nature portfolio

https://doi.org/10.1038/s41586-025-09820-3

Supplementary information

Mapping the genetic landscape across 14 psychiatric disorders

In the format provided by the authors and unedited

Supplementary Note
Mapping the Genetic Landscape across 14 Psychiatric Disorders

Table of Contents

Follow-up Analyses: Freeze 3 PTSD and MD

One surprising result was the LDSC estimated genetic correlation of .99 (*SE* = .03) across the Freeze 3 iterations of the PTSD and MD GWAS summary statistics. We ran several follow-up analyses to try and better understand this particular result. First, we note that the attenuation ratio for the HapMap3 SNPs used by LDSC—calculated as (univariate LDSC Intercept – 1) / (mean χ^2-1)—was similar and within a reasonable range for both PTSD (.041) and MD (.048), suggesting that neither disorder might be characterized by uncontrolled confounding that is driving the effects. Second, we examined the SNP-based heritabilities and genetic correlations across the three freezes of PTSD and MD (**Suppl. Fig. 2**). We find that, as would be expected given smaller sample sizes, estimates of this genetic correlation using earlier freezes were imprecise. At the same time, the 95% confidence intervals on these estimates did not include the current estimate of .99. This tentatively indicates that there has been a shift in the underlying population estimate, as opposed to a gradual homing in on a more precise estimate of the same underlying population.

To investigate the possibility of cohort-level confounding, we obtained a version of the summary statistics for PTSD excluding the largest cohort added in Freeze 3: the Million Veterans Program (MVP). We find that the inclusion of MVP is highly unlikely to be a cause for the upward trend in genetic correlation, as PTSD Freeze 3 excluding MVP was estimated to have a comparable genetic correlation to MD Freeze 3 (r_g = 1.02, *SE* = .04). We also examined the relationship between PTSD Freeze 3 and summary statistics that examined GWAS estimates for MD for individuals with and without self-reported trauma exposure in the UK Biobank[1]. These results revealed a similar level of genetic overlap across PTSD Freeze 3 and both MD with (r_g = .72, *SE* = .07) and without trauma exposure (r_g = .70, *SE* = .06). This suggests results are not being driven by MD cohorts, for which cases are more likely to have been exposed to trauma (e.g., veterans-based cohorts). We then examined results when pruning the MD Freeze 3 summary statistics with heterogeneity I^2 statistic *p*-values < 1×10^{-5}. These pruned MD summary statistics were also estimated to have a .99 r_g with PTSD. However, it is important to note that LAVA estimated an average local r_g of .6, with several loci that are uniquely associated with only one of the two disorders (**Suppl. Figs. 3-4**).

It is unclear, based on these results, what is driving the general increase in genetic correlation across PTSD and MD across freezes. Given decreasing SNP-based heritability estimates across the freezes for both disorders, it may be that with increasing numbers of contributing cohorts to the univariate GWAS meta-analysis that the psychiatric signal is being reduced to more transdiagnostic components (**Suppl. Fig. 2**). As the explicit purpose of CDG3 was to triangulate across methods, we considered the framework of this paper an opportunity to identify what is shared and unique across even the highly overlapping freezes of these two disorders. It will be of critical import for future work to further disentangle what may be driving this estimate with respect to particular cohorts, ascertainment strategies, and shared biology.

Univariate MiXeR

We applied univariate MiXeR[2] to estimate polygenicity, measuring the effective number of causal variants for each phenotype. Formally, polygenicity was defined as the number of causal variants that are estimated to account for 90% of a trait's SNP heritability. The estimates of a trait's polygenicity and discoverability allow one to project the sample size a future GWAS study would need to saturate the yield of genome-wide significant loci (i.e., where genome-wide significant loci are anticipated explain 90% of the trait's heritability). Such "power curves" are shown on **Suppl. Fig. 5**. The results are presented in Suppl. Table 5, with SCZ, CUD and TS were among the most "discoverable", with around 12M subjects needed to saturate the GWAS yield. MD, ANX and PTSD were the least discoverable, with estimated 80M subjects needed to saturate the GWAS yield. We note that in terms of polygenic prediction saturation may happen at a lower N, as in practice the best performing polygenic risk scores are not constrained to genome-wide significant variants. Also, more advanced PRS methods incorporating better priors, continuous shrinkage, and leveraging LD from cross-ancestry GWAS are likely to bring major improvements, and therefore much smaller Ns may be sufficient to reach saturation.

High discoverability in SCZ and TS can be largely attributed to their substantially high SNP heritability, with MiXeR's estimate approaching 50% on the observed scale. For comparison, observed-scale SNP heritability for ANX, PTSD and MD was between 4% and 8%. Most other disorders were in the "middle" range, with heritability between 20% and 30%.

For polygenicity, SCZ and BIP have similar polygenicity (12k and 11k variants, respectively), while MD has the highest polygenicity and ADHD the lowest polygenicity. It is worth stressing that, while there is some concordance between heritability and polygenicity estimates, they are not perfectly correlated. For example, MD has higher polygenicity than SCZ, but SCZ has higher SNP-based heritability.

Case Ascertainment Sensitivity Analyses

The impact of shallow versus deep ascertainment methods for psychiatric cases is an increasing topic of concern and discussion[3]. In support of this concern, ascertainment stratified analyses for major depression in the UK Biobank (UKB) sample have found that cases derived using shallow case definitions (e.g., self-report of prior help seeking for nerves, anxiety, or depression) capture broadly pleiotropic pathways that lack depression specific signal[4]. This was as compared to more deeply phenotyped cases from UKB, defined using self-report of complete diagnostic symptoms, that showed higher genetic overlap with MD defined in outside samples by a (semi-)structured clinical interview delivered by a trained rater. Recent findings for other disorders reveal a similar pattern of results, where the most recent PGC GWAS for BIP[6], MD[7], PTSD[8], ANX[9], and OCD[10] show that results stratified by either clinical, community, EHR, or biobank samples all evince high levels of genetic overlap, with minimal evidence for ascertainment-specific genetic signal. This was with the one exception that self-report of diagnosis (typically defined using the 23andMe sample) often showed lower genetic overlap with clinical and community samples. We note that the results in the main text of the current CDG3 project were generated using the univariate GWAS summary statistics that excluded 23andMe for all psychiatric disorders, which should already significantly reduce genetic heterogeneity due to ascertainment.

We investigated whether restricting analyses to even more strictly ascertained cases affects the conclusions from the correlated factors and hierarchical *p*-factor model at both the genome-wide and locus level of analyses. This involved making the following two exclusions for these disorders. First, we removed cases defined by self-report of diagnosis or short format questionnaires (1-4 questions) that do not assess complete diagnostic criteria. In addition, we excluded cohorts that were initially ascertained for another disorder, as this may upwardly bias genetic correlations with that same disorder. For SCZ, NIC, AN, ASD and TS this did not result in any change to the GWAS datasets. For CUD, power would be insufficient following the second exclusion criteria to include it in the model and results should be interpreted with this in mind. For the remaining six disorders the following exclusions were made.

PTSD. For PTSD, we excluded self-report cohorts that did not administer the clinician administered PTSD scale (CAPS) that covers full diagnostic criteria. In addition, we exclude cohorts initially ascertained for other disorders (e.g., iPSYCH). This resulted in a 27.27% reduction in effective sample size (Supplementary Table 7 for case/control numbers).

MD. For MD, we excluded the self-report questionnaire cohorts that included cohorts with self-report of prior diagnosis (e.g., HELIUS). This resulted in a 16.65% reduction in effective sample size.

ANX. For ANX, we excluded self-report of prior diagnosis cohorts (e.g., UKB) along with cohorts initially ascertained for another disorder (what is referred to as the comorbidity cohorts in the ANX study that included the QIMR and Utah Suicide cohorts). This resulted in a 47.56% reduction in effective sample size.

ADHD. The ADHD analyses excluded the Yale-Penn cohort as this was initially ascertained for substance use disorders. Analyses were otherwise left unchanged, as the 11 remaining cohorts all met

the inclusion criteria, with ADHD cases diagnosed by either semi-structured clinical interviews or psychiatrists. The exclusion of the Yale-Penn cohort resulted in an 0.62% reduction in the effective sample size.

OCD. For OCD we excluded only the iPSYCH cohorts as this was initially ascertained for five other psychiatric diagnoses. This resulted in a 18.71% reduction in effective sample size.

BIP. From bipolar we exclude only the UKB cohort as this reflects self-report of prior diagnosis. This resulted in a 5.56% reduction in effective sample size.

Sensitivity Analyses: LDSC and Genomic SEM. The genetic correlation matrix in the sensitivity sample continued to show high levels of genetic overlap across the 14 disorders, along with notable clustering among subsets of disorders (Suppl. Fig. 11; Suppl. Table 8). On average, there were minimal differences in the genetic correlation matrix between the full sample and sensitivity sample LDSC matrix, with genetic correlations often being slightly larger in the full LDSC sample: for genetic correlations where at least one disorder was updated for the sensitivity analyses, the mean difference between the sensitivity and full sample was -.03 (range: -.18: .06; SD= .05).

We went on to perform the same combination of exploratory factor analyses (EFAs) fit to the LDSC correlation matrix estimated in odd autosomes that were used to guide confirmatory factor analysis (CFAs) using LDSC estimated for even autosomes. As in the full sample, we examined the fit for a common factor model and for one to five factor model solutions identified using varimax or promax rotations at the EFA stage (see main text Method for additional details; Suppl. Table 9 for model fit in sensitivity sample). As in the full sample, the best fitting model derived from EFAs performed in the sensitivity sample was a five-factor correlated factors model (Comparative Fix Index [CFI] = .926, Standard Root Mean Square Residual [SRMR] = .088). This model also provided good fit when applied to the LDSC matrix for all autosomes (CFI = .907, SRMR = .086; Suppl. Table 10 for model output). The five factors identified using the EFA in the sensitivity analysis were defined by a similar subset of disorders relative to those identified using the EFA in the full sample. For example, PTSD, MD, and ANX (three of the disorders with the largest change in effective sample size across the full and sensitivity samples) continued to cluster together onto an internalizing factor. The largest shift was for the third factor, which was defined in the sensitivity sample by ASD and NIC, but was defined by ASD, ADHD, and TS in the full sample.

Given the similarity in factor structure identified using the EFA conducted in the sensitivity sample, we also examined whether the factor structure identified in the full sample also provided good fit in the sensitivity sample. We find that the five-factor correlated factors model identified in the full sample continued to provide the best fit to the data in the sensitivity sample when applied to both odd autosomes (CFI = .945, SRMR = .074) and all autosomes (CFI = .953, SRMR = .064). The point estimates for the five-factor correlated factors model applied to the sensitivity data were also highly similar to estimates from the full sample (Suppl. Fig. 12 for path diagram with comparisons of estimates). The hierarchical *p*-factor model also provided good fit to the sensitivity data in all autosomes (CFI = .937, SRMR = .074), and once again the loadings of the five psychiatric factors on the p-factor were highly similar in the full and sensitivity sample (Suppl. Fig. 12). The correlated factors model and hierarchical models identified in the full dataset were used to estimate multivariate GWAS in the sensitivity dataset as they both provided better fit to the sensitivity data and offer a more direct form of comparison to the full sample results.

Sensitivity Analyses: Multivariate GWAS. Multivariate GWAS results from the correlated factors and hierarchical models in the sensitivity dataset largely mirrored results from the full dataset. This was indicated, in part, by inspection of Miami and QQ-plots that directly compare the sensitivity and full GWAS factor and Q_{SNP} results; these results generally reveal similar overall patterns (Suppl. Figs. 13-18). In addition, sensitivity GWAS hits were largely subsumed by the original GWAS hits for the factors (Suppl. Table 11). We focus here on GWAS results for the Internalizing factor, which was defined by the three disorders with the largest reduction in effective sample size in the sensitivity sample (PTSD, MD, and

ANX). Among the 69 Internalizing factor hits from the sensitivity GWAS, 98.5% were also hits in the original GWAS. Among the 150 Internalizing hits from the original GWAS, 100 hits were not captured by the sensitivity GWAS; however, the average *p*-value among those 100 hits in the sensitivity GWAS was trending significant (average p = 2.37E-5) and still far lower than the average p-value across all SNPs in the sensitivity GWAS (average p = .422). This is visually depicted in Suppl. Fig. 19, which shows that the distribution of p-values among these 100 SNPs is in the outer right-tail of a distribution of -log10(*p*)-values. We take this to indicate that the greater number of hits from the original GWAS reflects increased power from the added participant samples in the original univariate GWAS. This as opposed to an alternative explanation that attributes these 100 hits to reflect signal driven by particular forms of case ascertainment being included in the univariate GWAS. For the QSNP heterogeneity statistic, we identify 17 hits in the original GWAS and no hits in the sensitivity GWAS for this same Internalizing factor. In this instance, these 17 SNPs were generally no more significant for QSNP (average p = .348) than the rest of the Q_{SNP} p-values for the sensitivity GWAS (average p = .510; Suppl. Fig. 19). This suggests that these SNPs may be picking up on heterogeneity introduced by different forms of case ascertainment being folded into the univariate GWAS of the psychiatric disorders. As Q_{SNP} is used as a quality control on the factor-level results, factor hits presented in the main text (and downstream analyses that use these hits as input) are not affected by this potential heterogeneity.

Sensitivity Analyses: Summary. At the genome-wide level, genetic correlations and the genomic factor structure in Genomic SEM (including the size of the factor loadings) were largely unchanged when restricting the psychiatric cases to those that were more strictly ascertained. At the locus level, the signal was slightly attenuated for the sensitivity sample, as would be expected given the reduction in sample size and corresponding decrease in statistical power. However, the overall trend in variant level findings was highly concordant across the original and sensitivity GWAS. The concordance in results across the sensitivity and original sample here is in line with the prior results from univariate GWAS of psychiatric disorders noted above that indicate concordance in results across most forms of ascertainment, coupled with the fact that we do not use the short format, self-report 23andMe sample found to be the most divergent in many of these prior analyses[6,7,9,10]. While future analyses should continue to examine the effects of case ascertainment, the current findings indicate that including the full univariate GWAS samples offers increased statistical power with minimal bias introduced by including other forms of ascertainment. To facilitate ascertainment focused questions in future work, we make available the full sensitivity GWAS summary statistics for the factors (see **Data Availability** statement in main text).

PheWAS in Mayo Clinic Participant Sample

Phecodes were ascertained using EHR data from 57,001 patients from the Mayo Clinic Biobank. EHR data for the participants was extracted on September 23, 2022, and included any diagnostics on or before April 6, 2020, the date patient consent was checked. The Institutional Review Board of Mayo Clinic approved this study. All samples were prepared for sequencing using a custom automated sample preparation workflow developed at the Regeneron Genetics Center (RGC). The RGC sequencing assay is a heavily modified version of the TWIST Comprehensive Exome custom capture. The custom design includes additional augmentation of the exome capture with "backbone" regions intended to capture common tagging variation for purposes of genome-wide association studies (GWAS). Specifically, these additional backbone regions are targeted at lower depth, which then undergo substantial post-processing using proprietary algorithms that can boost genotyping quality based on shared information via linkage disequilibrium and population allele frequencies. This is referred to by RGC as genotyping-by-sequencing (G×S).

The resulting G×S data was run through the Mayo Clinic Genotype QC pipeline. In this QC pipeline, SNPs were excluded using filters for call rate (<95%), minor allele frequency (<0.5%), and deviation from Hardy-Weinberg Equilibrium ($p < 1\times10^{-6}$). Individuals were excluded for excessive missing

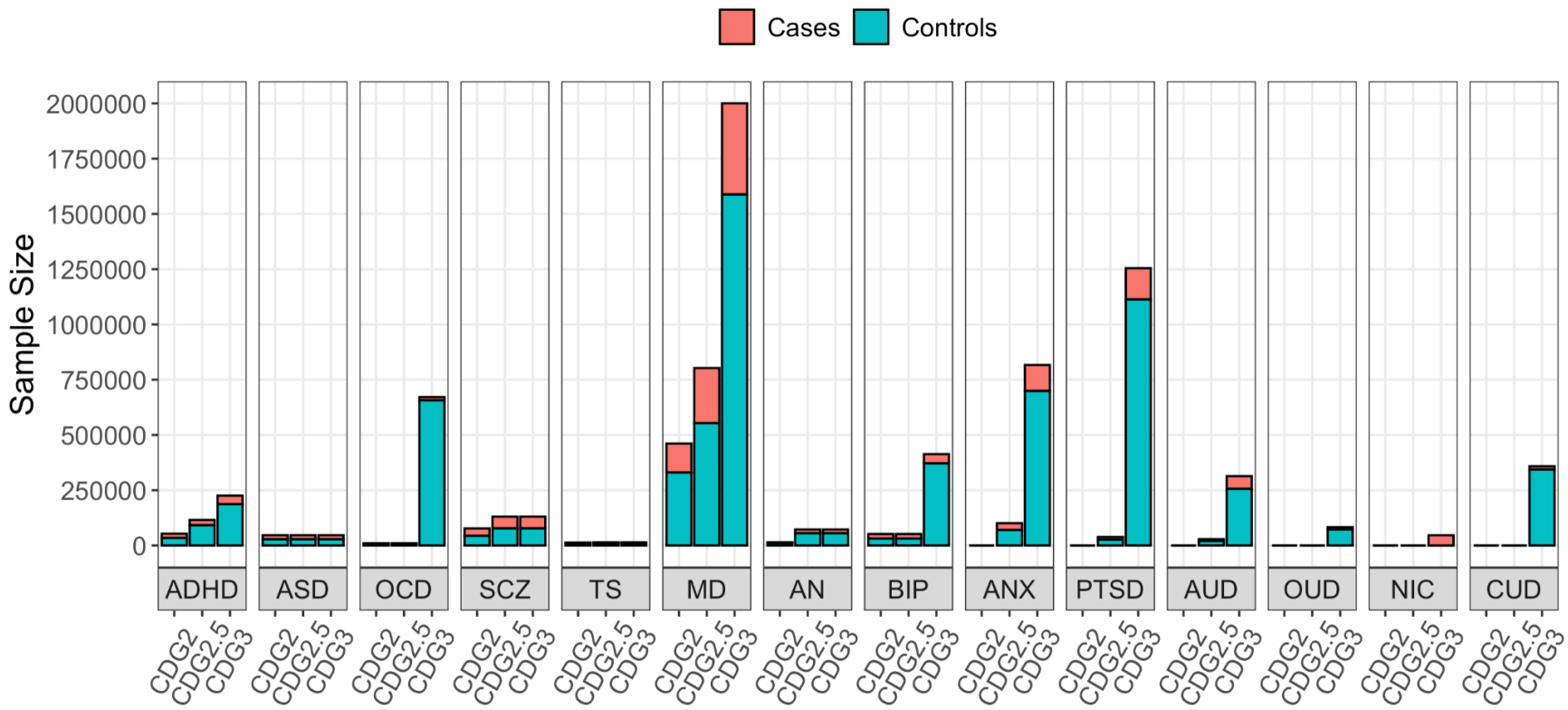
data (>5%), sex errors, and abnormal heterozygosity (< 70% on multiple chromosomes). An analysis of genetic ancestry was performed on a subset of 4874 high-quality HapMap3 SNPs. Principal components analysis (PCA) was first performed on the HapMap3 samples, and then MCB samples were projected onto these PCs, with the leading 20 PCs reported. RGC applied a kernel density estimator (KDE) approach to ancestry assignment based on the PCA output. KDEs were trained for each of the individual HapMap3 populations, which were then mapped to one of five main genetic ancestral super-populations (AFR-like=African-like; AMR-like=Admixed American-like; EAS-like=East Asian-like; EUR-like=European-like; SAS-like=South Asian-like) based on specific likelihood criteria for the individual populations such that the likelihood for a given ancestry group was greater than 0.3, the sample was assigned to that ancestry group. When two ancestry groups had a likelihood of greater than 0.3, RGC arbitrarily assigned AFR-like over EUR-like, AMR-like over EUR-like, AMR-like over EAS-like, SAS-like over EUR-like, and AMR-like over AFR-like. We restricted all analyses to include only those determined to be of EUR ancestries.

