e.g., antisocial behavior, substance use) and internalizing (INT; e.g., anxiety, depression) behaviors have shared genetic bases [11, 12]. Genome-wide association studies (GWAS), and, more recently, genomic structural equation modeling (gSEM), have provided a more nuanced understanding of the genetic architecture underlying these dimensions [13-17]. However, gaps remain. First, although subclinical traits have been included in studies of EXT behaviors [14, 18], their integration in the study of INT psychopathology has been limited, which risks overlooking genetic contributions across the full range of symptom severity. Second, previous studies have not typically focused

on genetic effects that are shared across *both* EXT and INT psychopathology. An exception is a GWAS of childhood behavior problems, which identified just two significant loci [19]. Thus, our understanding of the common genetic factors that drive comorbidities between these two spectra is limited, despite evidence that individuals who experience comorbid EXT and INT problems are uniquely vulnerable to poor economic and social outcomes [20], greater criminality [21], and suicide attempts [22].

Our study addresses these gaps by using a joint multivariate GWAS to explore shared genetic contributions across both clinical and subclinical traits within the EXT and INT spectra. We integrate the dual perspectives of HiTOP and RDoC by examining genetic associations both hierarchically, as suggested by HiTOP, and across multiple units of analysis, as emphasized by RDoC. Specifically, we investigate mechanisms across genes, molecules, cells, circuits, physiology, and behavior to determine whether associations are unique to each psychopathology spectrum or reflect common mechanisms across both. This approach aligns with HiTOP's hierarchical model, which posits that higher-order shared features of psychopathology coexist with specific features at more granular levels. Similarly, the RDoC framework informs our analysis by guiding the exploration of genetic mechanisms across biological levels. By bridging these models through a genetic lens, our study offers a novel integration aimed at deepening our understanding of the genetic architecture of EXT and INT psychopathology and inform more precise approaches to diagnosis, treatment, and classification of the phenomena that comprise them.

## Methods

### Unit of Analysis: Genes

*Summary Statistics.* The selection of EXT summary statistics was based on previous studies in European ancestry (EUR) [14, 18] and existing theory [7]; Supplementary Table 1). We included the largest GWAS of attention deficit hyperactivity disorder (ADHD; n = 225,534; [23] and four substance use disorders [SUDs; i.e., alcohol (AUD; n = 753,248; [24], cannabis

(CanUD; n = 886,025; [25], opioid (OUD; n = 425,944; [26], and tobacco (TUD; n = 495,005; [27]]. We also included subclinical measures reflecting disinhibition and antagonism [age of first sexual intercourse (AgeSex; reverse-coded; n = 317,694; [14], general risk tolerance (Risk; n = 431,126; [14, 18], number of sexual partners (NumSex; n = 370,711; [14, 18], antisocial behavior (ASB; n = 16,400; [28], and automobile speeding propensity (n = 404,291; [18]].

INT summary statistics were selected to capture the spectrum of eating pathology, fear, and distress [7]; Supplementary Table 1). The selected studies included the largest available GWAS of INT disorders: (1) anorexia nervosa (AN; n = 72,517; [29], (2) major depressive disorder (MDD; n = 1,074,629; [30], and (3) posttraumatic stress disorder (PTSD; n = 214,408; [31]. We also included irritability (http://www.nealelab.is/uk-biobank/; n = 345,231), loneliness (n = 490,689; [32], subjective wellbeing (SWB; reverse-coded; n = 298,420; [33], miserableness (http://www.nealelab.is/uk-biobank/; n = 355,182), and anxiety (ANX; n = 280,490). To detect variants associated with both anxiety disorders (i.e., panic disorder, agoraphobia, generalized anxiety disorder (GAD), specific phobia, and social phobia) and subclinical anxiety (i.e., GAD-2 scores), we combined three GWAS of these traits [34-36] via multi-trait analysis of GWAS [37].

*Genomic Structural Equation Modeling.* Using linkage disequilibrium score regression (LDSC; [38]) in GenomicSEM [16], we calculated genetic correlations ($r_g$) between the input traits. Single nucleotide polymorphisms (SNPs) were filtered using EUR HapMap3 reference panels [39], and SNPs with MAF < 0.01 were removed. LDSC uses ancestry-matched 1000 Genomes Phase 3 linkage disequilibrium (LD) scores [40]. When available in the summary statistics, SNP-level sample sizes were specified. Otherwise, the sample size was calculated by summing effective sample sizes across the GWAS cohorts [41]. Traits with an average $r_g$ < 0.20, representing a small effect size, were excluded from gSEM models.

We fit five confirmatory factor analyses (CFAs) based on existing theories of psychopathology [42-44]. Because our objective was to test theoretically driven models, we did not perform exploratory factor analyses. First, we evaluated a correlated factors model with two

factors representing EXT and INT psychopathology. Next, we tested a three-factor model to explore whether SUDs were distinct from other EXT behaviors. We also fit a bifactor model consisting of a general psychopathology factor on which all traits loaded, and orthogonal EXT and INT psychopathology factors. A higher-order model was tested with a second-order factor (EXT+INT) representing genetic effects shared by the two spectra. Finally, we fit a unidimensional, or *p*-factor, model where all traits loaded onto one latent factor. In CFA models (except the three-factor model), the residuals of the SUDs were allowed to correlate to account for shared variance not explained by the broader EXT factor. We evaluated models with chi-square, Akaike information criterion (AIC), comparative fit index (CFI; > 0.90), and standardized root mean squared residual (SRMR; < 0.08) fit statistics. We set an *a priori* minimum threshold of 0.35 for factor loadings to ensure that each trait meaningfully contributed to the factor.

*Genome-Wide Association Studies.* For GWAS, we excluded SNPs with MAF < 0.01 and imputation scores < 0.6. SNPs were regressed on the model latent variable(s) using diagonally weighted least squares estimation. We calculated factor-specific $Q_{SNP}$ values [13], a measure of heterogeneity that reflects the extent to which a SNP exerts effects through the common factor or, instead, differentially across a factor's indicators. SNPs with a Q $\chi^2$ p-value < $5x10^{-8}$ were removed prior to conducting downstream analyses. We identified lead SNPs using LD clumping in PLINK 1.9 [45] with a 3-Mb range, $r^2 > 0.10$, and the EUR 1000 Genomes reference panel [40].

*Gene Expression and Enrichment.* Gene-based tests, gene-set enrichment, and gene-tissue expression analyses were conducted using MAGMA [46] in FUMA [47]. We examined gene expression in BrainSpan [48] and GTEx v8 [49] tissues. Gene associations were identified based on (1) position, (2) expression quantitative trait loci (eQTLs) from PsychENCODE [50] and GTEx v8 [49] brain tissues, and (3) chromatin interactions.

*Transcriptome-Wide Association Studies.* We conducted two transcriptome-wide association studies (TWAS) on each factor [51]. First, we used S-MultiXcan [51], which

prioritizes likely causal genes by jointly predicting expression across multiple tissues. We examined expression across 13 brain tissues in GTEx v8 [49]. To complement this approach, we used S-PrediXcan [52] and weights trained on transcriptional differences in the frontal and temporal cortices of psychiatric cases and controls [53, 54]. A Bonferroni correction was applied to to account for multiple testing.

*Polygenicity and Genetic Overlap.* Univariate MiXeR analyses [55] were conducted to estimate polygenicity (i.e., the number of causal variants required to explain 90% SNP heritability) and discoverability (i.e., the average effect size of causal variants; [56]). Bivariate models identified the proportion of unique and shared causal variants for EXT and INT. In contrast to genetic correlations, MiXeR accounts for polygenic overlap regardless of the direction of effect of causal variants.

**Unit of Analysis: Molecules**

To identify drug repurposing targets, we used five gene annotation approaches: (1) MAGMA [46], (2) chromatin interactions, (3) eQTLs, (4) S-MultiXcan, and (5) S-PrediXcan [51, 52]. To avoid unreliable associations, we queried druggable genes [57] identified by multiple biological annotation sources using the Drug-Gene-Interaction Database (DGIdb; [58]). For the first-order factors, we limited drug repurposing to genes that showed specificity of association with *either* spectrum but not both.

**Unit of Analysis: Cells**

To investigate gene expression at the cellular level, we used single-cell RNA sequencing (scRNA-seq) datasets from 15 human brain cell expression profiles [59-62]. The analysis involved three steps: (1) identifying significant cell types within each dataset using MAGMA, (2) performing within-dataset conditional analyses to determine independent associations among correlated cell types, and (3) conducting cross-dataset conditional analyses to assess whether associations reflected shared genetic signals across different datasets.

**Unit of Analysis: Circuits**

We used BrainXcan [63] to examine associations between externalizing, internalizing, EXT+INT, and 327 brain image-derived phenotypes (IDPs) from structural (T1-weighted) and diffusion magnetic resonance images (dMRIs) via ridge regression. Effect sizes and p-values were adjusted using LD block-based permutation, and Bonferroni correction was used to account for multiple testing.

**Unit of Analysis: Physiology**

To evaluate potentially causal impacts on 15 physical health traits, we conducted Generalized Summary-data-based Mendelian Randomization (GSMR; [64]). Traits were chosen across four domains associated with psychopathology: (1) pain, (2) general health, (3) cardiovascular disease, and (4) chronic illness [65-67]. The pain domain comprised summary statistics from GWAS of pain intensity [68], multisite chronic pain [69], and back pain [70]. General health indices included summary statistics from the UK Biobank for longstanding illness, disability, or infirmity; hospitalization; and age at death. Cardiovascular disease comprised GWAS of five traits: (1) heart failure [71], (2) stroke [71], (3) myocardial infarction [72], (4) hypertension (http://www.nealelab.is/uk-biobank/), and (5) abdominal aortic aneurysm [71]. Finally, we selected four GWAS of chronic illnesses: (1) type 2 diabetes [73], (2) inflammatory bowel disease (IBD; [74], (3) chronic obstructive pulmonary disease (COPD; [71], and (4) asthma [71]. Genetic instruments with significant pleiotropic effects were removed using the heterogeneity in dependent instruments outlier method [64]. A Bonferroni-adjusted p-value was applied to account for multiple testing.

**Unit of Analysis: Behaviors**

Using the Complex Trait Genetics Virtual Lab [75], we calculated genetic correlations between each factor and 1,437 phenotypes. Genetic correlations were calculated using LDSC [38] and EUR 1000 Genomes Phase 3 data [40] as LD references. To account for multiple testing, a Benjamini-Hochberg false discovery rate (FDR) correction was applied. This analysis allowed us to explore how genetic factors underlying EXT and INT spectra are correlated with a

diverse range of phenotypes, including psychological traits and behaviors. By examining these genetic correlations across a broad set of phenotypes, we aimed to identify potential behavioral expressions or tendencies that are genetically linked to the two psychopathology spectra.

## Results

### Unit of Analysis: Genes

*Confirmatory Factor Analysis.* Of 18 traits examined (Supplementary Figure 1), two (automobile speeding propensity and anorexia nervosa) were excluded, as they were weakly associated with all others (mean $r_g < 0.20$). Using 16 traits, we tested several CFA models (Figure 1 and Supplementary Figures 2-4). A general psychopathology factor model did not provide adequate fit ($\chi^2(98) = 8965.28$, AIC = 9041.28, CFI = 0.79, and SRMR = 0.15). Although a bifactor model fit well ($\chi^2(88) = 3472.02$, AIC = 3568.02, CFI = 0.92, and SRMR = 0.07), it led to several weak (< 0.35) and one negative standardized loading, possibly due to overfitting the data. Because of its limited interpretability and several traits that did not adequately load onto the factors, we did not perform GWAS on factors from this model.

Two equivalent CFA models provided adequate fit and strong factor loadings: (1) a two-factor correlated model and (2) a higher-order factor model. Fit statistics for both were $\chi^2(97) = 3877.82$, AIC = 3955.82, CFI = 0.91, and SRMR = 0.09. Although a three-factor model that separated SUDs from other externalizing traits also provided a reasonable fit to the data ($\chi^2(101) = 4401.18$, AIC = 4471.18, CFI = 0.90, and SRMR = 0.09), it did not improve upon the fit of the simpler two-factor model and did not justify the added complexity. To ensure identification in the higher order model, the loadings onto the second-order factor were constrained equal to the square root of the genetic correlation between externalizing and internalizing [76]. This approach, though under-identified, is a commonly accepted one for testing a factor that explains shared variance between two correlated dimensions, facilitating interpretation of the shared genetic architecture of EXT and INT psychopathology.

*Genome-Wide Association Studies.* Eighty-three GWS SNPs exhibited heterogeneous effects across the EXT spectrum. Among these (Supplementary Table 2), a plurality was most strongly associated with AgeSex (47.62%), followed by TUD (16.67%). After removing heterogenous SNPs, a GWAS of EXT identified 409 GWS lead SNPs (Supplementary Table 3). Of these, 92 (22.49%) were not identified or within ±1000 kb of SNPs identified by any of the input GWAS, and four were not previously associated with any EXT trait using the same threshold. Three of the four novel SNPs were on chromosome 4 (rs1961547, rs9316, and rs7682762), with the fourth on chromosome 22 (rs1473811). These SNPs were associated with chronotype, schizophrenia, and social support, among other traits (Supplementary Table 4).

For INT, 41 GWS SNPs exhibited significant heterogeneity, with most (78.05%) showing the strongest associations with MDD (Supplementary Table 5). After removing heterogeneous SNPs, there were 85 lead SNPs (Supplementary Table 6). Of these, 23 (27.06%) were in loci that were not significant in the input GWAS, and two were not previously associated with any INT trait (Supplementary Table 7). The two novel associations were on chromosomes 3 and 4 (rs1381763 and rs4698408, respectively) and were associated with neuroticism and depression.

For EXT+INT, 39 GWS SNPs exhibited heterogeneous effects (Supplementary Table 8). Of these, a plurality (41.03%) was most strongly associated with AgeSex, followed by AUD (17.95%). After removing heterogeneous SNPs, there were 256 lead SNPs associated with EXT+INT (Supplementary Table 9), 38 of which (14.84%) were not in loci identified by any of the input GWAS. All were previously associated with at least one EXT or INT phenotype (Figure 2).

*Gene Expression and Enrichment.* MAGMA identified 493 genes associated with EXT, including 326 uniquely associated with EXT but not INT (Supplementary Table 10). These EXT-specific genes included *CADM2* and *FTO*. Although not differentially expressed during any developmental period, the EXT-specific genes were differentially expressed in four brain tissues: the hippocampus, amygdala, putamen, and caudate (Supplementary Figure 5). A gene-

set related to mRNA binding was the only significant association with EXT ($p_{bon}$ = 0.04; Supplementary Table 11). There were 146 genes associated with INT, 31 of which were specific to INT (Supplementary Table 12). The 31 genes were not differentially expressed in any developmental stages or tissue types. No gene-sets were significant after Bonferroni correction (Supplementary Table 11). There were 321 genes associated with EXT+INT (Supplementary Table 13), including 29 that were also identified by both first-order factors (e.g., *DRD2*, *NCAM1*, and *DCC*). Gene expression for EXT+INT was significantly upregulated during early and early mid-prenatal periods and downregulated during early childhood (Supplementary Figure 6). EXT+INT genes were upregulated in 10 brain regions, including the frontal cortex, amygdala, anterior cingulate cortex, and hippocampus (Supplementary Figure 6). No gene sets were significant after Bonferroni correction (Supplementary Table 11).

*Transcriptome-Wide Association Studies.* Using S-MultiXcan to predict SNP effects on gene expression across 13 brain tissues revealed 352 significant genes for EXT, 141 for INT, and 238 for EXT+INT (Figure 3, Supplementary Tables 14-16, and Supplementary Figure 7). TWAS using PsychENCODE data identified 207 genes for EXT, 52 for INT, and 124 for EXT+INT (Supplementary Tables 17-19 and Supplementary Figure 8). Forty-five genes were identified by both TWAS for EXT, 21 for INT, and 36 for EXT+INT (Supplementary Figures 9-11).