Cryptic relatedness analysis was performed in multiple steps using PLINK[11] and PRIMUS[12]. IBD sharing estimates were calculated among all individuals with the same ancestral superclass assignment based on a $\hat{\pi}$ threshold of 0.1875 to capture up to second-degree relationships. A second set of IBD estimates was calculated among all samples based on a $\hat{\pi}$ cutoff of 0.3 to capture 1st degree relationships that may span ancestral superclasses. Samples were then grouped into first-degree family networks based on the IBD estimates, which were processed using PRIMUS under default settings. Finally, samples with cryptic relatedness were removed from the analysis if they had >100 closely related samples ($\hat{\pi}$>0.1875) or >25000 related samples ($\hat{\pi}$>0.08), which was performed iteratively until no such samples remained. For each pair with an estimated 2nd degree or higher relatedness, we removed the individual with shorter length of EHR record. An additional 76 participants were excluded due to overlap with the univariate disorder GWAS for either AUD or BD used in this study.

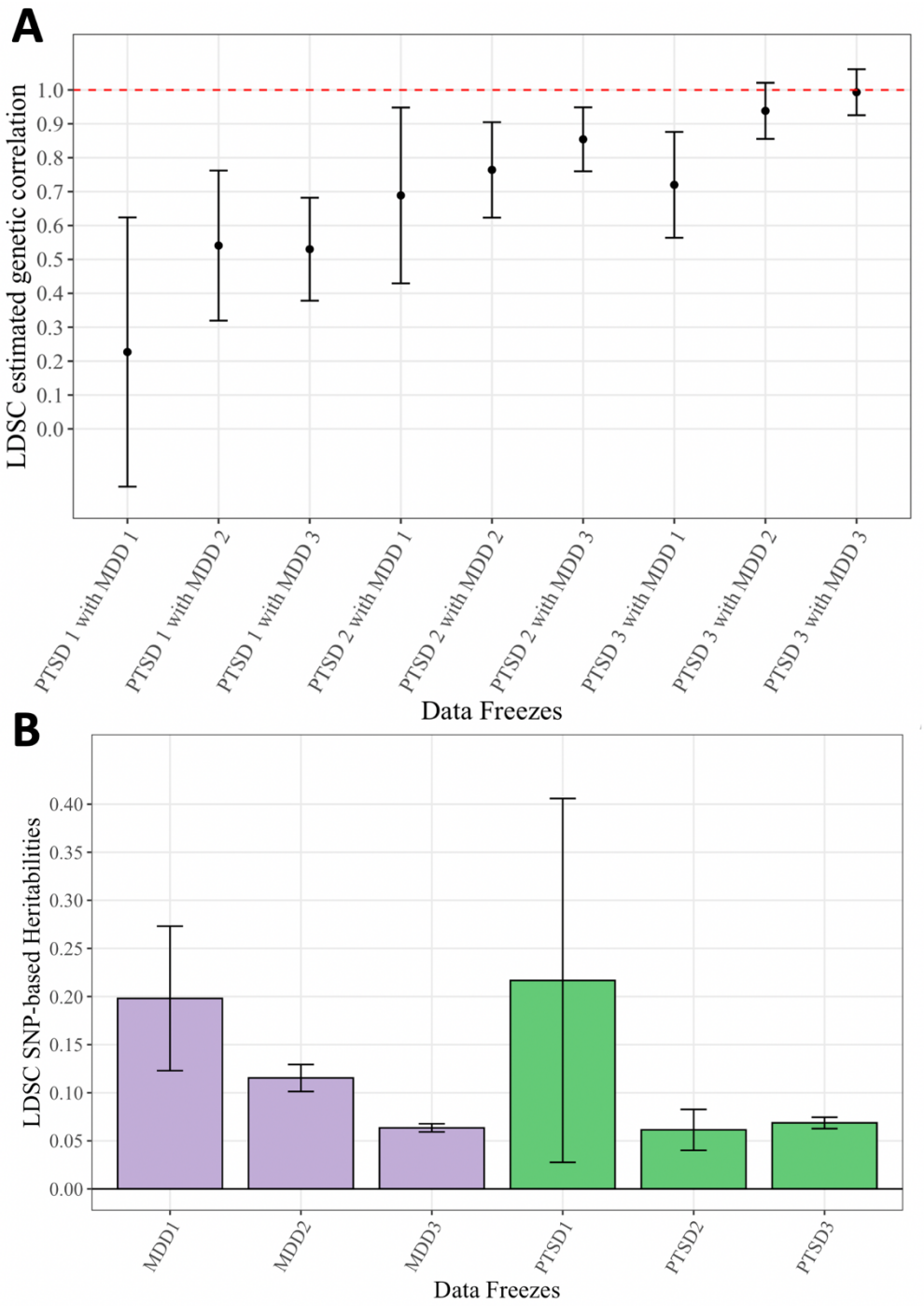
PRSs were then estimated within this quality-controlled sample using LDpred2-auto[13] along with the full set (i.e., no *p*-value thresholding) of multivariate GWAS summary statistics for the five factors from the correlated factors model and the transdiagnostic *p*-factor from the hierarchical model. PheWAS Results are reported in **Suppl. Table 38** and displayed for each factor in **Suppl. Fig. 38.** Significance was defined using a Bonferroni corrected threshold of $p < 4.7\times10^{-6}$, reflecting the standard threshold of $p < .05$ corrected for 10,626 statistical tests. This reflects the product of the number of examined phecodes (1,776) and psychiatric factors (6).

MAGMA Enrichment

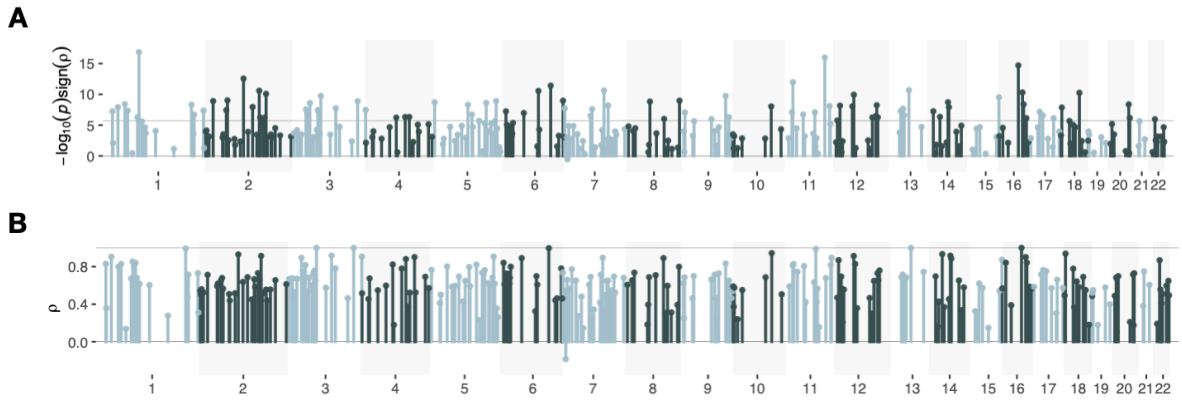
To examine the overlap between our GWAS signal and known disease-causing rare variations, we computed gene-level rare variant heritability using MAGMA[14]. First, we obtained and harmonized candidate gene lists from contemporary rare variant studies spanning autism spectrum disorder, the broader class of neurodevelopmental disorders, and SCZ.[15–20] Next, a gene set annotation was generated by pairing each source study with the genes passing significance in that study. This two-column file was fed into MAGMA as a set annotation, with the '*--set-annot gene-col=1 set-col=2'* flag, to be compared against gene-level results for the psychiatric factors. To generate continuous annotations, we divided the negative log matrix of gene and study *p*-values into 40 bins, whereby 40 represents a high degree of association between the gene and study and 0 a low one. Studies without *p*-values were assigned a value of 40 for each gene in the study's curated gene list, and 0 otherwise. This binned gene by study matrix was fed into MAGMA as a covariate file, with the '*--gene-covar*' and '*--model direcn=pos*' flags. Resulting enrichment statistics were Bonferroni corrected for multiple comparisons.



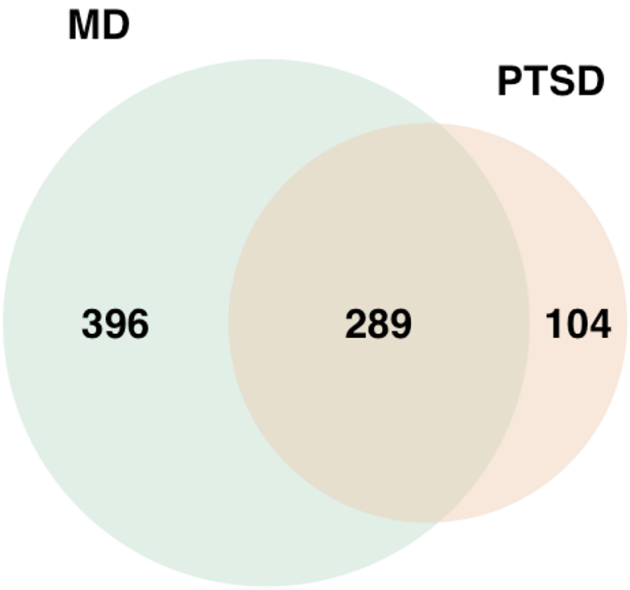
Supplementary Figure 1. Participant Sample Sizes for CDG2, CDG2.5, and CDG3. Figure displays the total sample sizes in the CDG2, what we term the CDG2.5 analyses from Grotzinger et al. 2022[21], and the current CDG3 analyses for the 14 major psychiatric disorders. Cases and controls are stacked on top of one another, with controls shown in blue and cases in red. Note that the right-most, six disorders on the x-axis are entirely new relative to the CDG2 analyses and the right-most three disorders entirely new relative to CDG2.5. ADHD = attention-deficit hyperactivity disorder; ASD = autism spectrum disorder; OCD = obsessive compulsive disorder; SCZ = schizophrenia; TS = Tourette's syndrome; MD = major depression AN = anorexia nervosa; BIP = bipolar disorder; ANX = anxiety disorder; PTSD = post-traumatic stress disorder; AUD = alcohol use disorder; OUD = opioid use disorder; NIC = nicotine dependence; CUD = cannabis use disorder



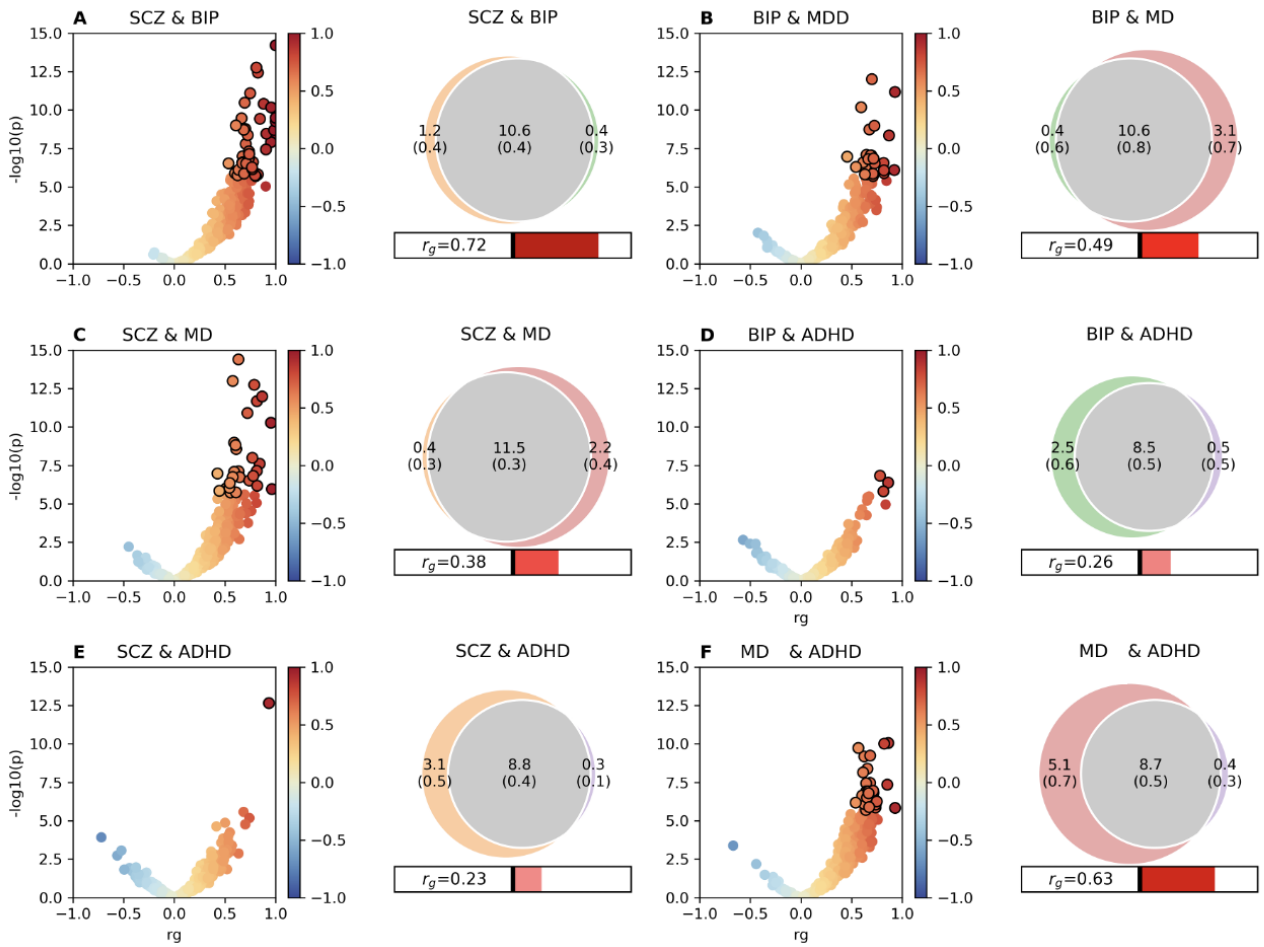
Supplementary Figure 2. Genetic Correlations and Heritabilities across Data Freezes for MD and PTSD. *Panel A* depicts the LDSC estimated genetic correlations across data freezes for MD and PTSD. Freezes are depicted sequentially, and as can be readily observed, the genetic correlation across this pair of disorders has been gradually increasing across freezes. *Panel B* depicts the SNP-based heritability estimates from LDSC for these same PTSD and MD data freezes. Error bars depict 95% confidence intervals across both panels that are centered around the heritability (Panel A) and genetic correlation (Panel B) point estimates.



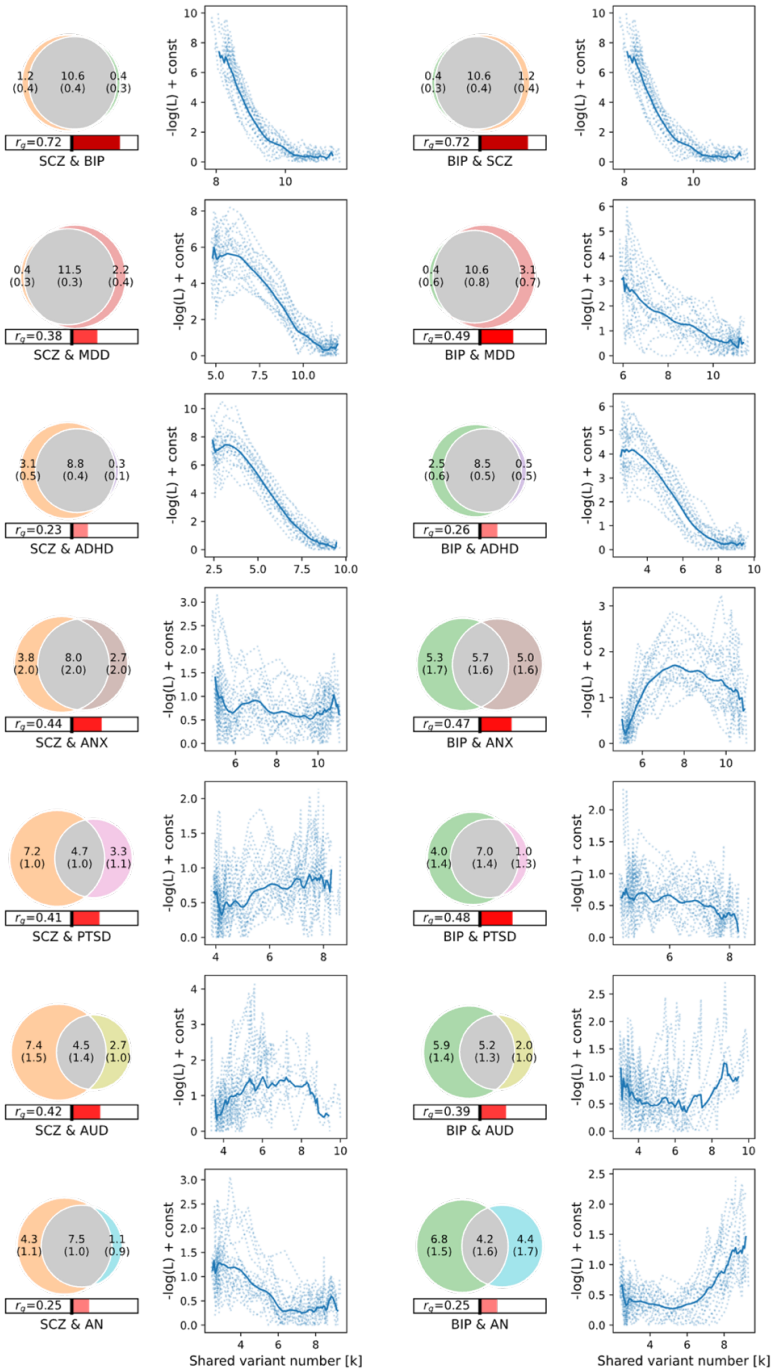
Supplementary Figure 3. Local genetic correlation patterns for MD and PTSD. Panel A shows the -log10 *p*-values for the local genetic correlations scaled by the sign in all evaluated loci (the line indicating the Bonferroni corrected significance threshold of 2.1×10^{-6}), while Panel B shows the corresponding local genetic correlation estimate across all evaluated loci (the line indicating an estimate of 1). Empirical p-values were obtained via a permutation procedure with partial integration, evaluating the two-sided hypothesis of no association using the estimated parameters as test statistics.



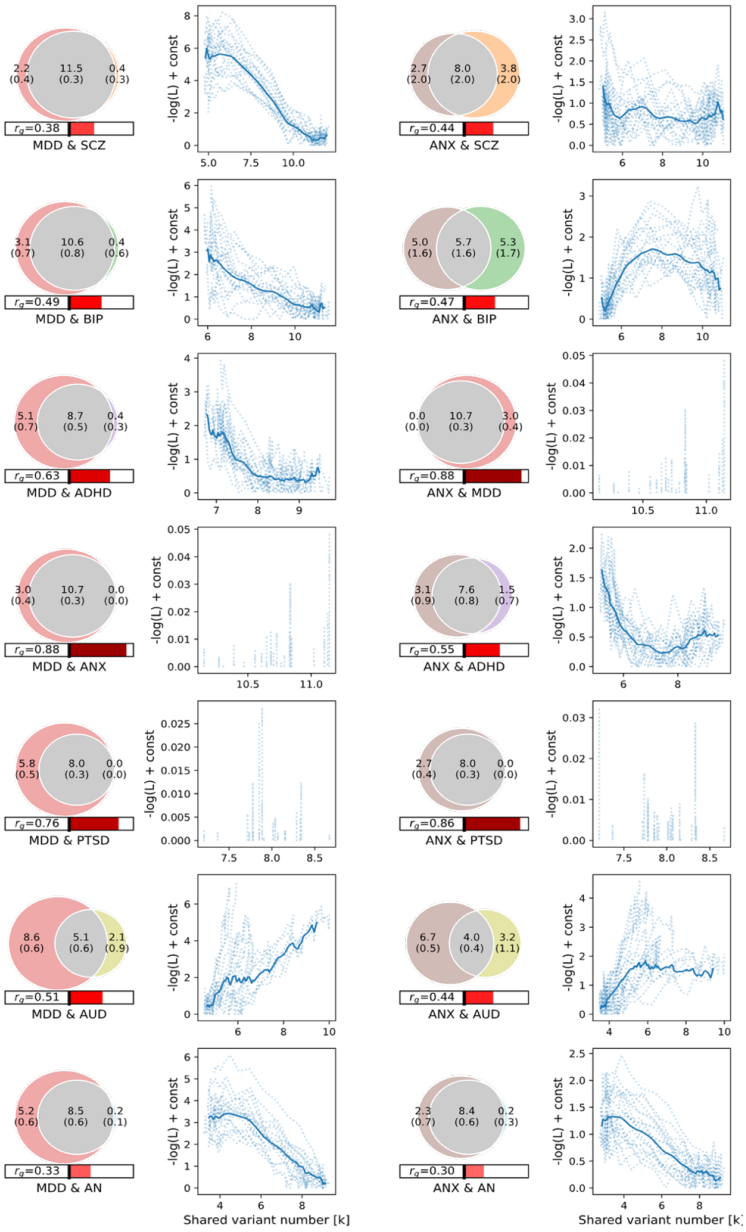
Supplementary Figure 4. LAVA PTSD and MD Venn Diagram. Venn diagram illustrating the number of unique and overlapping loci with univariate genetic signal detected at $p < .05 / 1{,}093$ for MD and PTSD.



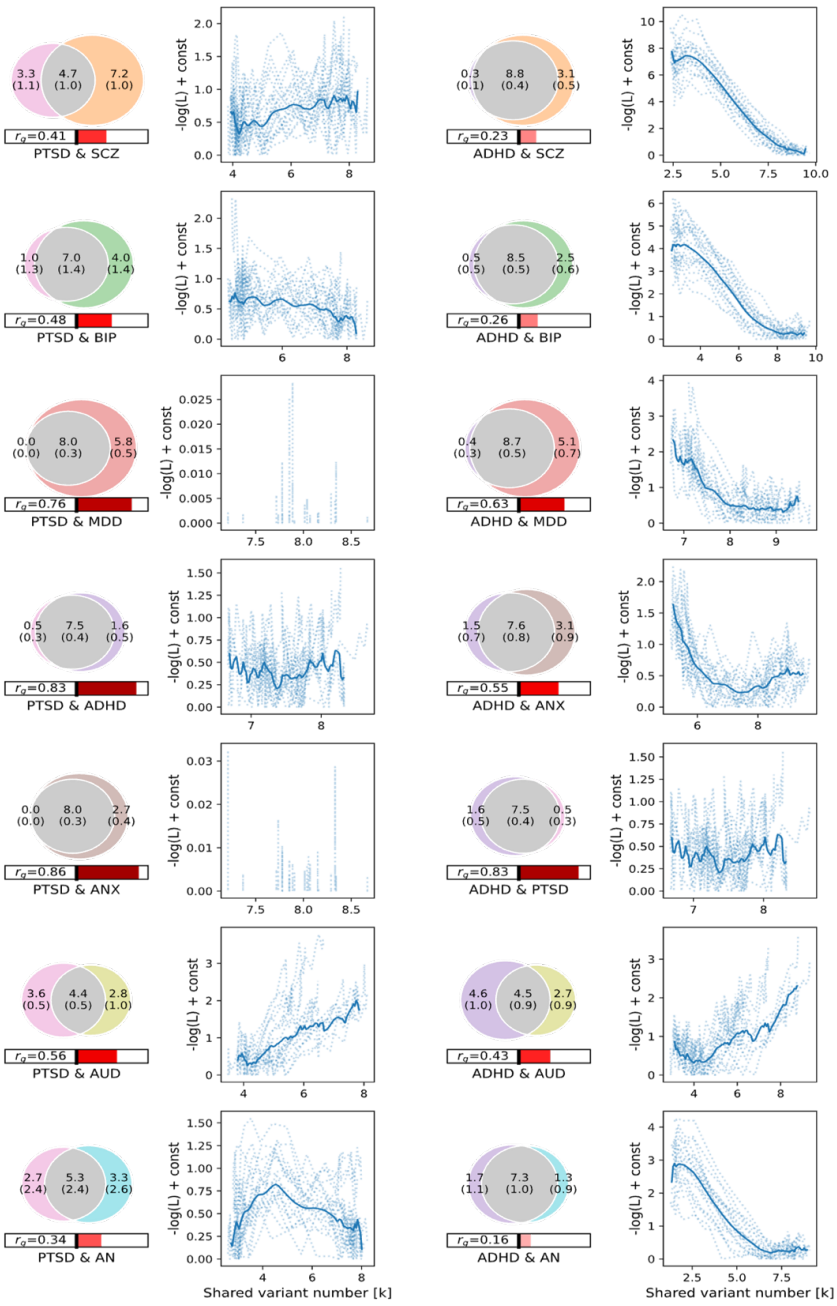
Supplementary Figure 5. LAVA and MiXeR results for select disorders. Results are shown for LAVA (volcano plots) and MiXeR (Venn diagrams) for four of the most well-powered disorders: schizophrenia (SCZ), bipolar disorder (BIP), major depression (MD), and attention-deficit/hyperactivity disorder (ADHD). LAVA results are shown as a volcano plot of the local genetic correlations (r_g) on the x-axis and -log10(p) values on the y-axis. Stronger negative r_g's are depicted by darker shades of blue, while stronger positive r_g's are shown in darker shades of red. The points for local r_g's that were significant at a Bonferroni corrected threshold are depicted with black outlines. MiXeR results are shown as adjacent Venn diagrams that convey unique and shared polygenic components at the causal level. The numbers within the Venn diagrams indicate the estimated quantity of causal variants (in thousands) per component, explaining 90% of SNP heritability in each phenotype, followed by the standard error. The size of the circles reflects the degree of polygenicity. r_g is represented in the horizontal bars beneath the Venn diagrams. Sample sizes for the disorders are reported in Extended Data Table 1.



Supplementary Figure 6. Cross-trait MiXeR results for Schizophrenia and Bipolar disorder.
Venn diagrams of unique and shared polygenic components at the causal level for SCZ (left column) and BIP (right column). The numbers within the Venn diagrams indicate the estimated quantity of causal variants (in thousands) per component, explaining 90% of SNP heritability in each phenotype, followed by the standard error. The size of the circles reflects the degree of polygenicity. Genetic correlation (r_g) is represented in the horizontal bars beneath the Venn diagrams. Right of the central bar (red) indicates positive r_g and left of the central bar (blue) indicates negative r_g. Next to each Venn diagram is a plot of the log-likelihood of the bivariate fit as a function of π12 (shared variant) parameter. These curves indicate the strength of the evidence of polygenic overlap as compared to minimal overlap. Positive AIC values (**Suppl. Table 5**) and a clear minima on the likelihood curves provide evidence of good model fit. Sample sizes for the disorders are reported in Extended Data Table 1.

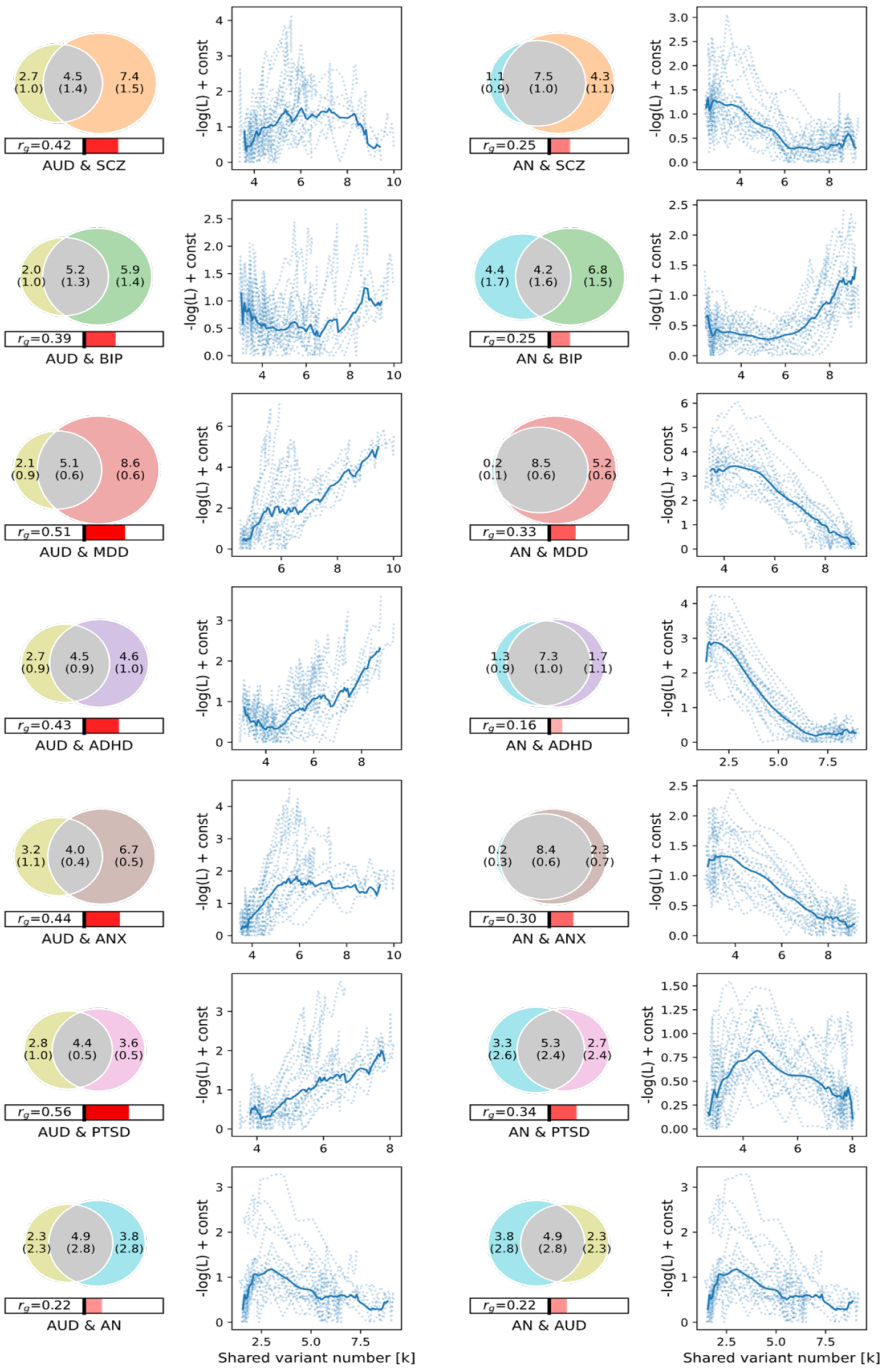


Supplementary Figure 7. Cross-trait MiXeR results for Major Depressive Disorder and Anxiety Disorders. Venn diagrams of unique and shared polygenic components at the causal level for MD (left column) and ANX (right column). The numbers within the Venn diagrams indicate the estimated quantity of causal variants (in thousands) per component, explaining 90% of SNP heritability in each phenotype, followed by the standard error. The size of the circles reflects the degree of polygenicity. Genetic correlation (r_g) is represented in the horizontal bars beneath the Venn diagrams. Right of the central bar (red) indicates positive r_g and left of the central bar (blue) indicates negative r_g. Next to each Venn diagram is a plot of the log-likelihood of the bivariate fit as a function of $\pi 12$ (shared variant) parameter. These curves indicate the strength of the evidence of polygenic overlap as compared to minimal overlap. Positive AIC values (Suppl. Table 5) and a clear minima on the likelihood curves provide evidence of good model fit. We note that for PTSD, ANX and MD the genetic correlations were high to an extent that there was little room for additional overlap beyond correlation, given MiXeR's modeling assumptions. More specifically, the range in size of the putative shared component is too small to allow for an accurate model fit this situation, as demonstrated by the range on the respective x-axes. Sample sizes for the disorders are reported in Extended Data Table 1.

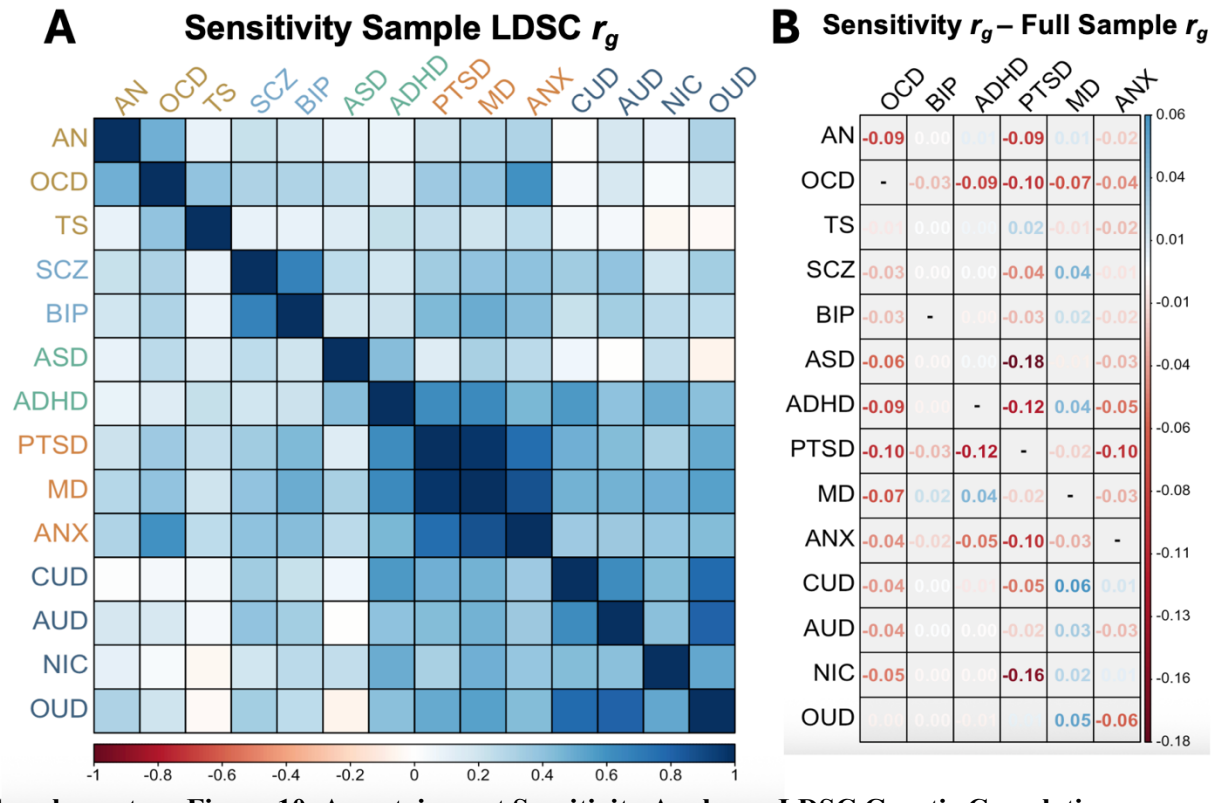


Supplementary Figure 8. Cross-trait MiXeR results for Post-traumatic Stress Disorder and Attention-deficit/hyperactivity Disorder.