*Polygenicity and Genetic Overlap.* The EXT and INT spectra displayed similar levels of polygenicity (12,600 and 13,200 causal variants, respectively). However, INT had lower discoverability ($\hat{\sigma}_{\beta}^{2}$ = 1.40 x $10^{-5}$) than EXT ($\hat{\sigma}_{\beta}^{2}$ = 1.44 x $10^{-4}$), suggesting that INT may require larger samples to allow detection of significant SNPs. Despite a MiXeR-estimated $r_g$ of 0.37, almost all causal variants (96.83% of EXT and 92.42% of INT; Supplementary Figure 12) overlapped (AIC = 12.30, BIC = 4.06, where positive values indicate the model explains the data better than the $r_g$). Of the shared causal variants, 62.92% were estimated to have concordant effect directions.

**Unit of Analysis: Molecules**

Among the 1,759 genes identified for EXT, 308 were druggable targets. Sixty were identified by at least two methods, and 52 were specific for EXT. When queried in DGIdb, the 52 genes yielded 492 drug-gene interactions (Supplementary Table 23), including with dextroamphetamine (used to treat ADHD), phenobarbital (used to prevent withdrawal symptoms from benzodiazepines and alcohol), baclofen (used to treat AUD), naltrexone (used to treat AUD and OUD), naloxone (used to reverse opioid overdose and to prevent intravenous abuse of buprenorphine when treating OUD), and methadone (used to treat OUD). Interactions with antimigraine, anti-inflammatory, and anticonvulsant drugs (e.g., topiramate and lamotrigine) were also identified. Most identified drugs have regulatory approval (64.84%).

Biological annotation identified 454 genes associated with INT, 60 of which were druggable targets. Fifteen were identified by at least two methods, and 7 were specific for INT, yielding 292 drug-gene interactions (Supplementary Table 24). Drug targets included antidepressants and antipsychotics. Unlike for EXT, drugs identified for INT (82.33%) are not currently approved, suggesting that they are potential candidates for treating INT psychopathology.

For EXT+INT 1,138 genes were identified, 204 of which were druggable targets, 47 of which were implicated by more than one method. Using DGIdb, we identified 460 drug-gene interactions (Supplementary Table 25), many of which were also present in the INT or EXT results. Most of these drugs (75.52%) are not currently approved.

**Unit of Analysis: Cells**

Using scRNA-seq datasets, EXT was significantly associated with dopaminergic and GABAergic neurons and neuroblasts from embryonic brain samples (GSE76381), human cortical neurons and hybrid cells (GSE67835), and pyramidal neurons from the cornu ammonis (CA1) region of the hippocampus[61]. After conditional analyses, independent significant associations remained for GABAergic neurons, cortical neurons, and hippocampal neurons

(Supplementary Figure 13). The only significant cell-type association for INT was with GABAergic neurons (GSE76381), which was not independently significant after conditional analyses. EXT+INT showed significant associations with dopaminergic neurons (GSE76381), GABAergic neurons (GSE76381), and cortical neurons (GSE67835), though these associations were not independently significant.

**Unit of Analysis: Circuits**

After Bonferroni correction, 8 T1 (Supplementary Figure 14) and 12 dMRI IDPs (Supplementary Figure 15) were associated with EXT (Supplementary Table 20), including greater gray matter volume in the thalamus, caudate nuclei, and occipital pole, and lower gray matter volume in the right ventral striatum and left amygdala. Significant associations were also found with intra-cellular volume fraction or orientation dispersion indices (ODI) in the medial lemniscus, cerebral peduncle, and middle cerebellar peduncle, among others. INT was significantly associated with lower gray matter volume in the left subcallosal cortex (Supplementary Table 21 and Supplementary Figures 16-17). EXT+INT showed negative associations with gray matter volume in the left amygdala and subcallosal cortex (Figure 4), positive associations with ODI in the medial lemniscus and left cerebellar peduncle, and negative associations with ODI in the right external capsule (Supplementary Table 22 and Supplementary Figures 18-19).

**Unit of Analysis: Physiology**

EXT had significant positive causal effects on all physical health traits, except age at death and IBD. INT was causally associated with all traits except age at death and abdominal aortic aneurysm. INT had protective effects on IBD ($b_{xy}$ = -0.32, SE = 0.09, $p$ = 0.0004), but stronger positive associations than EXT with all three pain phenotypes, all five cardiovascular diseases, and three of four chronic illnesses (Figure 5). Like EXT, EXT+INT had positive causal effects on all physical health traits except age at death and IBD (Supplementary Table 29).

**Unit of Analysis: Behaviors**

After Bonferroni correction, there were 413 significant $r_g$ with EXT (Supplementary Table 26 and Supplementary Figure 20). Tobacco phenotypes were among the strongest (current smoking: $r_g$ = 0.79, SE = 0.02; maternal smoking around birth: $r_g$ = 0.71, SE = 0.03; and ever smoked: $r_g$ = 0.62, SE = 0.02), along with lower socioeconomic status (Townsend deprivation index: $r_g$ = 0.68, SE = 0.03; experiencing financial difficulties: $r_g$ = 0.58, SE = 0.03) and educational attainment ($r_g$ = -0.44, SE = 0.02). After Bonferroni correction, 311 phenotypes were genetically correlated with INT (Supplementary Table 27 and Supplementary Figure 21). Among the strongest of these were psychiatric phenotypes, such as mood swings ($r_g$ = 0.90, SE = 0.01), neuroticism ($r_g$ = 0.89, SE = 0.01), and feeling fed-up ($r_g$ = 0.82, SE = 0.01). INT was also genetically correlated with several types of pain (abdominal: $r_g$ = 0.60, SE = 0.04; facial: $r_g$ = 0.51, SE = 0.08; chest: $r_g$ = 0.49, SE = 0.03; and multisite chronic pain: $r_g$ = 0.49, SE = 0.03). There were 474 significant $r_g$ with EXT+INT, with most being like the first-order factors (Supplementary Table 28 and Supplementary Figure 22).

## Discussion

Our findings demonstrate the value of integrating the HiTOP and RDoC frameworks to advance our understanding of the biological underpinnings of EXT and INT psychopathology. By combining HiTOP's dimensional, hierarchical approach with RDoC's focus on functional domains, we identified both shared and distinct mechanisms across multiple units of analysis, from genes to behaviors. This integrative perspective revealed common pathways contributing to these spectra and uncovered unique biological signatures that distinguish EXT and INT traits, offering a more comprehensive view of how the two dimensions of psychopathology converge or diverge at various biological levels.

We identified considerable overlap in genetic variants contributing to EXT and INT spectra in line with HiTOP's hierarchical model, which posits that higher-order psychopathology factors emerge from shared features between lower-level spectra. For example, we found that genes like *DRD2, NCAM1,* and *DCC* mapped to both spectra, highlighting potential genetic

mechanisms contributing to shared liability. Prior research also showed that *DRD2* and *DCC* exhibit pleiotropic effects across multiple forms of psychopathology and are linked to psychopathology through effects shared with cognitive performance [77]. In addition, we identified genes with greater specificity: *CADM2* and *FTO* were uniquely associated with EXT traits, suggesting the involvement of pathways related to impulsivity, risk-taking, and reward sensitivity—key components of RDoC domains, such as the cognitive and positive valence systems.

At the molecular level, drug repurposing analyses identified several potential therapeutic targets. For EXT psychopathology, we identified perturbagens, including dextroamphetamine and naltrexone, that target dopaminergic and opioid pathways. These medications, commonly used to treat conditions like ADHD and SUDs, may also be effective for managing other EXT conditions. For INT, we identified perturbagens like antidepressants and antipsychotics, which target neurotransmitter systems consistent with those in the RDoC's negative valence system domain. Notably, while most identified drugs for EXT psychopathology are approved for clinical use, most identified for INT are not, suggesting new therapeutic avenues for treating INT traits and disorders.

Genes associated with each spectrum also showed specific expression patterns in brain cell types. For EXT-related genes, there was enriched expression in GABAergic, cortical, and pyramidal hippocampal neurons, cell types that align with RDoC constructs related to cognitive control and acute threat. In contrast, genes linked to INT psychopathology were more narrowly associated with GABAergic neurons involved in inhibitory signaling, consistent with RDoC's negative valence systems, particularly the acute threat construct.

The neural signatures at the cellular level were further reflected in broader structural and functional brain alterations linked to each spectrum. Among other regions, EXT was associated with lower volume in the ventral striatum and amygdala. These regions, central to emotional regulation and impulse control, are both implicated in the reward learning construct of RDoC's

positive valence systems. INT was linked to lower gray matter volume in the left subcallosal cortex, a region involved in mood regulation. Consistent with cross-spectra risk, the EXT+INT factor was associated with lower grey matter volume in both the amygdala (implicated by EXT) and subcallosal cortex (implicated by INT). Lower left amygdala volume has been shown to mediate the relation between childhood threat exposure and the development of EXT and INT symptoms [78]. Similarly, lower left subcallosal cortical volume is a potential mediator of associations between personality and emotional states (e.g., subjective well-being) and psychiatric disorders (e.g., alcohol dependence; [79]. EXT+INT liability was also related to white matter fiber differences, which may signify disruptions in neural communication that contribute to psychopathology [80].

Mendelian Randomization analyses highlighted potentially causal effects of genetic liability for EXT and INT traits on physical health. INT liability was causally associated with risk for chronic pain, cardiovascular disease, and other chronic illnesses. Previous research also supports likely causal effects of genetic risk for INT traits/disorders on localized pain [81, 82] and disease outcomes [83]. EXT liability also exhibited associations, albeit generally weaker, with adverse health outcomes, indicating that liability for both spectra contribute to physical health risk.

Genetic correlations of the two spectra revealed distinct but overlapping phenotypes. EXT was predominantly genetically correlated with substance use and socioeconomic status, while most of the genetic correlations with INT were for anxiety and mood-related traits. Despite this, there was some overlap in the top genetic correlations for the two spectra, particularly for pain-related phenotypes. EXT+INT was also genetically correlated with several pain-related phenotypes, as well as substance use and mood. These patterns support differentiation between the two spectra, yet also highlight areas of overlap where common genetic factors exhibit pleiotropy with other behavioral manifestations.

**Limitations**

A key limitation of this study is the inclusion of only European-ancestry individuals. Although GWAS of EXT- and INT-related *disorders* are available for some non-European ancestry groups, they are limited for more precise, *non-disorder* psychiatric phenotypes. Nonetheless, some research suggests that a similar factor structure exists in African-ancestry individuals. At the disorder level, a gSEM of African-ancestry individuals identified substance use and psychiatric disorder factors that roughly aligned with EXT and INT, respectively [84]. Additionally, the analyses showed that a higher-order factor accounted for genetic variance shared by substance use and psychiatric disorders, which to a degree corresponds to our EXT+INT factor. With the growth of diverse biobanks [85] and deep phenotyping [86], direct replication in other ancestries should soon be possible and given high priority.

Because we focused specifically on traits within the EXT and INT spectra, the current study does not fully capture the broad scope of psychopathology. Future research could benefit from incorporating a wider range of traits, including those related to dimensions such as thought disorders or personality disorders, to provide a more comprehensive understanding of the genetic architecture underlying psychopathology. Expanding the range of included traits and exploring additional domains will allow for more nuanced examinations of shared and unique genetic factors across different psychopathology dimensions.

**Conclusions**

Together, these findings illustrate the power of integrating HiTOP's dimensional, hierarchical approach with RDoC's functional domain framework to deepen our understanding of the mechanisms underlying EXT and INT psychopathology. By identifying the convergence and divergence of biological pathways across multiple levels of analysis, this approach provides a more nuanced view of psychopathology, which can guide targeted interventions to address the overlapping and unique aspects of these complex conditions. This integrated framework offers promising avenues for advancing precision mental health care and treatment development.

**Financial Support:** This work was supported by the Veterans Integrated Service Network 4 Mental Illness Research, Education and Clinical Center and by Department of Veterans Affairs grants I01 BX004820 to H.R.K., National Institute on Alcohol Abuse and Alcoholism grant AA028292 to R.L.K, and KNAW (Royal Netherlands Academy of Arts and Sciences) Academy Professor Award (PAH/6635) to D.I.B. The funders had no role in study design, data collection or analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Dr. Kranzler is a member of advisory boards for Altimmune, Clearmind Medicine, Dicerna Pharmaceuticals, Enthion Pharmaceuticals, Lilly Pharmaceuticals, and Sophrosyne Pharmaceuticals; a consultant to Altimmune and 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. Drs. Kranzler and Gelernter hold U.S. patent 10,900,082 titled: "Genotype-guided dosing of opioid agonists," issued 26 January 2021. The other authors have no disclosures to make.

**References**

1. Kessler, R.C., et al., *Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication.* Archives of General Psychiatry, 2005. **62**(6): p. 617-627.
2. Kessler, R.C., et al., *Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States: Results From the National Comorbidity Survey.* Archives of General Psychiatry, 1994. **51**(1): p. 8-19.
3. Lee, P.H., Y.-C.A. Feng, and J.W. Smoller, *Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders.* Biological Psychiatry, 2021. **89**(1): p. 20-31.
4. Galatzer-Levy, I.R. and R.A. Bryant, *636,120 Ways to Have Posttraumatic Stress Disorder.* Perspect Psychol Sci, 2013. **8**(6): p. 651-62.
5. Kozak, M.J. and B.N. Cuthbert, *The NIMH Research Domain Criteria Initiative: Background, Issues, and Pragmatics.* Psychophysiology, 2016. **53**(3): p. 286-297.
6. Waszczuk, M.A., et al., *Redefining phenotypes to advance psychiatric genetics: Implications from hierarchical taxonomy of psychopathology.* Journal of Abnormal Psychology, 2020. **129**(2): p. 143-161.
7. Kotov, R., et al., *The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies.* Journal of Abnormal Psychology, 2017. **126**(4): p. 454-477.
8. Cuthbert, B.N., *Research Domain Criteria: toward future psychiatric nosologies.* Dialogues in Clinical Neuroscience, 2015. **17**(1): p. 89-97.
9. Kotov, R., et al., *The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence.* Annual Review of Clinical Psychology, 2021. **17**(1): p. 83-108.
10. Michelini, G., et al., *Linking RDoC and HiTOP: A new interface for advancing psychiatric nosology and neuroscience.* Clinical Psychology Review, 2021. **86**: p. 102025.
11. Hewitt, J.K., et al., *The analysis of parental ratings of children's behavior using LISREL.* Behavior Genetics, 1992. **22**(3): p. 293-317.

12. Silberg, J.L., et al., *The application of structural equation modeling to maternal ratings of twins' behavioral and emotional problems.* J Consult Clin Psychol, 1994. **62**(3): p. 510-21.

13. Grotzinger, A.D., et al., *Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis.* Nature Genetics, 2022. **54**(5): p. 548-559.

14. Karlsson Linnér, R., et al., *Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction.* Nature Neuroscience, 2021. **24**(10): p. 1367-1376.

15. Lee, P.H., et al., *Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders.* Cell, 2019. **179**(7): p. 1469-1482.e11.

16. Grotzinger, A.D., et al., *Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits.* Nature Human Behaviour, 2019. **3**(5): p. 513-525.

17. Mallard, T.T., A.D. Grotzinger, and J.W. Smoller, *Examining the shared etiology of psychopathology with genome-wide association studies.* Physiological Reviews, 2023. **103**(2): p. 1645-1665.

18. Karlsson Linnér, R., et al., *Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences.* Nat Genet, 2019. **51**(2): p. 245-257.

19. Neumann, A., et al., *A genome-wide association study of total child psychiatric problems scores.* PLOS ONE, 2022. **17**(8): p. e0273116.