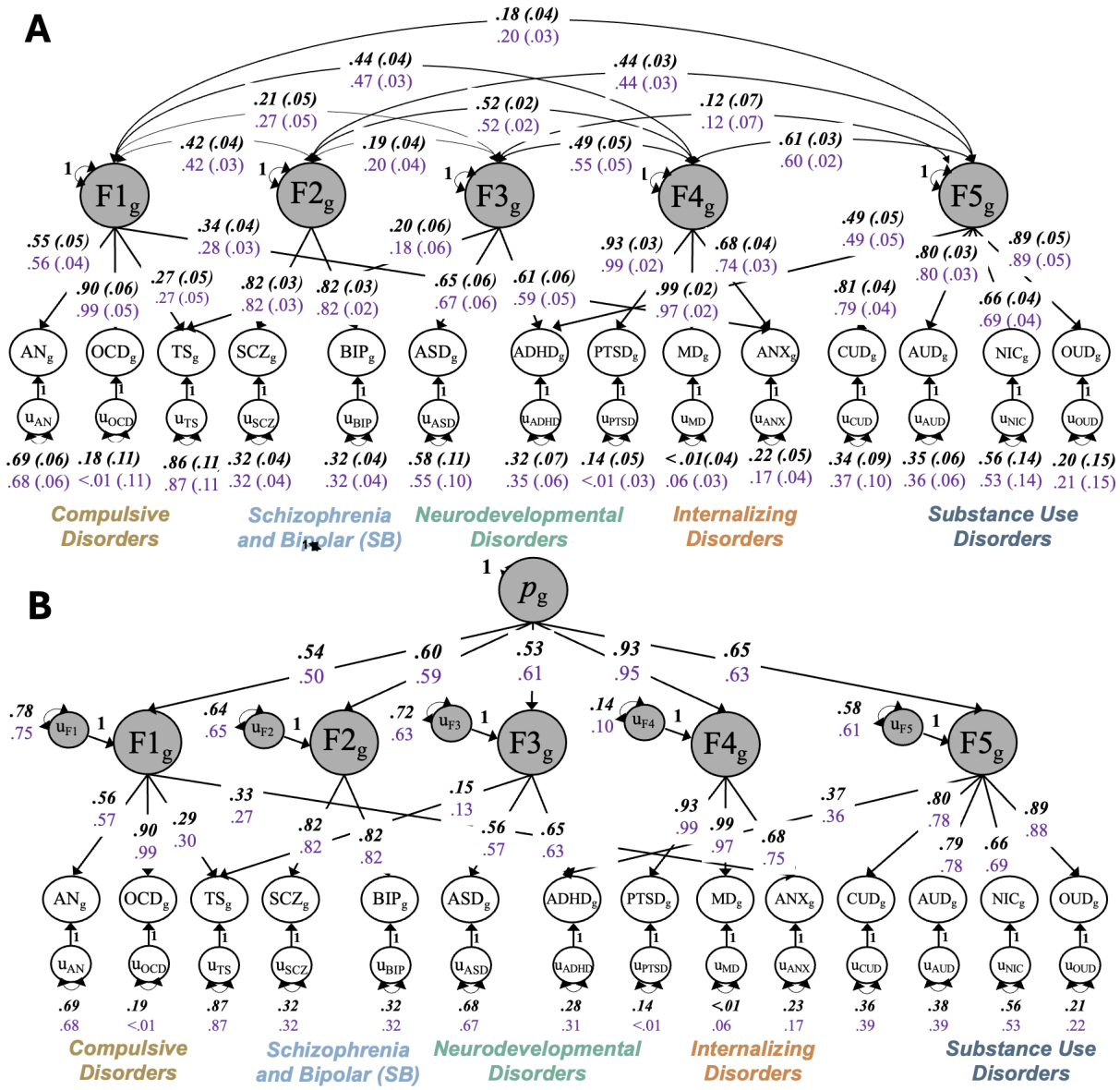
Venn diagrams of unique and shared polygenic components at the causal level for PTSD (left column) and ADHD (right column). The numbers within the Venn diagrams indicate the estimated quantity of causal variants (in thousands) per component, explaining 90% of SNP heritability in each phenotype, followed by the standard error. The size of the circles reflects the degree of polygenicity. Genetic correlation (rg) is represented in the horizontal bars beneath the Venn diagrams. Right of the central bar (red) indicates positive rg and left of the central bar (blue) indicates negative rg. Next to each Venn diagram is a plot of the log-likelihood of the bivariate fit as a function of $\pi 12$ (shared variant) parameter. These curves indicate the strength of the evidence of polygenic overlap as compared to minimal overlap. Positive AIC values (Suppl. Table 5) and a clear minima on the likelihood curves provide evidence of good model fit. We note that for PTSD, ANX and MD the genetic correlations were high to an extent that there was little room for additional overlap beyond correlation, given MiXeR's modeling assumptions. More specifically, the range in size of the putative shared component is too small to allow for an accurate model fit this situation, as demonstrated by the range on the respective x-axes. Sample sizes for the disorders are reported in Extended Data Table 1.



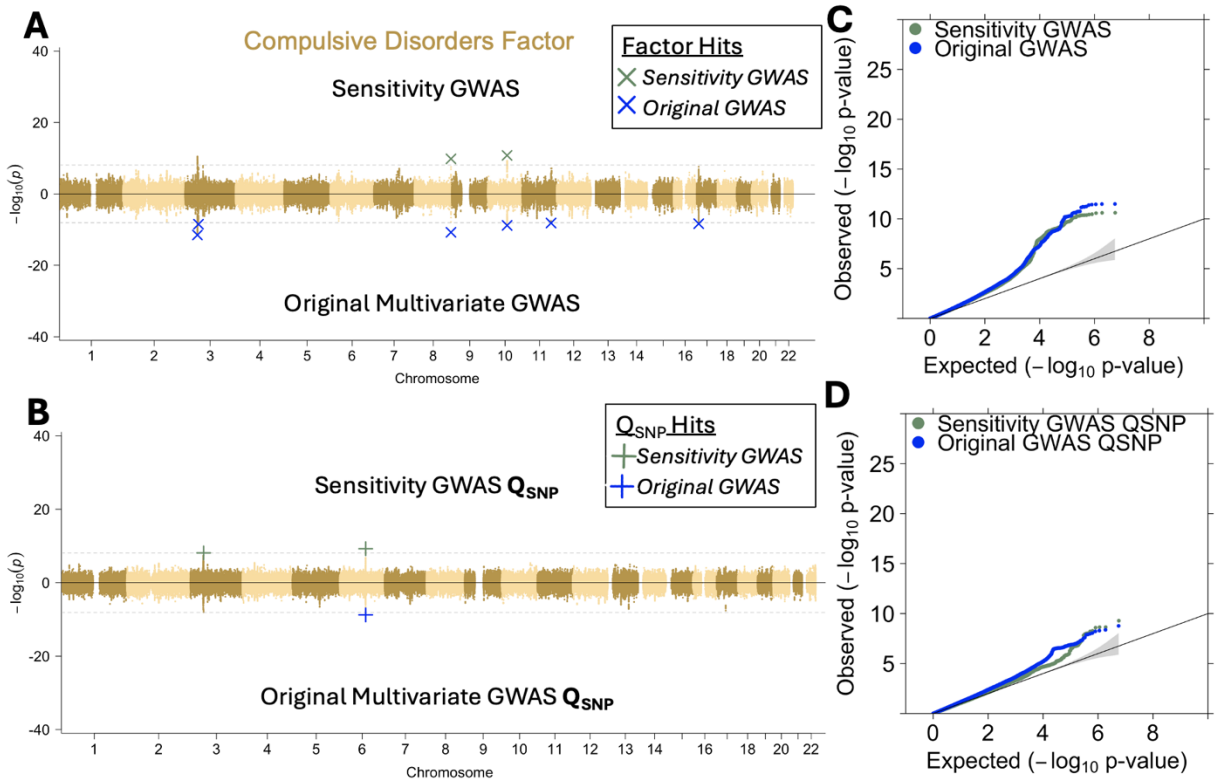
Supplementary Figure 9. Cross-trait MiXeR results for Alcohol Use Disorder and Anorexia Nervosa. Venn diagrams of unique and shared polygenic components at the causal level for AUD (left column) and AN (right column). The numbers within the Venn diagrams indicate the estimated quantity of causal variants (in thousands) per component, explaining 90% of SNP heritability in each phenotype, followed by the standard error. The size of the circles reflects the degree of polygenicity. Genetic correlation (rg) is represented in the horizontal bars beneath the Venn diagrams. Right of the central bar (red) indicates positive rg and left of the central bar (blue) indicates negative rg. Next to each Venn diagram is a plot of the log-likelihood of the bivariate fit as a function of $\pi 12$ (shared variant) parameter. These curves indicate the strength of the evidence of polygenic overlap as compared to minimal overlap. Positive AIC values (Suppl. Table 5) and a clear minima on the likelihood curves provide evidence of good model fit. Sample sizes for the disorders are reported in Extended Data Table 1.



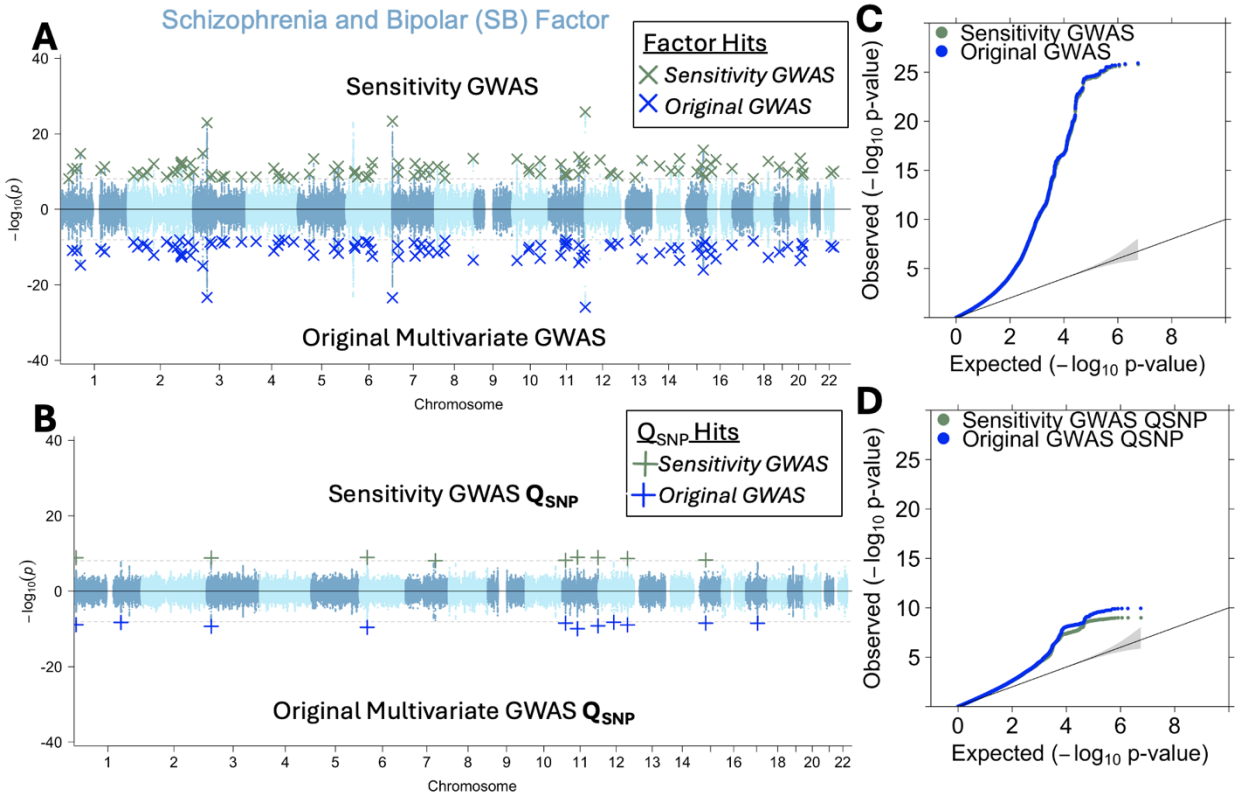
Supplementary Figure 10. Ascertainment Sensitivity Analyses: LDSC Genetic Correlations. Panel A depicts the LDSC estimates from the sensitivity samples that excluded certain cohorts based on ascertainment. Panel B reports the difference in the LDSC genetic correlation point estimates between the sensitivity sample and full sample estimates (i.e., negative numbers indicate larger genetic correlation estimates in the full sample), with the six disorders for which sensitivity data was obtained displayed across the six columns. Sample sizes for the disorders are reported in Extended Data Table 1. ADHD = attention-deficit hyperactivity disorder; ASD = autism spectrum disorder; OCD = obsessive compulsive disorder; SCZ = schizophrenia; TS = Tourette's syndrome; MD = major depression; AN = anorexia nervosa; BIP = bipolar disorder; ANX = anxiety disorder; PTSD = post-traumatic stress disorder; AUD = alcohol use disorder; OUD = opioid use disorder; NIC = nicotine dependence; CUD = cannabis use disorder.



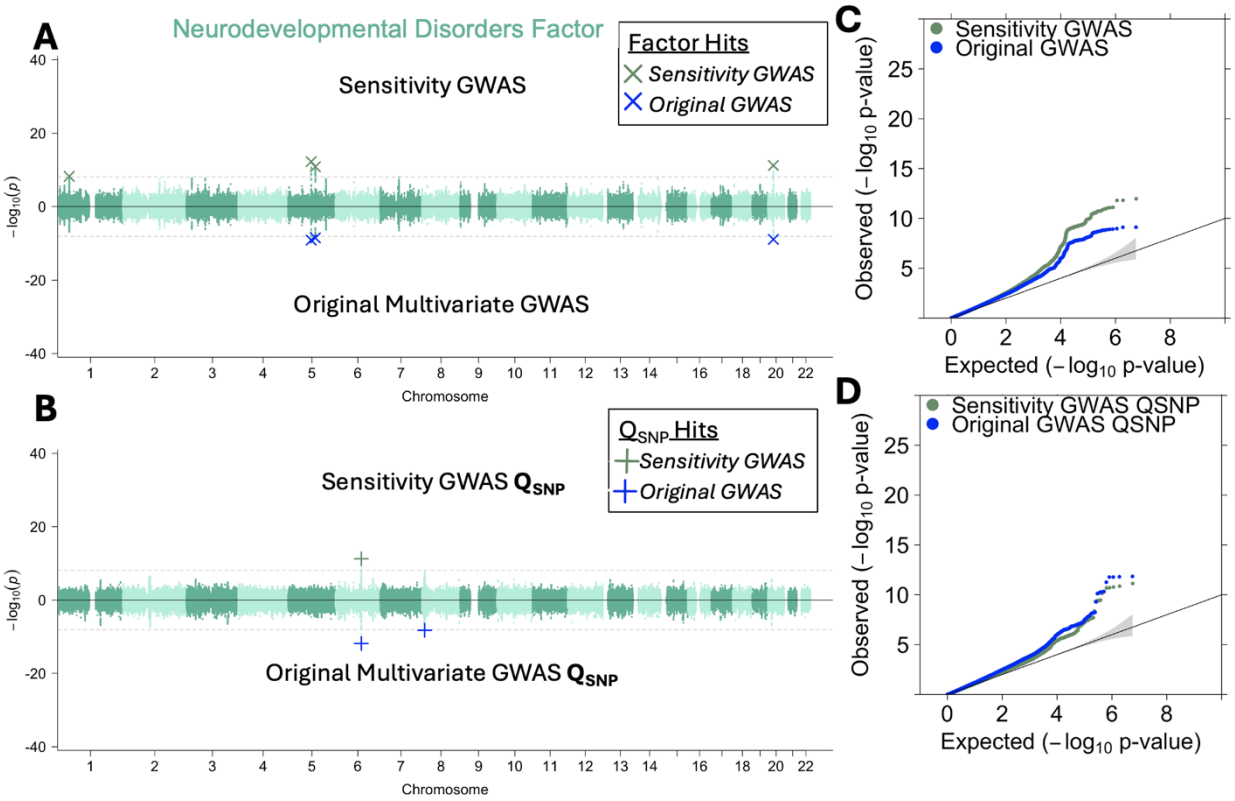
Supplementary Figure 11. Ascertainment Sensitivity Analyses: Factor Model.
Panel A depicts estimates from the five-factor model along with standard errors in parentheses. Across both panels, the lower values in purple reflect the estimates obtained using the full sample (i.e., they mirror what is presented in Figure 1). The upper values in ***black*** **bold italics** reflect the estimates obtained using the exclusions outlined in the ascertainment sensitivity analyses section. Estimates are standardized relative to the SNP-based heritabilities, where the total variance in the disorders is equal to the sum of the squared factor loading (the single-headed arrow(s) from the factor to the disorder) and the residual variance (the values on the double-headed arrows on the single-color circles). Panel B displays the standardized estimates from the p-factor model for both the full sample and the **ascertainment sensitivity sample**. ADHD = attention-deficit hyperactivity disorder; ASD = autism spectrum disorder; OCD = obsessive compulsive disorder; SCZ = schizophrenia; TS = Tourette's syndrome; MD = major depression; AN = anorexia nervosa; BIP = bipolar disorder; ANX = anxiety disorder; PTSD = post-traumatic stress disorder; AUD = alcohol use disorder; OUD = opioid use disorder; NIC = nicotine dependence; CUD = cannabis use disorder.



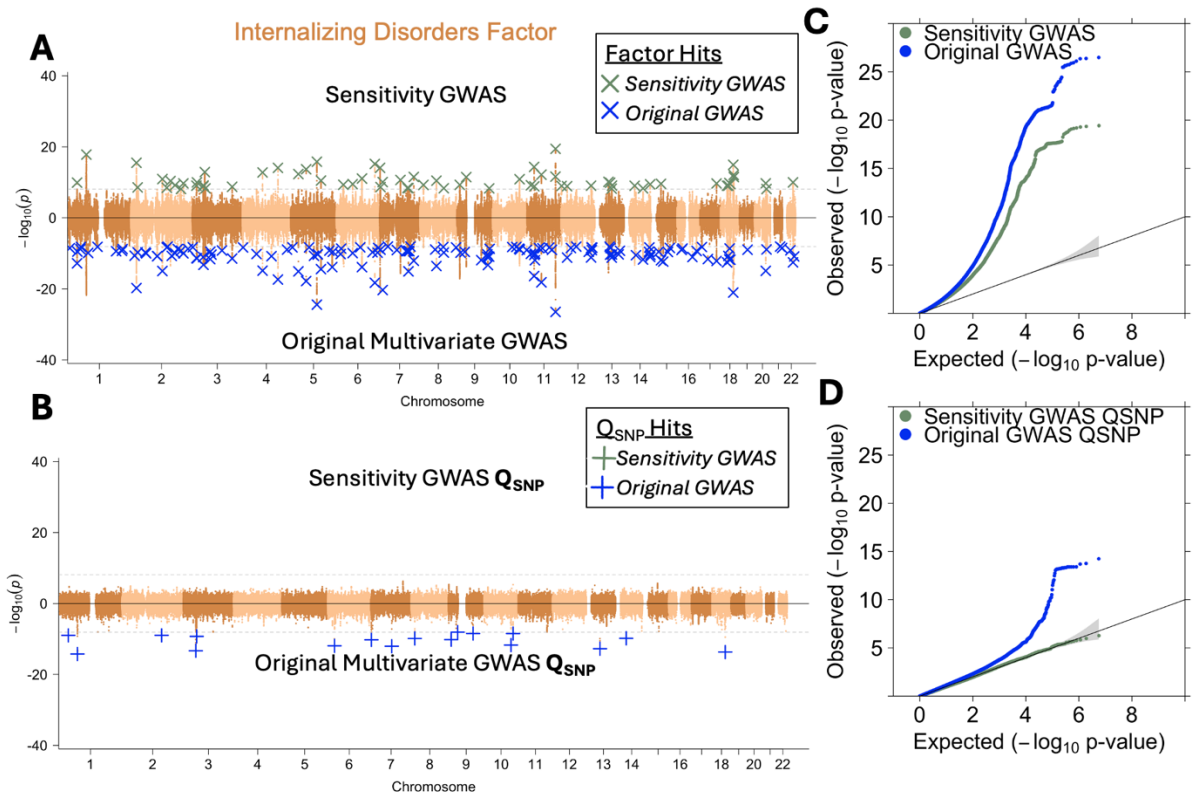
Supplementary Figure 12. Ascertainment Sensitivity Analyses: Compulsive disorders Factor Miami Plots. Panel A depicts multivariate GWAS results for the Compulsive disorders factor as a Miami plot with -log10(p) values for the sensitivity GWAS results on the top half and the log10(p) values of the original GWAS displayed on the bottom half. Hits at a Bonferroni corrected, genome-wide significance threshold that were not within 100kb of a QSNP are depicted with a green X for the sensitive GWAS on the top half and a blue X for the original GWAS on the bottom half. Panel B shows the -log10(p) values for the QSNP results for the sensitivity GWAS and the log(10) p values for the original GWAS. Hits at a Bonferroni corrected, genome-wide significance threshold are depicted with a green + for the sensitive GWAS on the top half and a blue + for the original GWAS on the bottom half. The gray dashed line indicates the Bonferroni corrected, genome-wide significance threshold in both panels. The GWAS *p*-values were two-sided and calculated from Z-statistics (the estimate over the standard error) from the multivariate GWAS. Panel Cs and D depict QQ-plots for the Compulsive disorders factor and QSNP statistic, respectively. In both panels, results for the original GWAS are shown in blue and results from the sensitivity GWAS in green.