20. Vergunst, F., et al., *Association of Childhood Externalizing, Internalizing, and Comorbid Symptoms With Long-term Economic and Social Outcomes.* JAMA Network Open, 2023. **6**(1): p. e2249568-e2249568.

21. Commisso, M., et al., *Association of childhood externalizing, internalizing, comorbid problems with criminal convictions by early adulthood.* Journal of Psychiatric Research, 2024. **172**: p. 9-15.

22. Commisso, M., et al., *Childhood externalizing, internalizing and comorbid problems: distinguishing young adults who think about suicide from those who attempt suicide.* Psychological Medicine, 2023. **53**(3): p. 1030-1037.

23. Demontis, D., et al., *Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains.* Nat Genet, 2023. **55**(2): p. 198-208.

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

25. 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. **55**(12): p. 2094-2103.

26. 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, 2022. **25**(10): p. 1279-1287.

27. Toikumo, S., et al., *Multi-ancestry meta-analysis of tobacco use disorder prioritizes novel candidate risk genes and reveals associations with numerous health outcomes.* medRxiv, 2023.

28. Tielbeek, J.J., et al., *Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior.* JAMA Psychiatry, 2017. **74**(12): p. 1242-1250.

29. Watson, H.J., et al., *Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa.* Nature Genetics, 2019. **51**(8): p. 1207-1214.

30. Als, T.D., et al., *Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses.* Nature Medicine, 2023. **29**(7): p. 1832-1844.

31. Stein, M.B., et al., *Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program.* Nature Genetics, 2021. **53**(2): p. 174-184.

32. Abdellaoui, A., et al., *Phenome-wide investigation of health outcomes associated with genetic predisposition to loneliness.* Human Molecular Genetics, 2019. **28**(22): p. 3853-3865.

33. Okbay, A., et al., *Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses.* Nature Genetics, 2016. **48**(6): p. 624-633.

34. Otowa, T., et al., *Meta-analysis of genome-wide association studies of anxiety disorders.* Molecular Psychiatry, 2016. **21**(10): p. 1391-1399.

35. 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, 2020. **177**(3): p. 223-232.

36. Purves, K.L., et al., *A major role for common genetic variation in anxiety disorders.* Molecular Psychiatry, 2020. **25**(12): p. 3292-3303.

37. Turley, P., et al., *Multi-trait analysis of genome-wide association summary statistics using MTAG.* Nature Genetics, 2018. **50**(2): p. 229-237.

38. Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.* Nature Genetics, 2015. **47**(3): p. 291-295.

39. The International HapMap 3 Consortium, *Integrating common and rare genetic variation in diverse human populations.* Nature, 2010. **467**(7311): p. 52-8.

40. The 1000 Genomes Project Consortium, *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

41. Grotzinger, A.D., et al., *Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution.* Biological Psychiatry, 2023. **93**(1): p. 29-36.

42. Caspi, A., et al., *The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders?* Clinical Psychological Science, 2013. **2**(2): p. 119-137.

43. Markon, K.E., *Bifactor and Hierarchical Models: Specification, Inference, and Interpretation.* Annual Review of Clinical Psychology, 2019. **15**(Volume 15, 2019): p. 51-69.

44. Krueger, R.F. and K.E. Markon, *Reinterpreting Comorbidity: A Model-Based Approach to Understanding and Classifying Psychopathology.* Annual Review of Clinical Psychology, 2006. **2**(Volume 2, 2006): p. 111-133.

45. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* American journal of human genetics, 2007. **81**(3): p. 559-575.

46. de Leeuw, C.A., et al., *MAGMA: Generalized Gene-Set Analysis of GWAS Data.* PLOS Computational Biology, 2015. **11**(4): p. e1004219.

47. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nat Commun, 2017. **8**(1): p. 1826.

48. Li, M., et al., *Integrative functional genomic analysis of human brain development and neuropsychiatric risks.* Science, 2018. **362**(6420): p. eaat7615.

49. The GTEx Consortium, et al., *The GTEx Consortium atlas of genetic regulatory effects across human tissues.* Science, 2020. **369**(6509): p. 1318-1330.

50. Wang, D., et al., *Comprehensive functional genomic resource and integrative model for the human brain.* Science, 2018. **362**(6420).

51. Barbeira, A.N., et al., *Integrating predicted transcriptome from multiple tissues improves association detection.* PLOS Genetics, 2019. **15**(1): p. e1007889.

52. Barbeira, A.N., et al., *Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics.* Nature Communications, 2018. **9**(1): p. 1825.

53. Gandal, M.J., et al., *Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder.* Science, 2018. **362**(6420): p. eaat8127.

54. Jourdon, A., et al., *PsychENCODE and beyond: transcriptomics and epigenomics of brain development and organoids.* Neuropsychopharmacology, 2021. **46**(1): p. 70-85.

55. Frei, O., et al., *Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation.* Nature Communications, 2019. **10**(1): p. 2417.

56. Holland, D., et al., *Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model.* PLOS Genetics, 2020. **16**(5): p. e1008612.

57. Finan, C., et al., *The druggable genome and support for target identification and validation in drug development.* Science Translational Medicine, 2017. **9**(383): p. eaag1166.

58. Freshour, S.L., et al., *Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts.* Nucleic Acids Research, 2021. **49**(D1): p. D1144-D1151.

59. Watanabe, K., et al., *Genetic mapping of cell type specificity for complex traits.* Nature Communications, 2019. **10**(1): p. 3222.

60. La Manno, G., et al., *Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells.* Cell, 2016. **167**(2): p. 566-580.e19.

61. Habib, N., et al., *Massively parallel single-nucleus RNA-seq with DroNc-seq.* Nature Methods, 2017. **14**(10): p. 955-958.

62. Darmanis, S., et al., *A survey of human brain transcriptome diversity at the single cell level.* Proc Natl Acad Sci U S A, 2015. **112**(23): p. 7285-90.

63. Liang, Y., et al., *BrainXcan identifies brain features associated with behavioral and psychiatric traits using large scale genetic and imaging data.* medRxiv, 2022: p. 2021.06.01.21258159.

64. Zhu, Z., et al., *Causal associations between risk factors and common diseases inferred from GWAS summary data.* Nature Communications, 2018. **9**(1): p. 224.

65. Waszczuk, M.A., et al., *General v. specific vulnerabilities: polygenic risk scores and higher-order psychopathology dimensions in the Adolescent Brain Cognitive Development (ABCD) Study.* Psychological Medicine, 2023. **53**(5): p. 1937-1946.

66. Isvoranu, A.M., et al., *Extended network analysis: from psychopathology to chronic illness.* BMC Psychiatry, 2021. **21**(1): p. 119.

67. Zhang, F., et al., *Causal influences of neuroticism on mental health and cardiovascular disease.* Human Genetics, 2021. **140**(9): p. 1267-1281.

68. Toikumo, S., et al., *The genetic architecture of pain intensity in a sample of 598,339 U.S. veterans.* medRxiv, 2023.

69. Johnston, K.J.A., et al., *Genome-wide association study of multisite chronic pain in UK Biobank.* PLOS Genetics, 2019. **15**(6): p. e1008164.

70. Freidin, M.B., et al., *Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals.* PAIN, 2019. **160**(6).

71. Zhou, W., et al., *Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease.* Cell Genom, 2022. **2**(10): p. 100192.

72. Hartiala, J.A., et al., *Genome-wide analysis identifies novel susceptibility loci for myocardial infarction.* European Heart Journal, 2021. **42**(9): p. 919-933.

73. Mahajan, A., et al., *Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation.* Nat Genet, 2022. **54**(5): p. 560-572.

74. Liu, Z., et al., *Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries.* Nature Genetics, 2023. **55**(5): p. 796-806.

75. Cuéllar-Partida, G., et al., *Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses.* bioRxiv, 2019: p. 518027.

76. Loehlin, J.C., *The Cholesky approach: A cautionary note.* Behavior Genetics, 1996. **26**(1): p. 65-69.

77. Lam, M., et al., *Collective genomic segments with differential pleiotropic patterns between cognitive dimensions and psychopathology.* Nature Communications, 2022. **13**(1): p. 6868.

78. Picci, G., et al., *Left amygdala structure mediates longitudinal associations between exposure to threat and long-term psychiatric symptomatology in youth.* Human Brain Mapping, 2022. **43**(13): p. 4091-4102.

79. Wendt, F.R., et al., *Multivariate genome-wide analysis of education, socioeconomic status and brain phenome.* Nature Human Behaviour, 2021. **5**(4): p. 482-496.

80. Kraguljac, N.V., et al., *Neurite Orientation Dispersion and Density Imaging in Psychiatric Disorders: A Systematic Literature Review and a Technical Note.* Biological Psychiatry Global Open Science, 2023. **3**(1): p. 10-21.

81. Yao, C., et al., *Exploring the bidirectional relationship between pain and mental disorders: a comprehensive Mendelian randomization study.* The Journal of Headache and Pain, 2023. **24**(1): p. 82.

82. Williams, F.M.K., et al., *Causal effects of psychosocial factors on chronic back pain: a bidirectional Mendelian randomisation study.* European Spine Journal, 2022. **31**(7): p. 1906-1915.

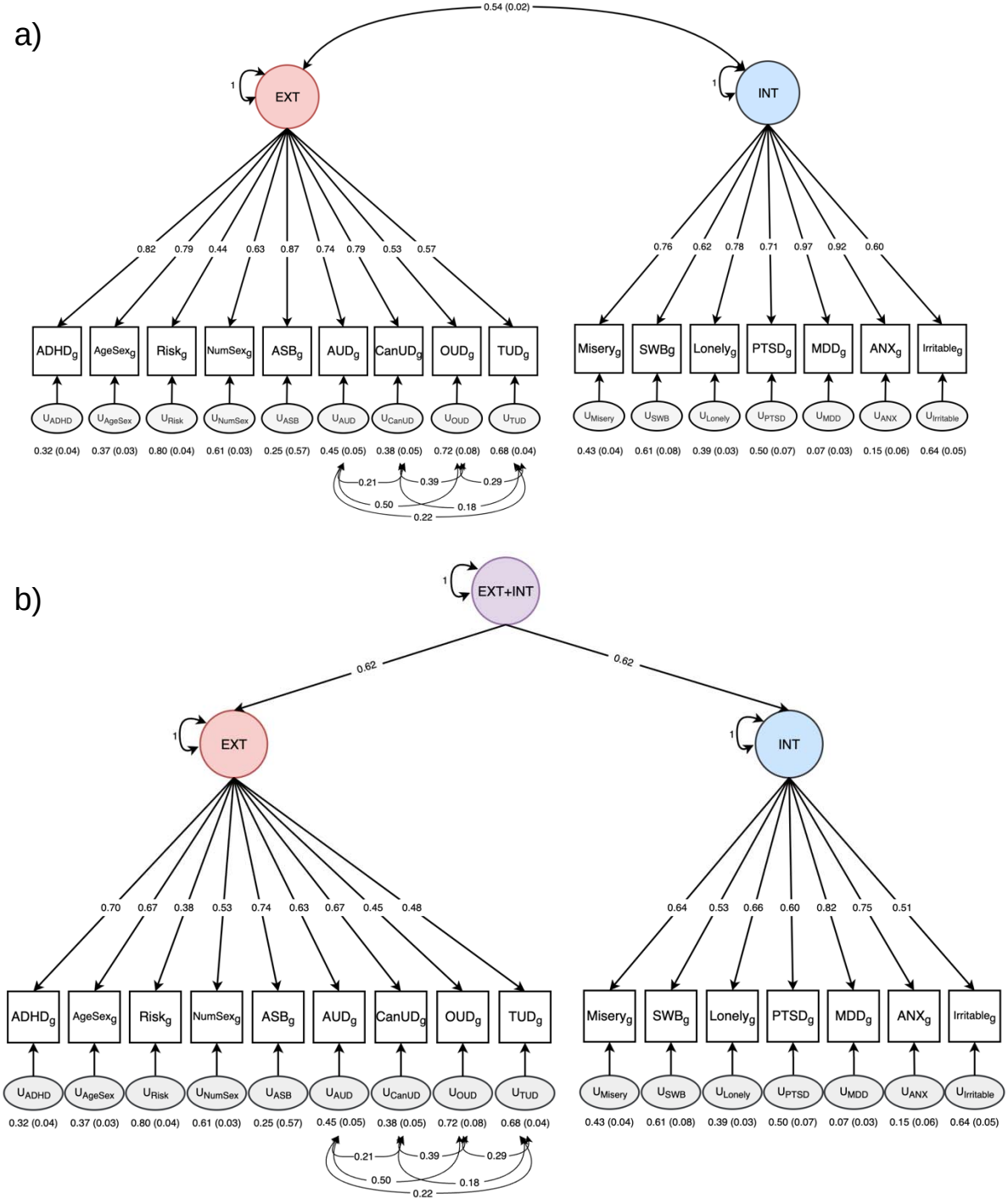
83. Mulugeta, A., et al., *Association between major depressive disorder and multiple disease outcomes: a phenome-wide Mendelian randomisation study in the UK Biobank.* Molecular Psychiatry, 2020. **25**(7): p. 1469-1476.

84. Khan, Y., et al., *Combining Transdiagnostic and Disorder-Level GWAS Enhances Precision of Psychiatric Genetic Risk Profiles in a Multi-Ancestry Sample.* medRxiv, 2024.

85. Bianchi, D.W., et al., *The All of Us Research Program is an opportunity to enhance the diversity of US biomedical research.* Nature Medicine, 2024. **30**(2): p. 330-333.

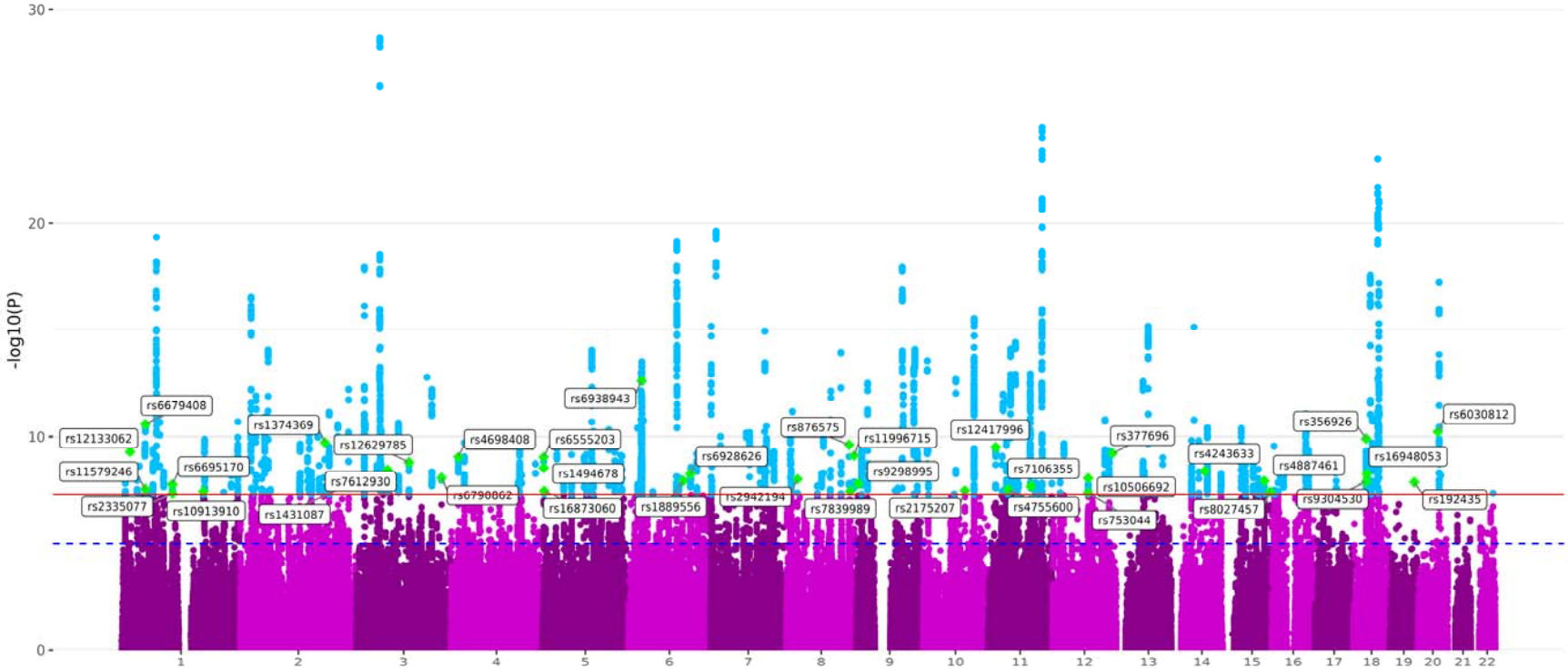
86. Sanchez-Roige, S. and A.A. Palmer, *Emerging phenotyping strategies will advance our understanding of psychiatric genetics.* Nature Neuroscience, 2020. **23**(4): p. 475-480.