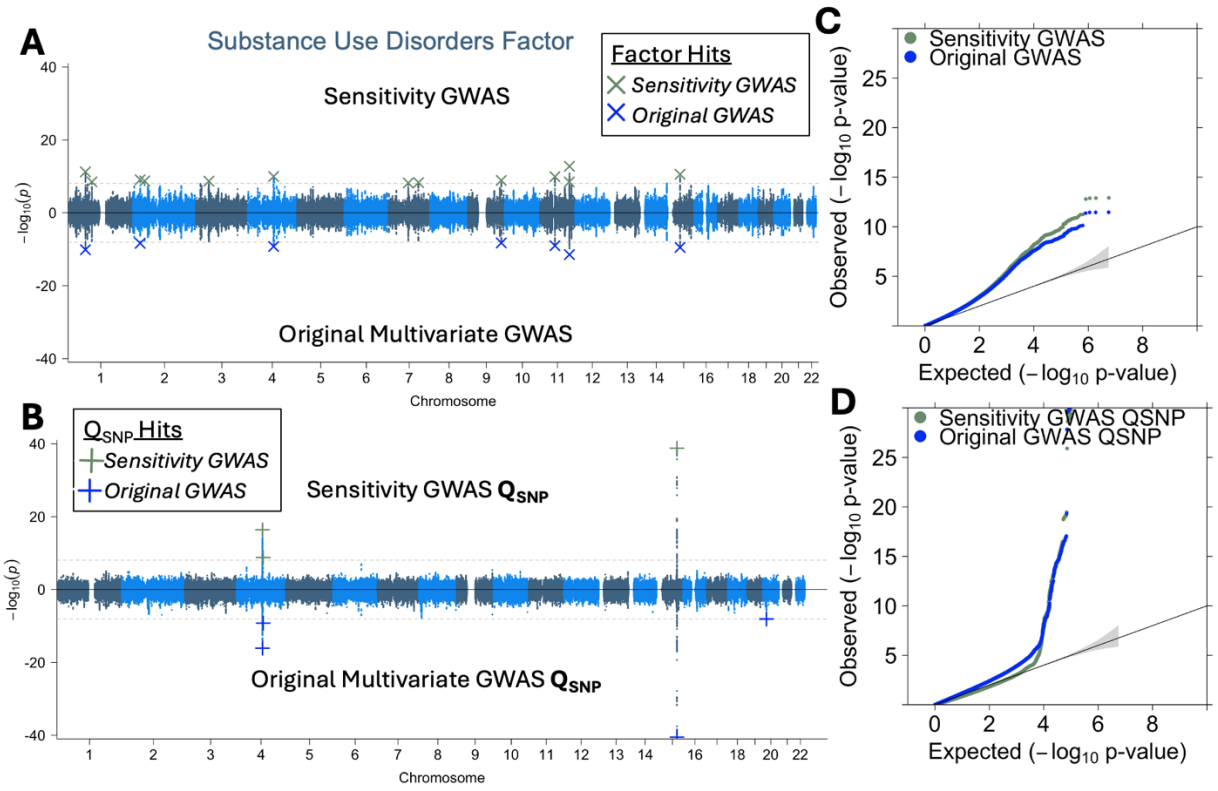
Supplementary Figure 13. Ascertainment Sensitivity Analyses: Schizophrenia and Bipolar (SB) Factor Miami Plots. Panel A depicts multivariate GWAS results for the Schizophrenia and Bipolar (SB) factor as a Miami plot with -log10(p) values for the sensitivity GWAS results on the top half and the log10(p) values of the original GWAS displayed on the bottom half. Hits at a Bonferroni corrected, genome-wide significance threshold that were not within 100kb of a QSNP are depicted with a green X for the sensitive GWAS on the top half and a blue X for the original GWAS on the bottom half. Panel B shows the -log10(p) values for the QSNP results for the sensitivity GWAS and the log(10) p values for the original GWAS. Hits at a Bonferroni corrected, genome-wide significance threshold are depicted with a green + for the sensitive GWAS on the top half and a blue + for the original GWAS on the bottom half. The gray dashed line indicates the Bonferroni corrected, genome-wide significance threshold in both panels. The GWAS *p*-values were two-sided and calculated from Z-statistics (the estimate over the standard error) from the multivariate GWAS. Panel Cs and D depict QQ-plots for the SB factor and QSNP statistic, respectively. In both panels, results for the original GWAS are shown in blue and results from the sensitivity GWAS in green.



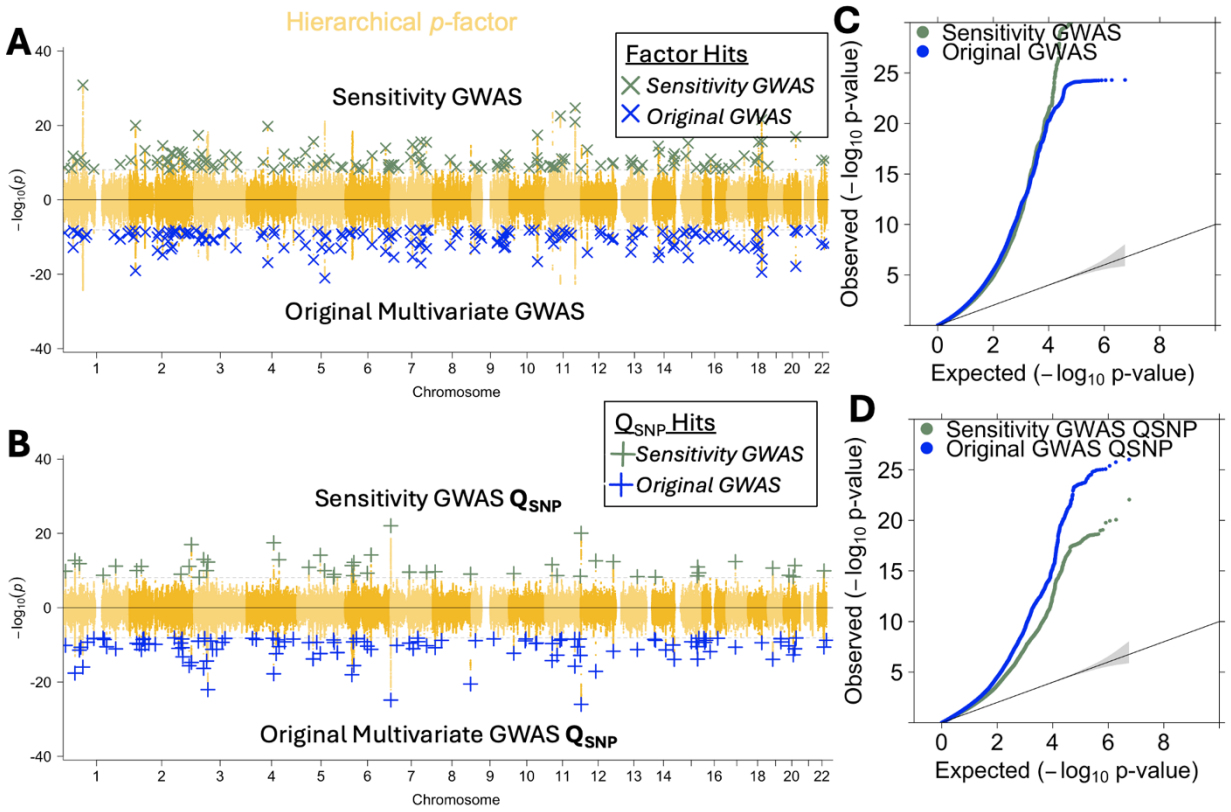
Supplementary Figure 14. Ascertainment Sensitivity Analyses: Neurodevelopmental Disorders Factor Miami Plots. Panel A depicts multivariate GWAS results for the Neurodevelopmental disorders factor as a Miami plot with -log10(p) values for the sensitivity GWAS results on the top half and the log10(p) values of the original GWAS displayed on the bottom half. Hits at a Bonferroni corrected, genome-wide significance threshold that were not within 100kb of a QSNP are depicted with a green X for the sensitive GWAS on the top half and a blue X for the original GWAS on the bottom half. Panel B shows the -log10(p) values for the QSNP results for the sensitivity GWAS and the log(10) p values for the original GWAS. Hits at a Bonferroni corrected, genome-wide significance threshold are depicted with a green + for the sensitive GWAS on the top half and a blue + for the original GWAS on the bottom half. The gray dashed line indicates the Bonferroni corrected, genome-wide significance threshold in both panels. The GWAS *p*-values were two-sided and calculated from Z-statistics (the estimate over the standard error) from the multivariate GWAS. Panel Cs and D depict QQ-plots for the Neurodevelopmental disorders factor and QSNP statistic, respectively. In both panels, results for the original GWAS are shown in blue and results from the sensitivity GWAS in green.



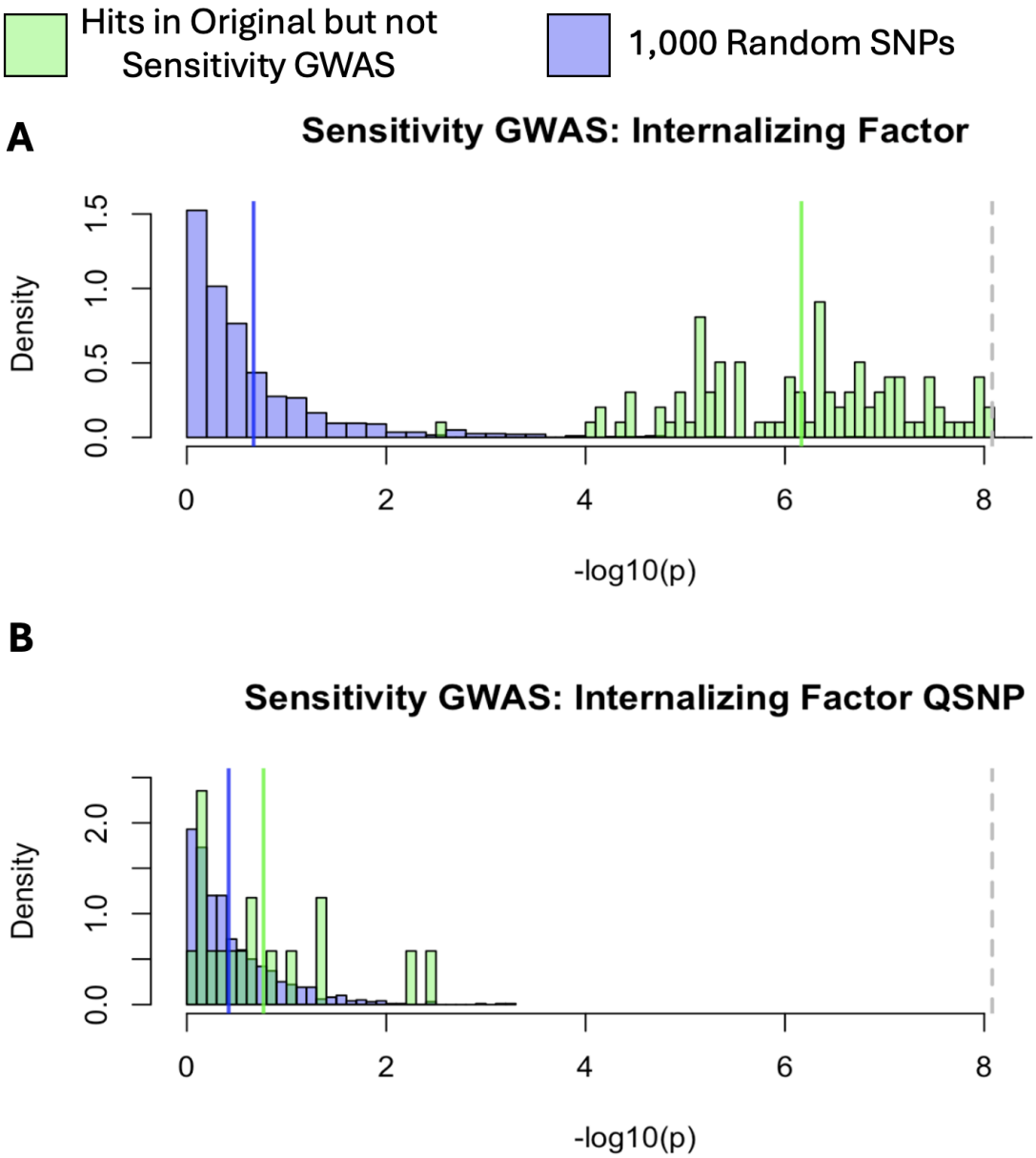
Supplementary Figure 15. Ascertainment Sensitivity Analyses: Internalizing Disorders Factor Miami Plots. Panel A depicts multivariate GWAS results for the Internalizing disorders factor as a Miami plot with -log10(p) values for the sensitivity GWAS results on the top half and the log10(p) values of the original GWAS displayed on the bottom half. Hits at a Bonferroni corrected, genome-wide significance threshold that were not within 100kb of a QSNP are depicted with a green X for the sensitive GWAS on the top half and a blue X for the original GWAS on the bottom half. Panel B shows the -log10(p) values for the QSNP results for the sensitivity GWAS and the log(10) p values for the original GWAS. Hits at a Bonferroni corrected, genome-wide significance threshold are depicted with a green + for the sensitive GWAS on the top half and a blue + for the original GWAS on the bottom half. The gray dashed line indicates the Bonferroni corrected, genome-wide significance threshold in both panels. The GWAS *p*-values were two-sided and calculated from Z-statistics (the estimate over the standard error) from the multivariate GWAS. Panel Cs and D depict QQ-plots for the Internalizing disorders factor and QSNP statistic, respectively. In both panels, results for the original GWAS are shown in blue and results from the sensitivity GWAS in green.



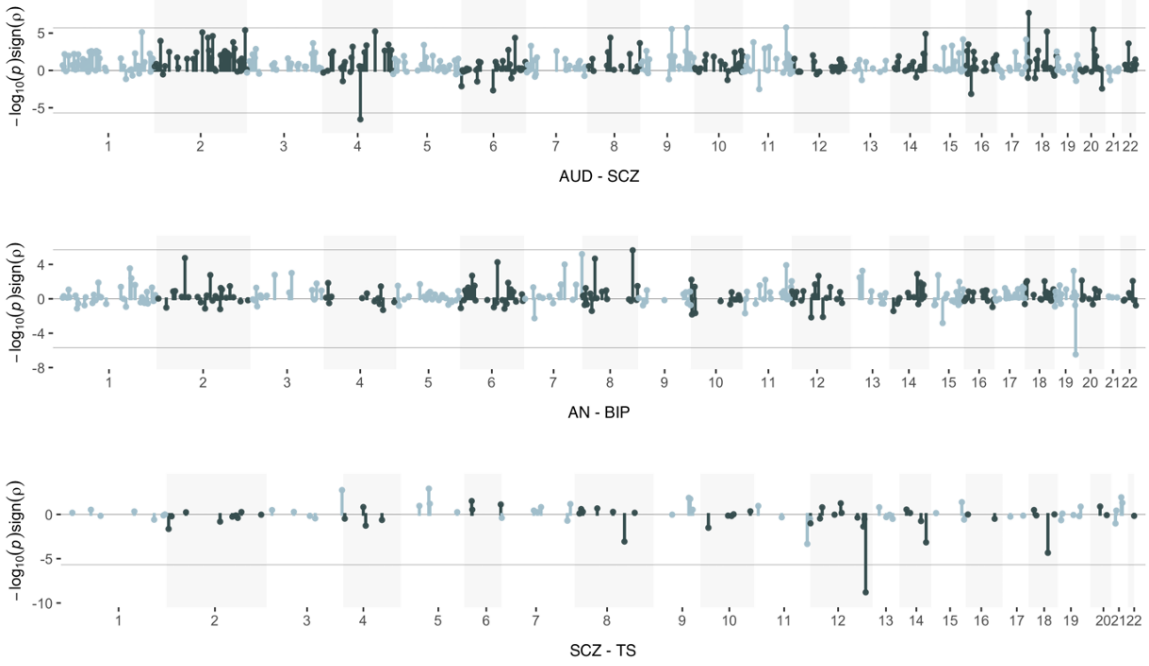
Supplementary Figure 16. Ascertainment Sensitivity Analyses: Substance Use Disorders Factor Miami Plots. Panel A depicts multivariate GWAS results for the Substance Use disorders factor as a Miami plot with -log10(p) values for the sensitivity GWAS results on the top half and the log10(p) values of the original GWAS displayed on the bottom half. Hits at a Bonferroni corrected, genome-wide significance threshold that were not within 100kb of a QSNP are depicted with a green X for the sensitive GWAS on the top half and a blue X for the original GWAS on the bottom half. Panel B shows the -log10(p) values for the QSNP results for the sensitivity GWAS and the log(10) p values for the original GWAS. Hits at a Bonferroni corrected, genome-wide significance threshold are depicted with a green + for the sensitive GWAS on the top half and a blue + for the original GWAS on the bottom half. The gray dashed line indicates the Bonferroni corrected, genome-wide significance threshold in both panels. The GWAS *p*-values were two-sided and calculated from Z-statistics (the estimate over the standard error) from the multivariate GWAS. Panel Cs and D depict QQ-plots for the Substance Use disorders factor and QSNP statistic, respectively. In both panels, results for the original GWAS are shown in blue and results from the sensitivity GWAS in green.



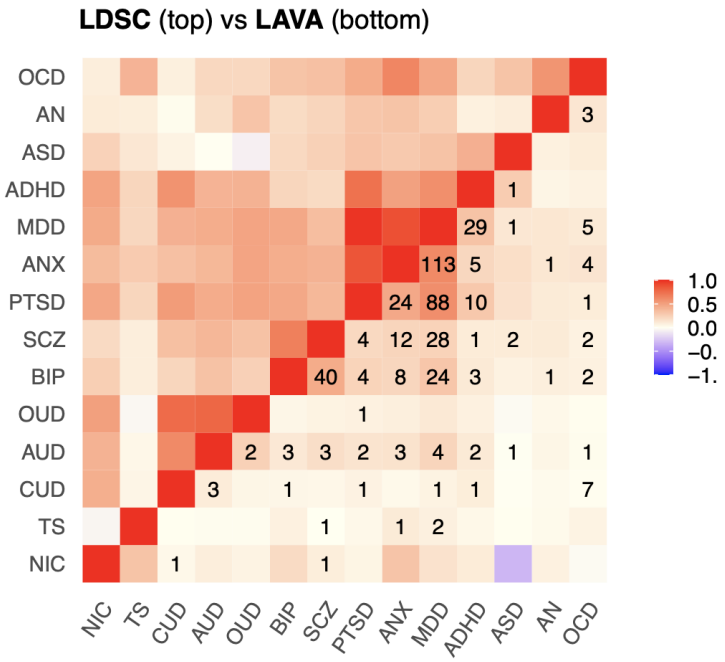
Supplementary Figure 17. Ascertainment Sensitivity Analyses: Hierarchical p-factor Miami Plots. Panel A depicts multivariate GWAS results for the Hierarchical p-factor as a Miami plot with -log10(p) values for the sensitivity GWAS results on the top half and the log10(p) values of the original GWAS displayed on the bottom half. Hits at a Bonferroni corrected, genome-wide significance threshold that were not within 100kb of a QSNP are depicted with a green X for the sensitive GWAS on the top half and a blue X for the original GWAS on the bottom half. Panel B shows the -log10(p) values for the QSNP results for the sensitivity GWAS and the log(10) p values for the original GWAS. Hits at a Bonferroni corrected, genome-wide significance threshold are depicted with a green + for the sensitive GWAS on the top half and a blue + for the original GWAS on the bottom half. The gray dashed line indicates the Bonferroni corrected, genome-wide significance threshold in both panels. The GWAS *p*-values were two-sided and calculated from Z-statistics (the estimate over the standard error) from the multivariate GWAS. Panel Cs and D depict QQ-plots for the Hierarchical p-factor and QSNP statistic, respectively. In both panels, results for the original GWAS are shown in blue and results from the sensitivity GWAS in green.