*Figure 1.* Confirmatory factor analyses of externalizing and internalizing psychopathology.



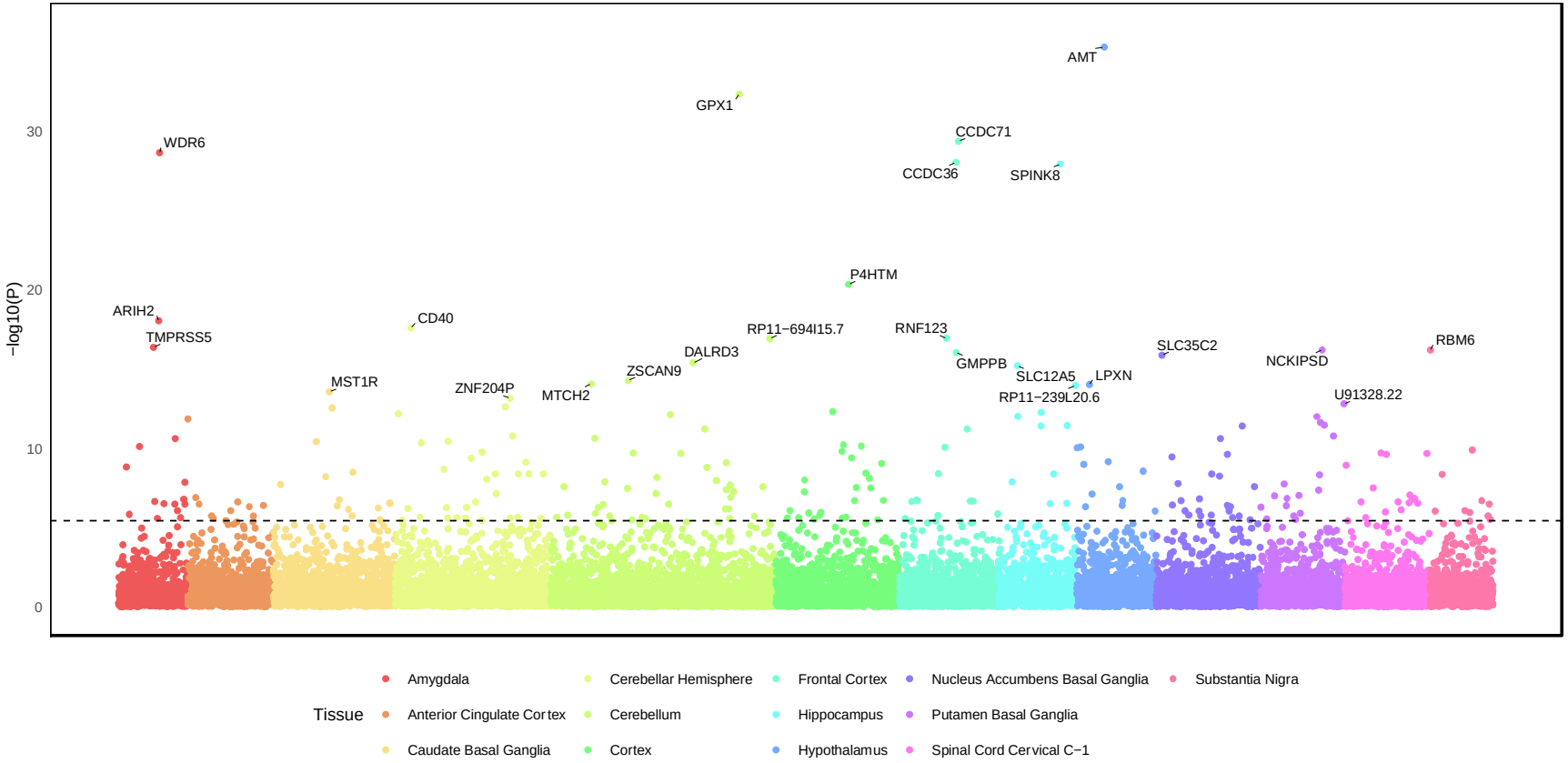

**a)** correlated two-factor model, **b**) higher order factor model. Fit of both: (97) = 3877.82, AIC = 3955.82, CFI = 0.91, SRMR = 0.09. EXT = externalizing, INT = internalizing, ADHD = attention deficit hyperactivity disorder, AgeSex = age at first sex (reverse coded), NumSex = number of sexual partners, ASB = antisocial behavior, AUD = alcohol use disorder, CanUD = cannabis use disorder, OUD = opioid use disorder, TUD = tobacco use disorder, SWB = subjective wellbeing (reverse coded), PTSD = posttraumatic stress disorder, MDD = major depressive disorder, ANX = anxiety.

*Figure 2.* Manhattan plot of the GWAS results for externalizing and internalizing (EXT+INT) liability.