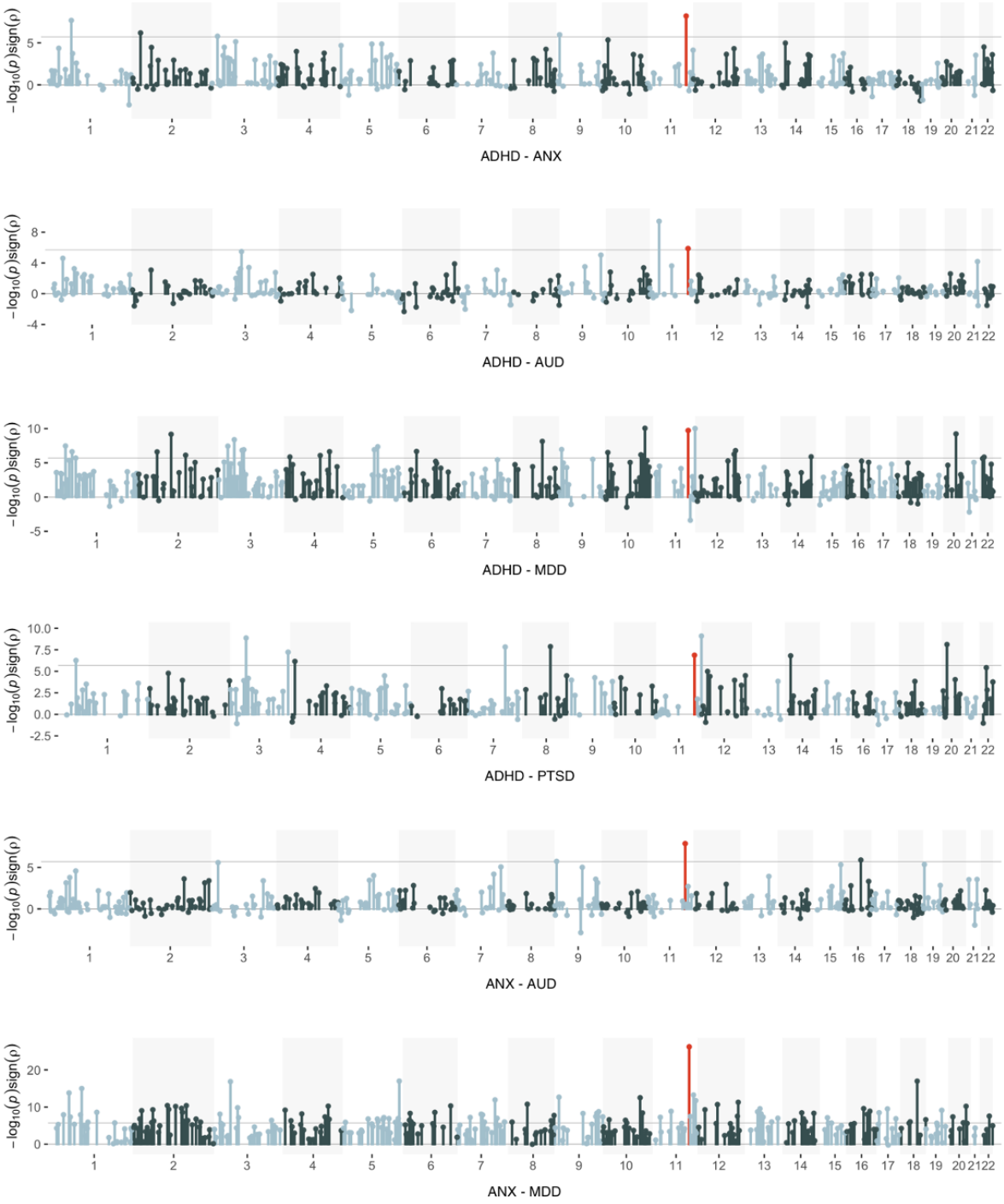
Supplementary Figure 18. Ascertainment Sensitivity Analyses: Internalizing Factor Histograms. Both panels depict in green the sensitivity GWAS -log10(*p*-values) for hits that were identified in the original GWAS, but not the sensitivity GWAS. The solid green vertical line indicates the average p-value across these SNPs. This is benchmarked against the -log10(*p*-values) for 1,000 randomly selected SNPs from the sensitivity GWAS for the internalizing factor in blue. The solid blue line indicates the average p-value across all analyzed SNPs (i.e., this average is not restricted to the 1,000 random SNPs). *Panel A* depicts the sensitivity results for the Internalizing factor. *Panel B* depicts the sensitivity results for the Internalizing factor Q_{SNP} heterogeneity metric. The GWAS *p*-values were two-sided and calculated from Z-statistics (the estimate over the standard error) from the multivariate GWAS.



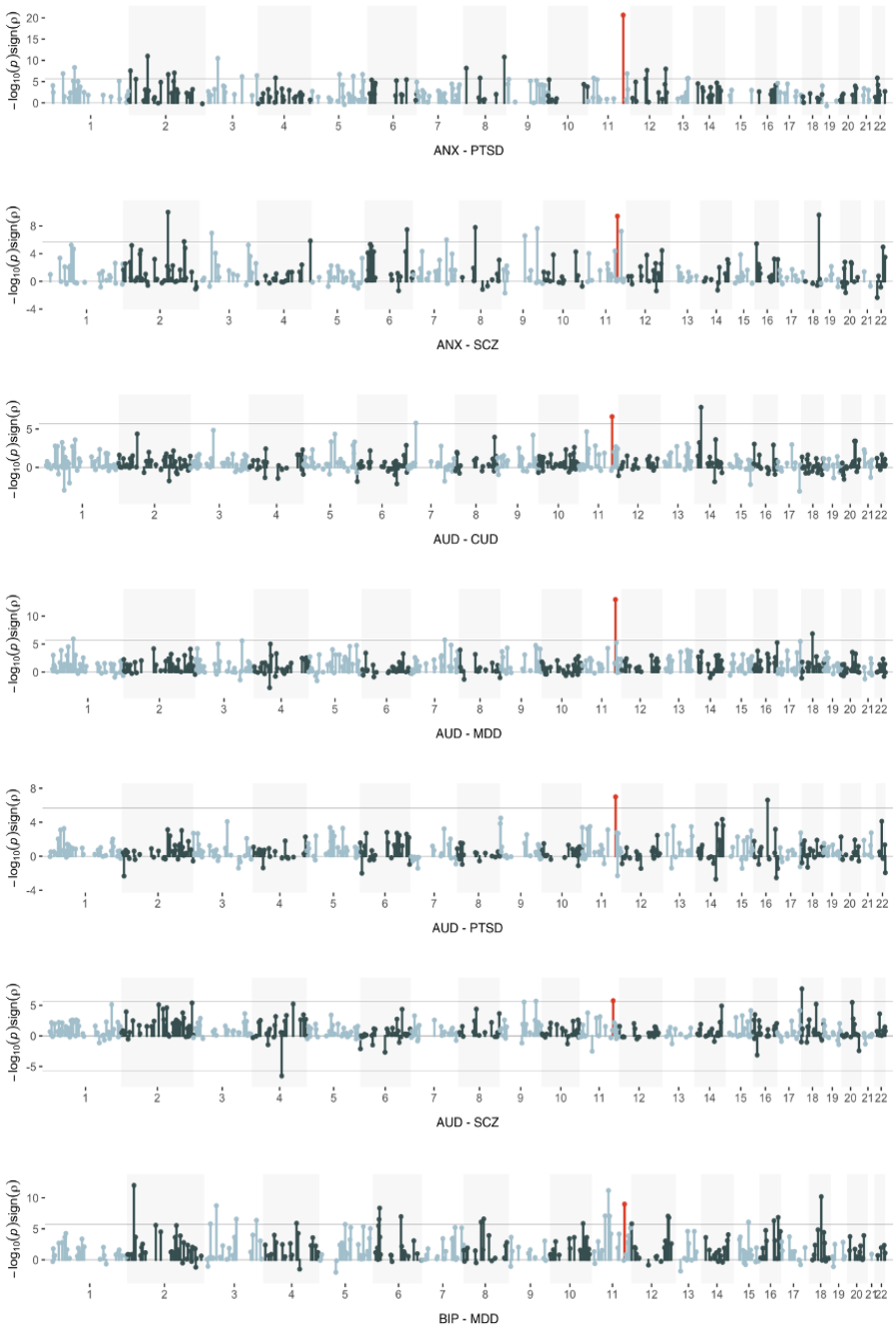
Supplementary Figure 19. Miami plots of local rg for disorder pairs that exhibited significant negative local r_g. All analyzed loci have been plotted in order of genomic position on the X-axis, binned by chromosome. The Y-axis shows the -log10 *p*-values for all tested local genetic correlations scaled by the direction of the association. Empirical p-values were obtained via a permutation procedure with partial integration, evaluating the two-sided hypothesis of no association using the estimated parameters as test statistics. The line indicates the Bonferroni corrected significance threshold of 2.1×10^{-6}.



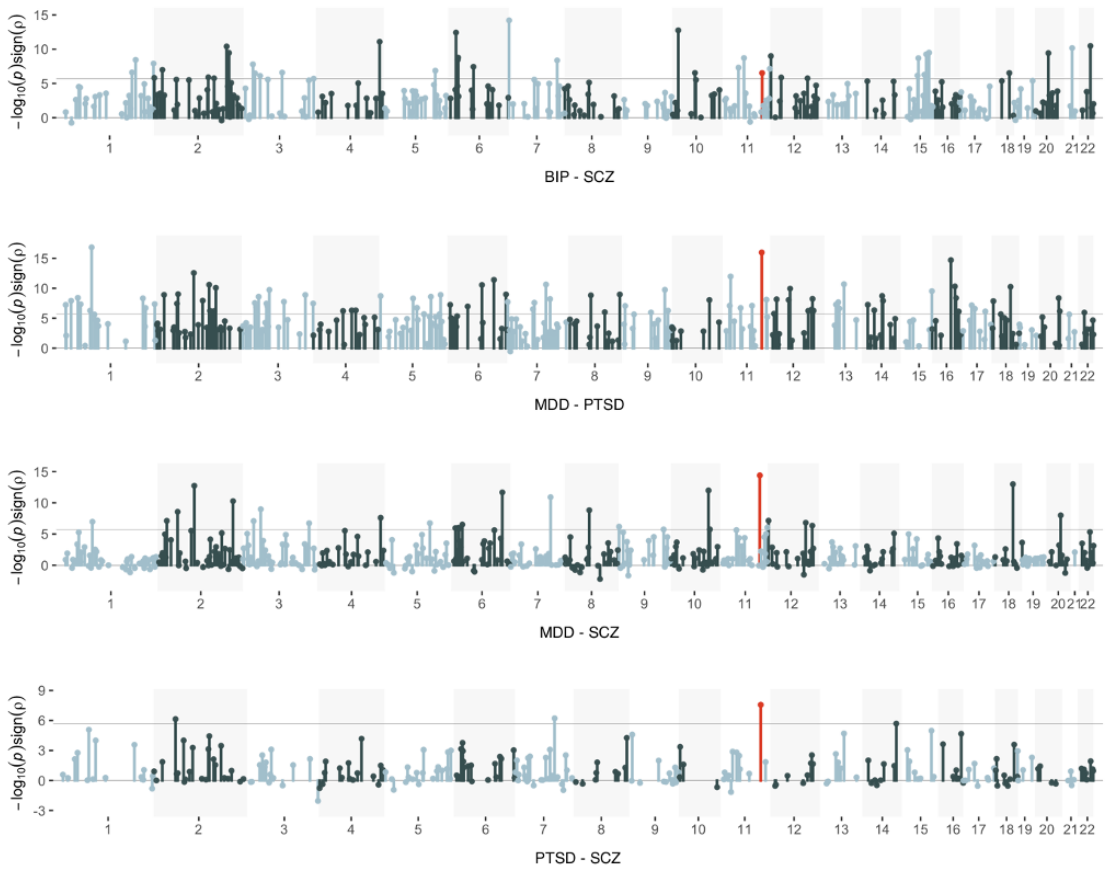
Supplementary Figure 20. Concordance between global and local genetic correlations. The global genetic correlations from LDSC have been plotted above the diagonal. This is contrasted against the average local genetic correlations across all tested loci from LAVA below the diagonal. The numbers represent the total number of significant local genetic correlations that were detected.



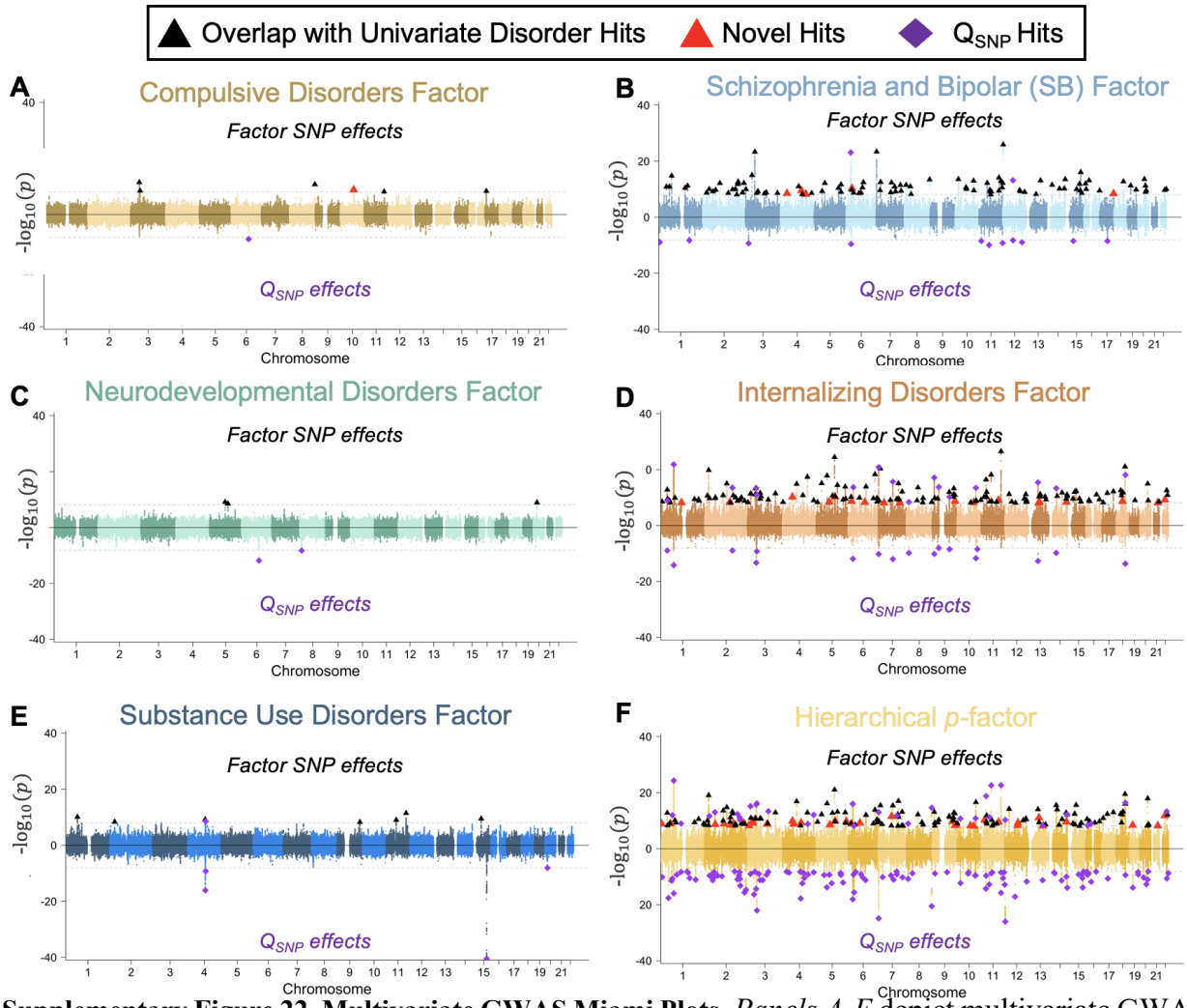
Supplementary Figure 21a. Miami plots of local rg for disorder pairs with significant rg correlation in top hotspot locus on chromosome 11. All analyzed loci have been plotted in order of genomic position on the X-axis, binned by chromosome. The Y-axis shows the -log10 *p*-values for all tested local genetic correlations scaled by the direction of the association. Empirical p-values were obtained via a permutation procedure with partial integration, evaluating the two-sided hypothesis of no association using the estimated parameters as test statistics. The line indicates the Bonferroni corrected significance threshold of 2.1×10^{-6}. The hotspot locus has been colored red.



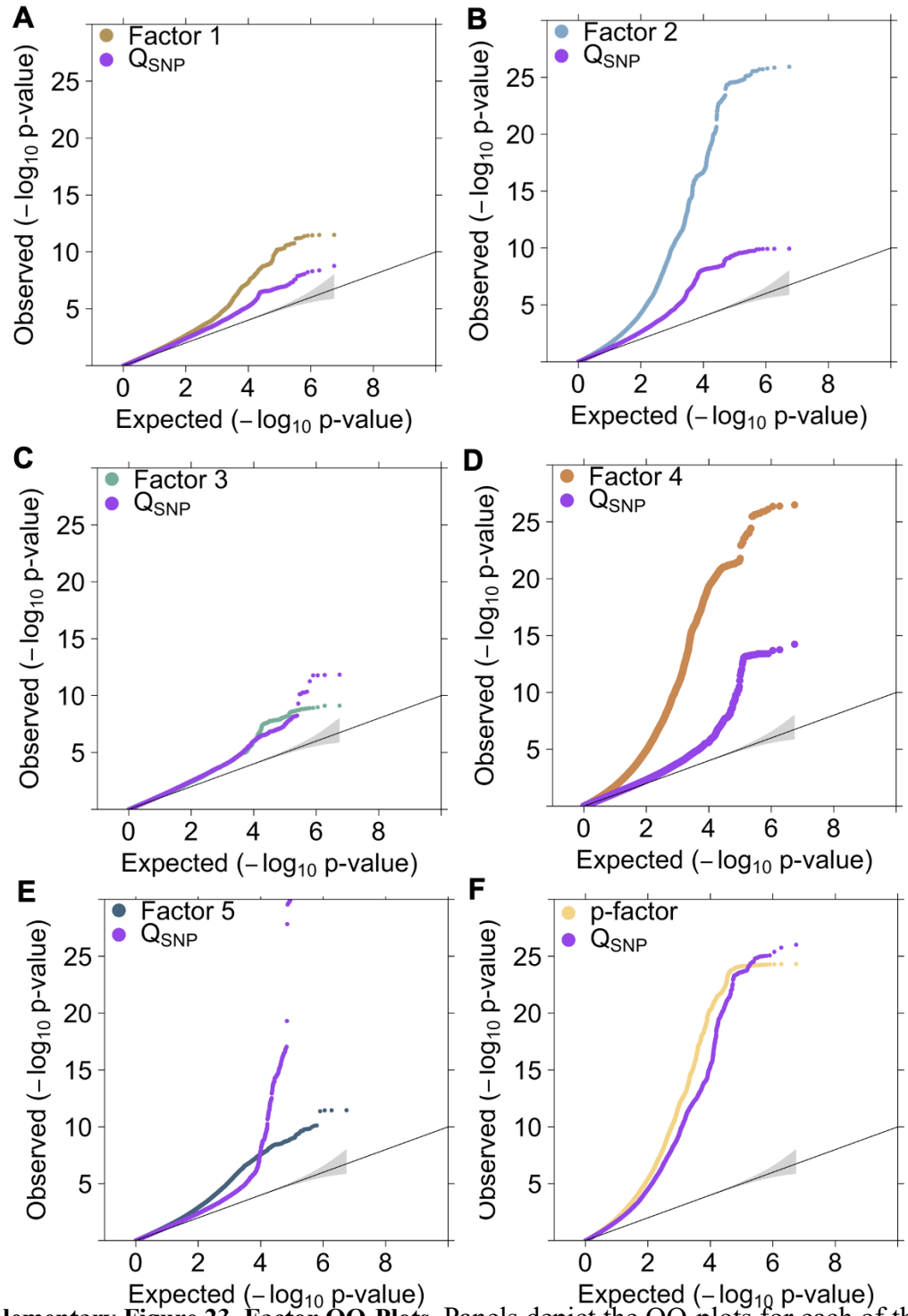
Supplementary Figure 21b. Miami plots of local rg for disorder pairs with significant rg correlation in top hotspot locus on chromosome 11. All analyzed loci have been plotted in order of genomic position on the X-axis, binned by chromosome. The Y-axis shows the -log10 *p*-values for all tested local genetic correlations scaled by the direction of the association. Empirical p-values were obtained via a permutation procedure with partial integration, evaluating the two-sided hypothesis of no association using the estimated parameters as test statistics. The line indicates the Bonferroni corrected significance threshold of 2.1×10^{-6}. The hotspot locus has been colored red.



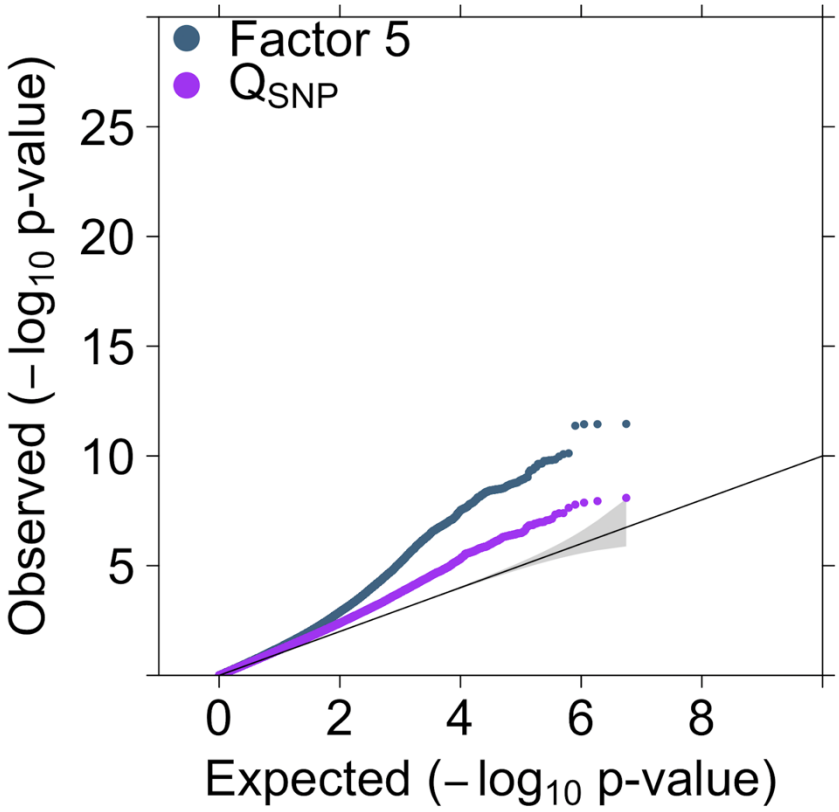
Supplementary Figure 21c. Miami plots of local rg for disorder pairs with significant rg correlation in top hotspot locus on chromosome 11. All analyzed loci have been plotted in order of genomic position on the X-axis, binned by chromosome. The Y-axis shows the -log10 *p*-values for all tested local genetic correlations scaled by the direction of the association. Empirical p-values were obtained via a permutation procedure with partial integration, evaluating the two-sided hypothesis of no association using the estimated parameters as test statistics. The line indicates the Bonferroni corrected significance threshold of 2.1×10^{-6}. The hotspot locus has been colored red.



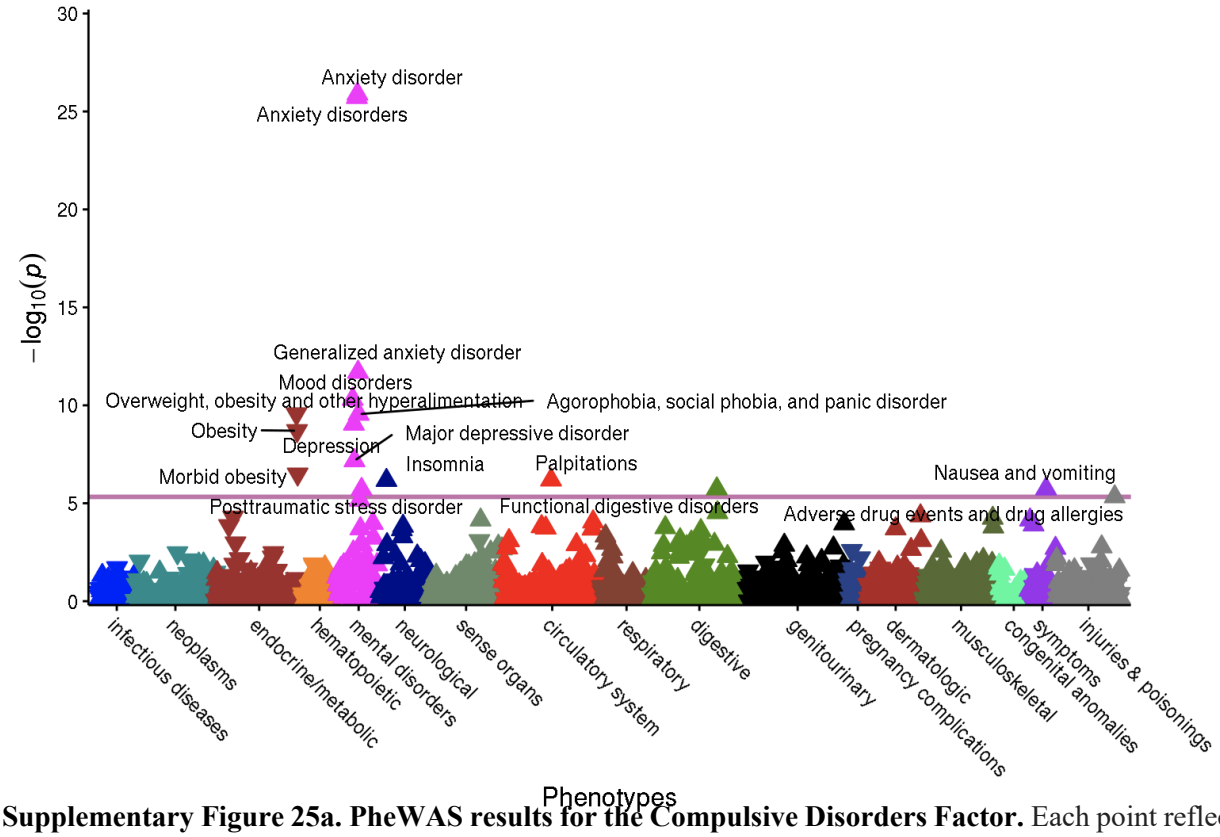
Supplementary Figure 22. Multivariate GWAS Miami Plots. *Panels A-E* depict multivariate GWAS results for the five factors from the correlated factors model and *Panel F* the results for the *p*-factor from the hierarchical model. All panels depict results for the -log10, two-tailed *p*-values for the factor on the top half of the Miami plot and the log10, one-tailed *p*-values for Q_{SNP} on the bottom half. Per the legend at the top of the figure, factors hits that were within 100kb of univariate hits are shown in black triangles, novel hits for the factors that were not within 100kb of a univariate or Q_{SNP} are depicted as red triangles, and Q_{SNP} hits as purple diamonds. The gray dashed line indicates the Bonferroni corrected, genome-wide significance threshold.



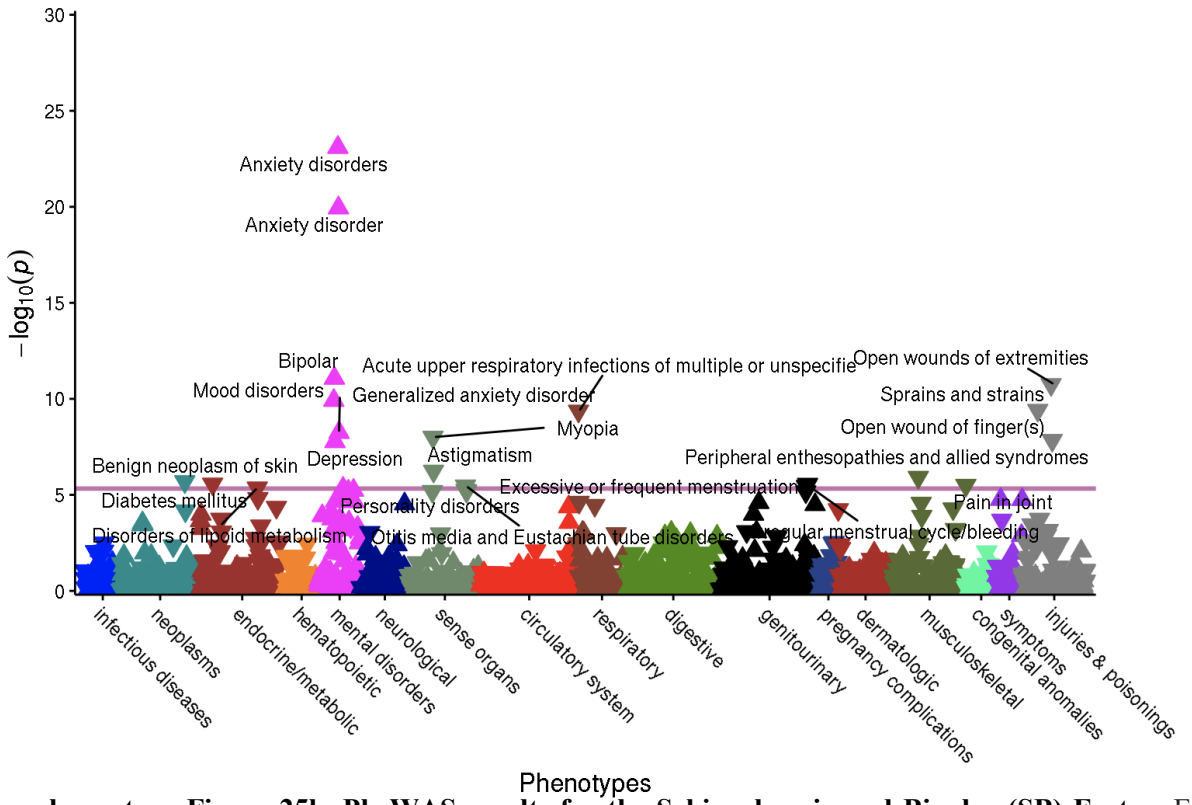
Supplementary Figure 23. Factor QQ-Plots. Panels depict the QQ-plots for each of the psychiatric factors from the multivariate GWAS analyses conducted in Genomic SEM. All panels depict factor-specific -log10, one-tailed *p*-values Q_{SNP} results in purple. *Panels A-E* depict results for the five factors from the correlated factors model and *Panel F* depicts results for the *p*-factor from the hierarchical model, with results for the factors reflecting two-tailed, -log10 *p*-values.



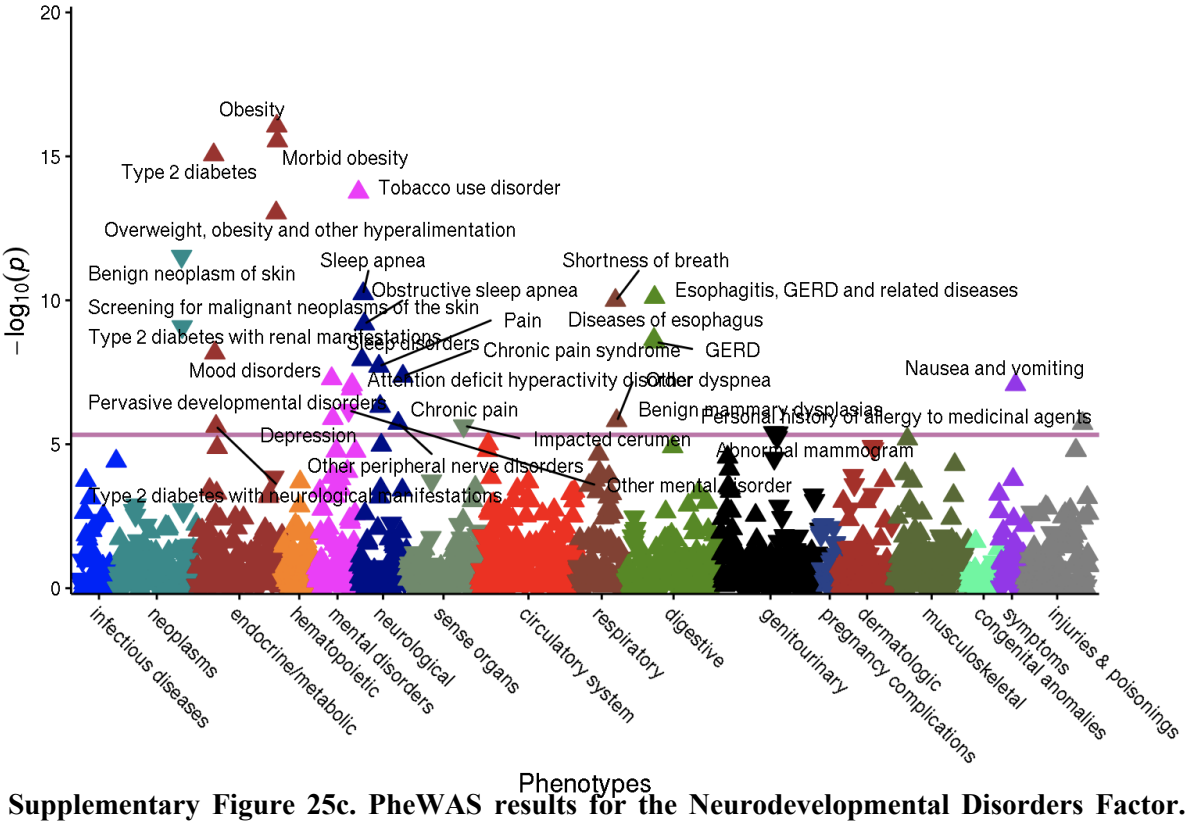
Supplementary Figure 24. Factor 5 QQ-Plot excluding top QSNP hits. Figure depicts the QQ-plot for the Substance Use disorders factor when excluding the three loci with the strongest Q_{SNP} signal in the alcohol dehydrogenase and nicotinic acetylcholine receptor genes. When these three loci in substance metabolizing pathways are removed the factor signal (depicted in blue) is stronger relative to Q_{SNP} (depicted in purple).



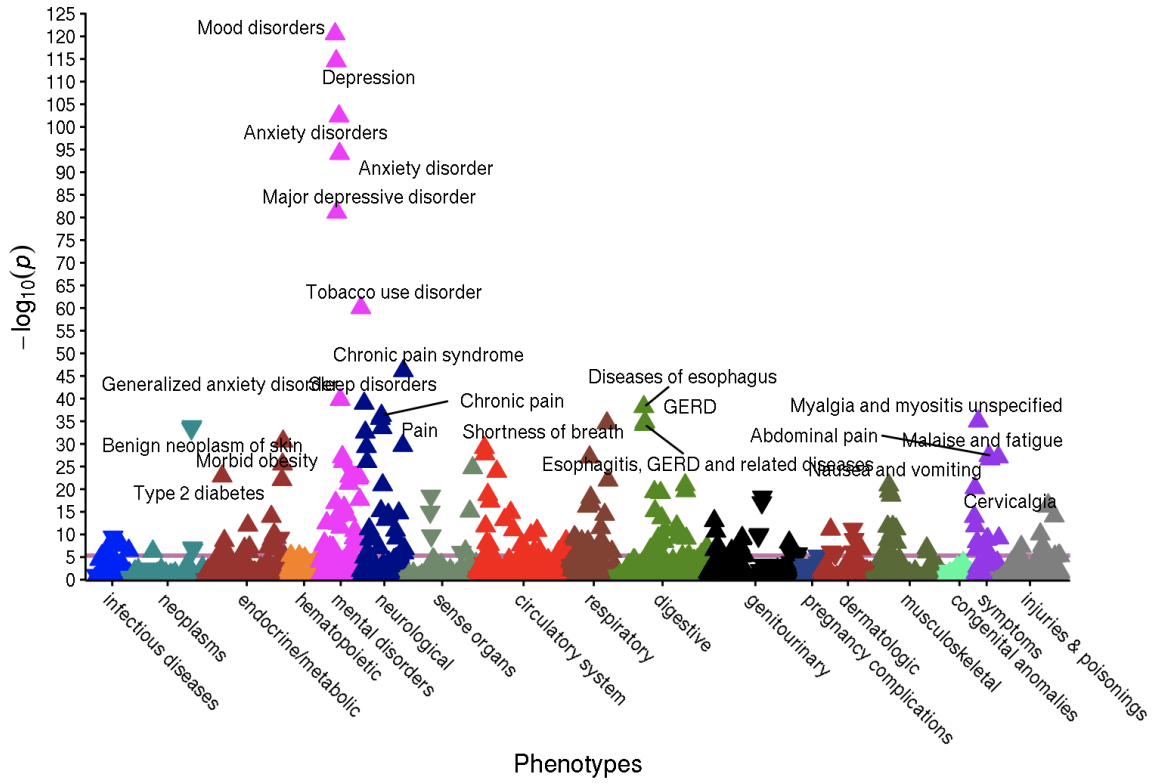
Supplementary Figure 25a. PheWAS results for the Compulsive Disorders Factor. Each point reflects associations between the different phecodes in the European-like genetic ancestry subset of the Mayo Clinic participant sample (N = 46,329) and the Compulsive disorders factor polygenic risk score (PRS). Results are grouped and color-coded on the x-axis according to different phenotype groups and vertically positioned on the y-axis according to their -log10, two-tailed p-values in the PheWAS. Upward and downward triangles indicate positive and negative PRS-phecode associations, respectively. The purple line indicates the Bonferroni corrected threshold used to define statistical significance in the current analyses (i.e., $p < 4.7\times10^{-6}$).