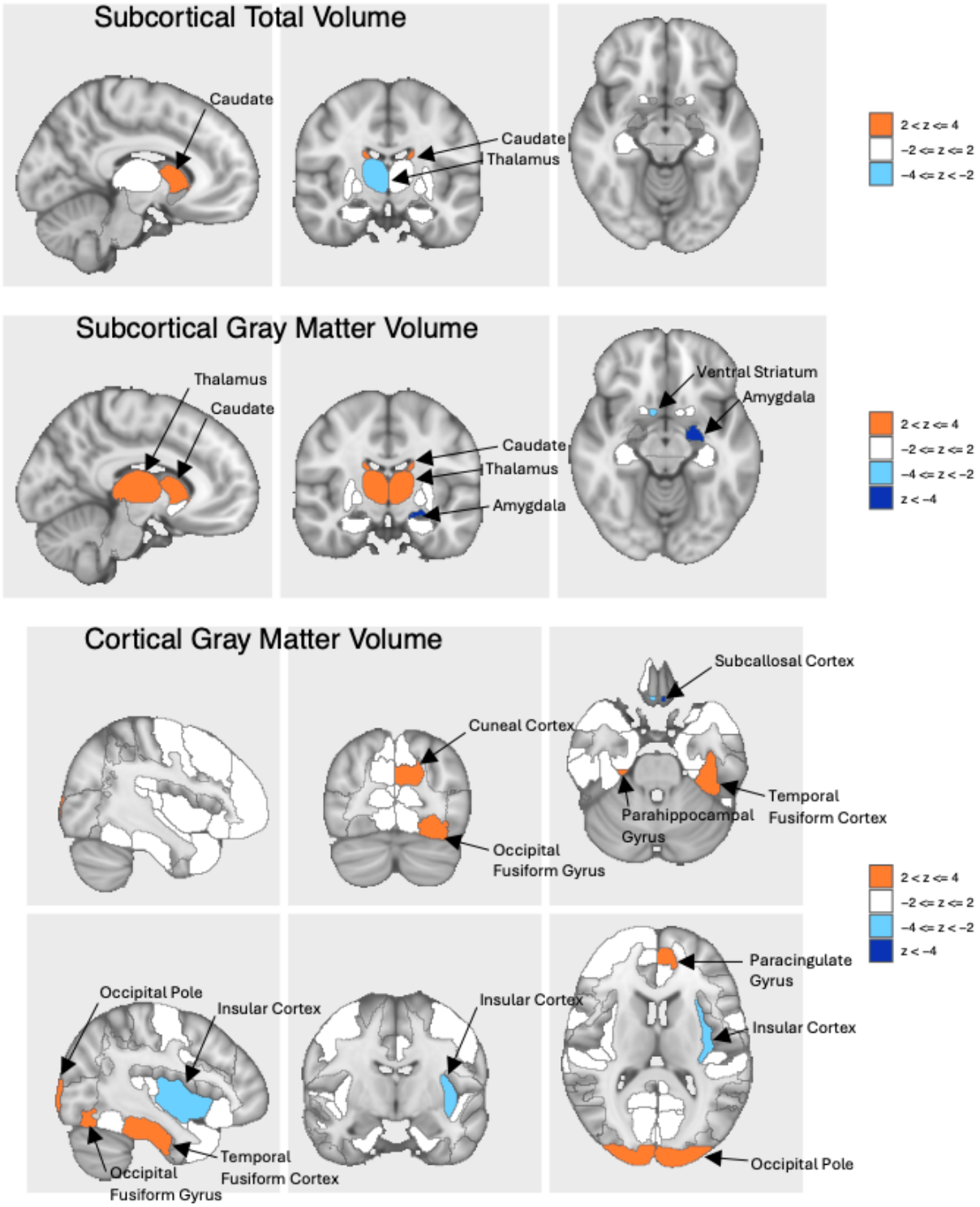

Significant single nucleotide polymorphisms (SNPs) are highlighted in blue. Green diamonds and annotations denote the lead SNPs in loci not identified in the input GWAS for either of the two spectra.

*Figure 3.* Transcriptome wide association study (TWAS) results for the externalizing and internalizing factor across 13 brain tissues using S-MultiXcan.



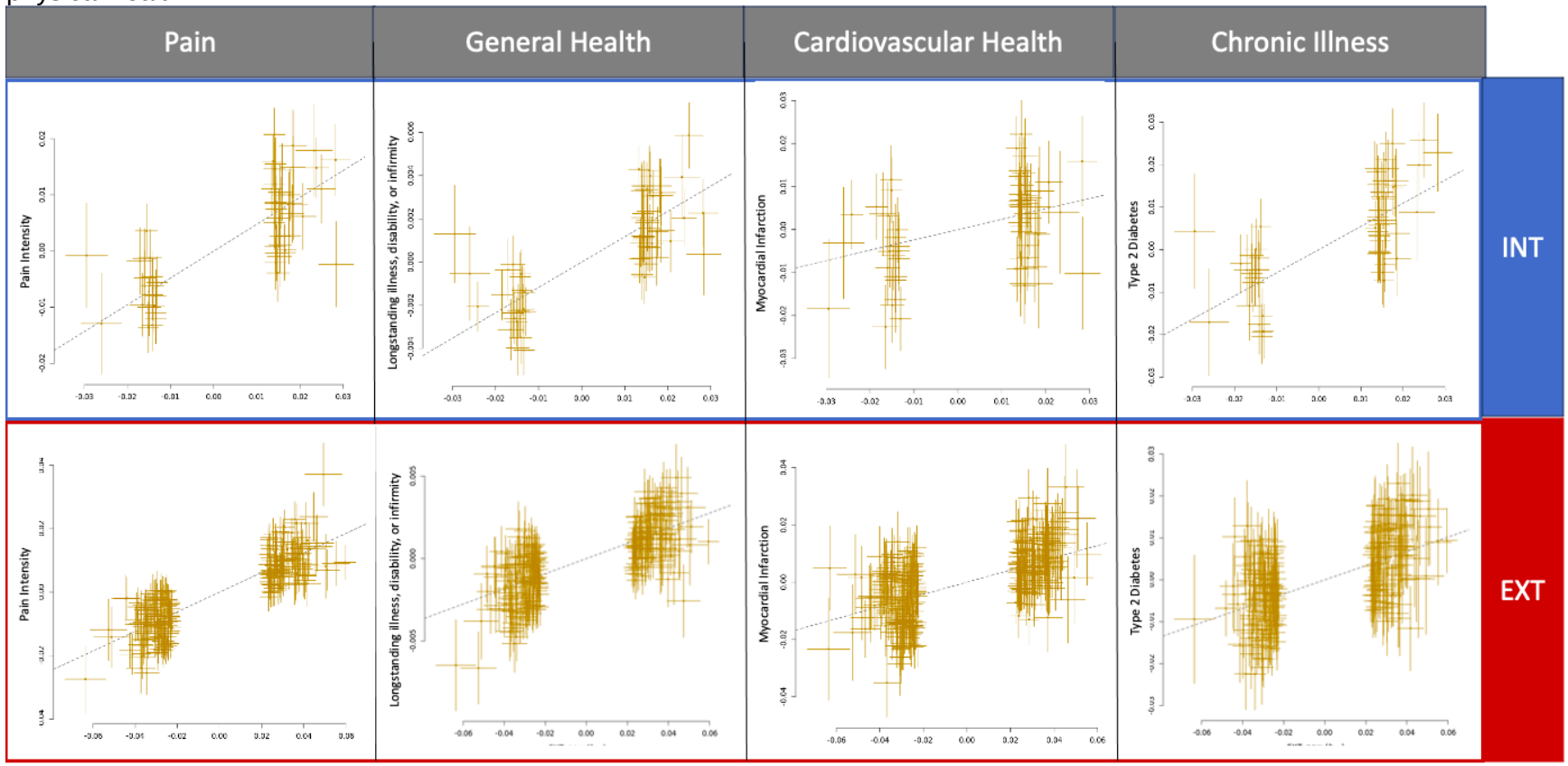

The gene names for the top 25 significant associations are annotated. Significance was determined using a Bonferroni-adjusted p-value of $3.73 \times 10^{-6}$ (0.05/13,406 tests). The dashed line at 5.43 indicates the significance level (-log10($3.73 \times 10^{-6}$). A total of 236 associations were significant after multiple testing correction. Full TWAS results for all the factors can be found in Supplementary Tables 14-19.

*Figure 4.* Partial results of BrainXcan association analysis for the second order externalizing + internalizing factor.




Associations shown are for image-derived phenotypes from structural (T1) magnetic resonance imaging. Full results from BrainXcan are in Supplementary Figures 14-19 and Supplementary Tables 20-22. Blue colors represent reduced volume, and orange represents increased volume.

*Figure 5.* Representative results of generalized summary-data-based Mendelian Randomization (GSMR) across four domains of physical health.




Pain intensity results: INT: $b_{xy}$ = 0.48, SE = 0.04, $p$ = 1.11E-28; EXT: $b_{xy}$ = 0.31, SE = 0.01, $p$ = 1.44E-155. Longstanding illness, disability, or infirmity results: INT: OR = 1.12, SE = 0.01, $p$ = 9.83E-33; EXT: OR = 1.05, SE = 0.002, $p$ = 2.27E-75. Myocardial infarction results: INT: OR = 1.27, SE = 0.07, $p$ = 0.0003; EXT: OR = 1.24, SE = 0.02, $p$ = 6.32E-38. Type 2 Diabetes results: INT: OR = 1.72, SE = 0.06, $p$ = 7.28E-20; EXT: OR = 1.19, SE = 0.02, $p$ = 6.94E-28. All results depicted are significant at a Bonferroni corrected p-value of 0.001. In all analyses, INT/EXT is the exposure, and the physical health trait is the outcome. INT = internalizing, EXT = externalizing, OR = odds ratio. Full results of GSMR analyses, including results for EXT+INT, can be found in Supplementary Table 29.