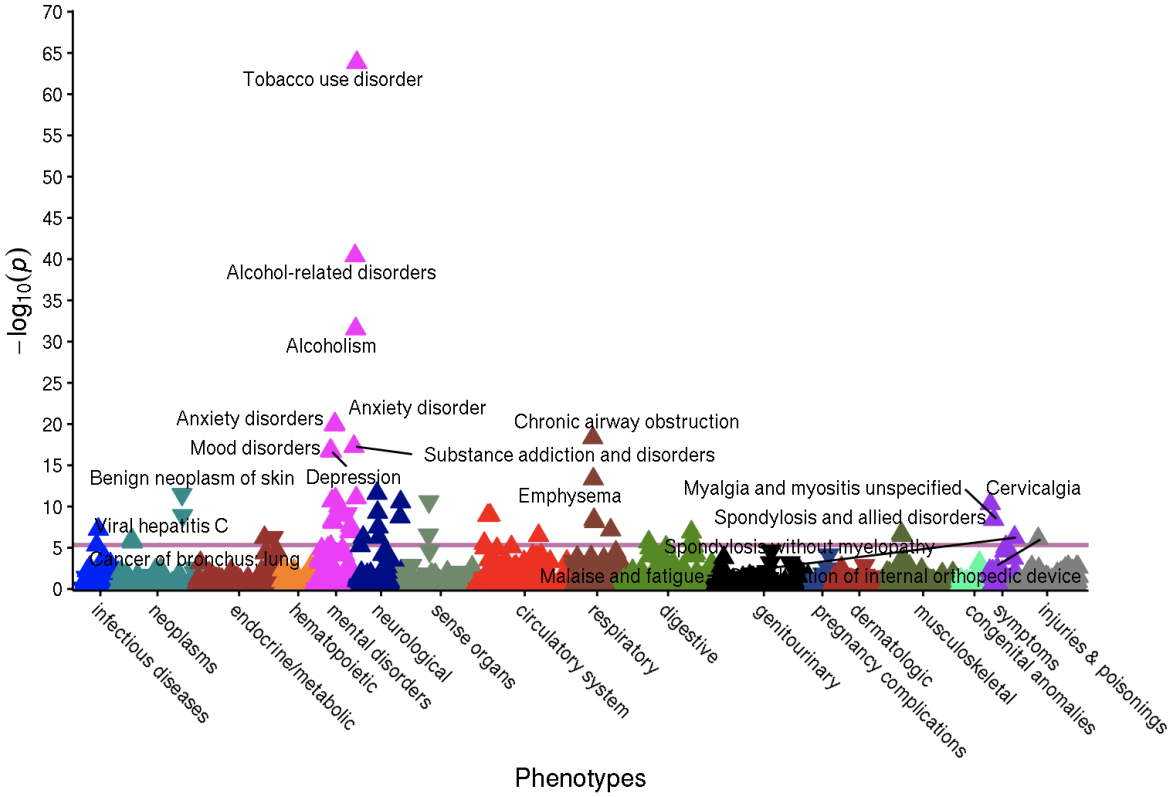
Supplementary Figure 25b. PheWAS results for the Schizophrenia and Bipolar (SB) Factor. Each point reflects associations between the different phecodes in the European-like genetic ancestry subset of the Mayo Clinic participant sample ($N = 46{,}329$) and the Schizophrenia and Bipolar factor (SB) polygenic risk score (PRS). Results are grouped and color-coded on the x-axis according to different phenotype groups and vertically positioned on the y-axis according to their -log10, two-tailed p-values in the PheWAS. Upward and downward triangles indicate positive and negative PRS-phecode associations, respectively. The purple line indicates the Bonferroni corrected threshold used to define statistical significance in the current analyses (i.e., $p < 4.7\times10^{-6}$).



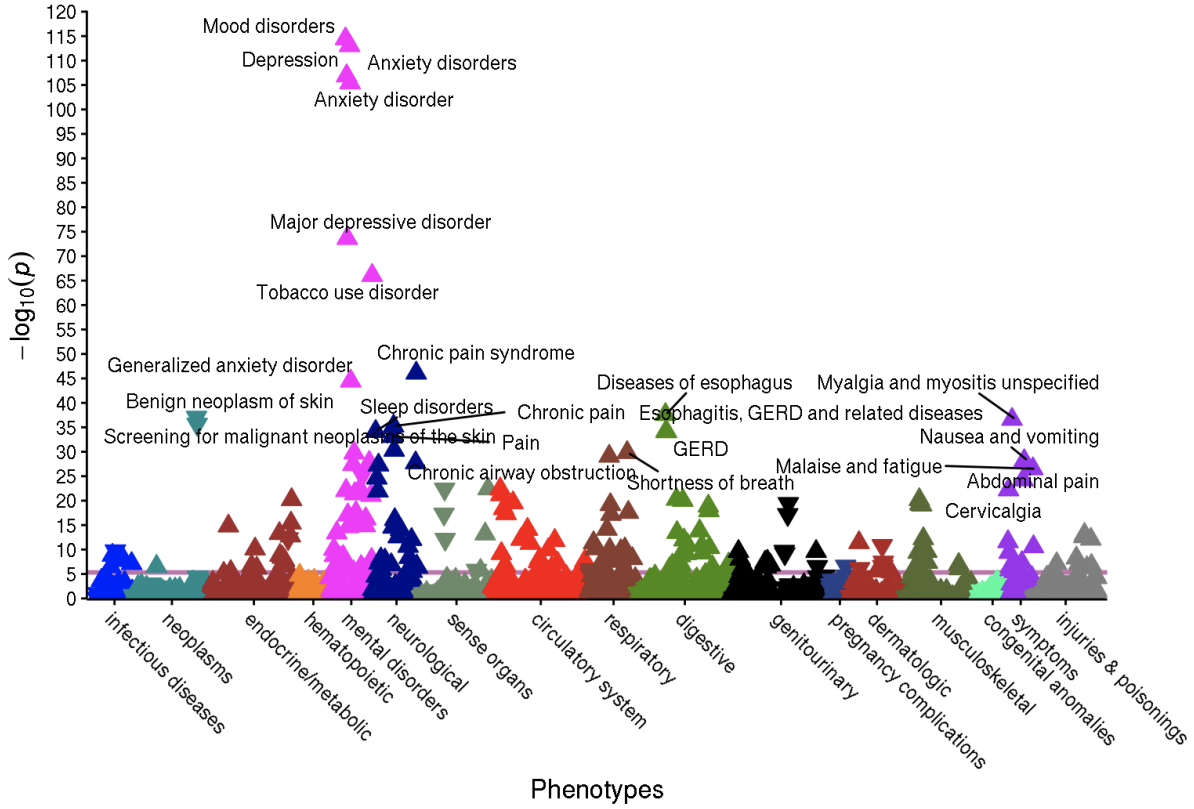
Supplementary Figure 25c. PheWAS results for the Neurodevelopmental Disorders Factor. Each point reflects associations between the different phecodes in the European-like genetic ancestry subset of the Mayo Clinic participant sample (N = 46,329) and the Neurodevelopmental disorders factor polygenic risk score (PRS). Results are grouped and color-coded on the x-axis according to different phenotype groups and vertically positioned on the y-axis according to their -log10, two-tailed p-values in the PheWAS. Upward and downward triangles indicate positive and negative PRS-phecode associations, respectively. The purple line indicates the Bonferroni corrected threshold used to define statistical significance in the current analyses (i.e., $p < 4.7\times10^{-6}$).



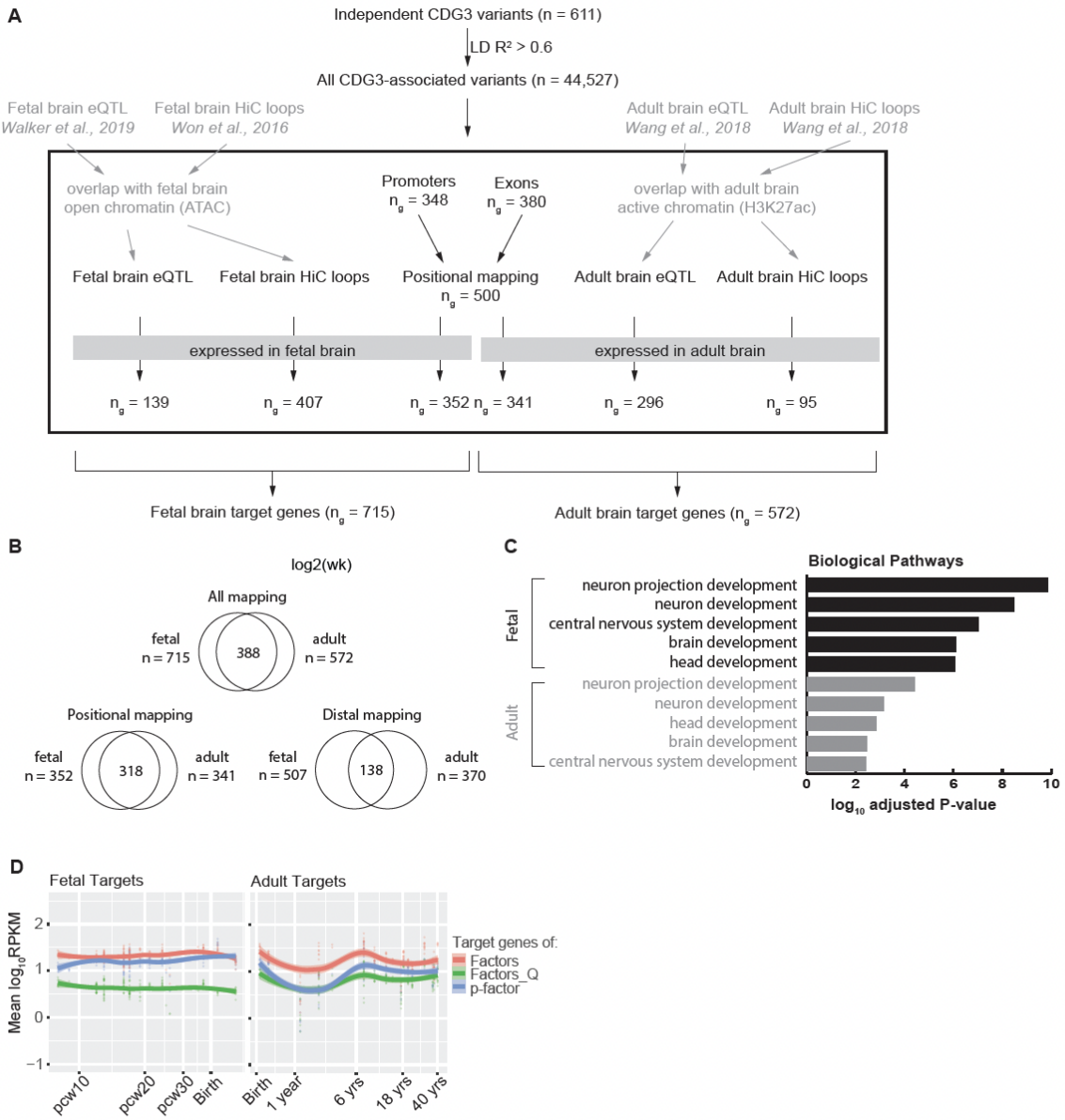
Supplementary Figure 25d. PheWAS results for the Internalizing Disorders Factor. Each point reflects associations between the different phecodes in the European-like genetic ancestry subset of the Mayo Clinic participant sample (N = 46,329) and the Internalizing disorders factor polygenic risk score (PRS). Results are grouped and color-coded on the x-axis according to different phenotype groups and vertically positioned on the y-axis according to their -log10, two-tailed p-values in the PheWAS. Upward and downward triangles indicate positive and negative PRS-phecode associations, respectively. The purple line indicates the Bonferroni corrected threshold used to define statistical significance in the current analyses (i.e., $p < 4.7\times10^{-6}$).



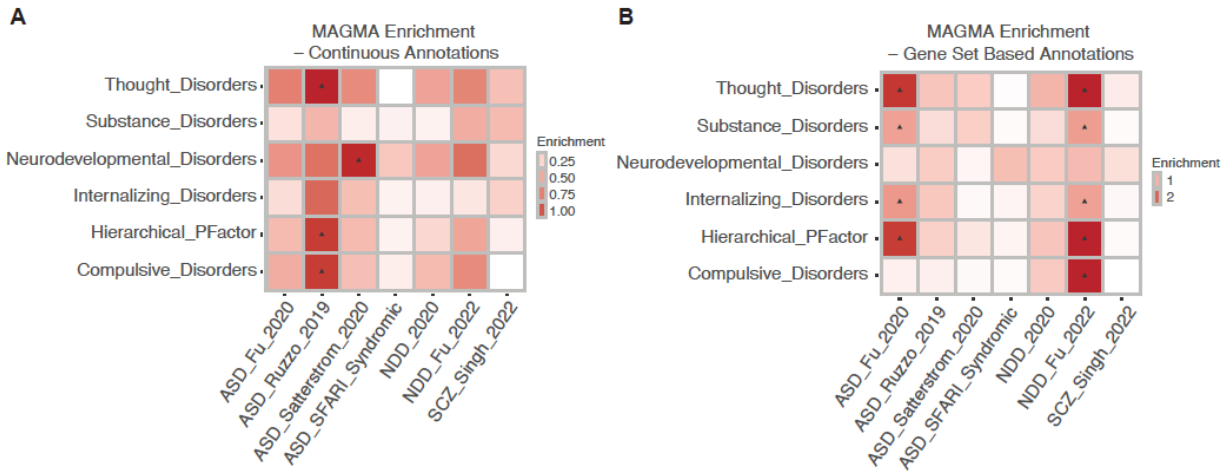
Supplementary Figure 25e. PheWAS results for the Substance Use Disorders Factor. Each point reflects associations between the different phecodes in the European-like genetic ancestry subset of the Mayo Clinic participant sample ($N = 46{,}329$) and the Substance Use disorders factor polygenic risk score (PRS). Results are grouped and color-coded on the x-axis according to different phenotype groups and vertically positioned on the y-axis according to their -log10, two-tailed p-values in the PheWAS. Upward and downward triangles indicate positive and negative PRS-phecode associations, respectively. The purple line indicates the Bonferroni corrected threshold used to define statistical significance in the current analyses (i.e., $p < 4.7\times10^{-6}$).



Supplementary Figure 25f. PheWAS results for the Hierarchical p-Factor. Each point reflects associations between the different phecodes in the European-like genetic ancestry subset of the Mayo Clinic participant sample (N = 46,329) and the Hierarchical p-factor polygenic risk score (PRS). Results are grouped and color-coded on the x-axis according to different phenotype groups and vertically positioned on the y-axis according to their -log10, two-tailed p-values in the PheWAS. Upward and downward triangles indicate positive and negative PRS-phecode associations, respectively. The purple line indicates the Bonferroni corrected threshold used to define statistical significance in the current analyses (i.e., $p < 4.7\times10^{-6}$).



Supplementary Figure 26. Predicting target genes of CDG variants and temporal expression patterns along the developmental trajectory of the human developing brain. (A) Schematics showing the procedure of target genes prediction. The independent variants were first expanded in their respective LD block, before either direct mapping based on TSS/exon overlap, or distal mapping with eQTL or HiC loops. The distal mapping was further filtered by corresponding open chromatin data. Ultimately, only genes expressed in the related tissues were retained. See Methods for more details. (B) Overlap of fetal and adult target genes. Most of the overlap is attributable to positional mapping result (bottom-left). See more details in Methods. (C) Gene ontology enrichment analysis of predicted target genes of all the target genes mapped with fetal or adult epigenomics data sets. (D) Averaged and normalized expression levels of target genes of the indicated classes along the temporal trajectory of human brain development. 'p-factor', all predicted target genes of class variants for the *p*-factor from the hierarchical model; 'Factors', all predicted target genes of variants associated with any of the five psychiatric factors from the correlated factors model; 'Factors_Q', all predicted targets of variants associated with any of the 5 factor-specific Q_{SNP} estimates from the correlated factors model.



Supplementary Figure 27. Enrichment of CDG variants near risk genes of related disorders. (A and B) Enrichment of the indicated variants near genes from the studies indicated using MAGMA. Enrichment using (A) the continuous annotation mode and (B) the gene set based annotation.

Supplementary Figure 28. Cell type specificity of CDG variants target genes. (A and B) EWCE enrichment by expression of CDG3 variants target genes associated with the indicated factors in (A) fetal brain cell types using two independent scRNA-Seq data sets[22,23] or (B) adult brain cell types using three independent snRNA-Seq data sets[24–26]. F2_Q = Schizophrenia and Bipolar (SB) $Q_{SNP;}$ F2 = Schizophrenia and Bipolar (SB) (SMI); F4_Q = Internalizing Disorders $Q_{SNP;}$ F4 = Internalizing Disorders. The target genes were segregated into those linked to "Fetal" only, "Adult" only or shared between the two categories. (C and D) Enrichment of target genes of CC-GWAS hits between MD and SCZ based on the single cell data sets as mentioned above. Note that the MD CC-GWAS hits are also MD GWAS hits, and the SCZ CC-GWAS hits are also SCZ GWAS hits. EWCE two-tailed p-values are empirically derived using a permutation test and significance was defined using an FDR corrected threshold for multiple comparisons. The implied sample size for the three depicted psychiatric factors was: Schizophrenia/Bipolar ($\hat{\boldsymbol{n}}$=127,202); Internalizing ($\hat{\boldsymbol{n}}$=1,637,337); *p*-factor ($\hat{\boldsymbol{n}}$=2,168,621).

References for Online Supplement

1. Coleman, J. R. I. *et al.* Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *Mol. Psychiatry* **25**, 1430–1446 (2020).
2. Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 1–11 (2019).
3. Cai, N. *et al.* Assessment and ascertainment in psychiatric molecular genetics: challenges and opportunities for cross-disorder research. *Mol. Psychiatry* **30**, 1627–1638 (2025).
4. Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat. Genet.* 1–11 (2020).
5. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
6. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
7. Major Depressive Disorder Working Group *et al.* Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell,* **188,** 640-652 (2025).
8. Nievergelt, C. M. *et al.* Genome-wide association analyses identify 95 risk loci and provide insights into the neurobiology of post-traumatic stress disorder. *Nat. Genet.* (2024) doi:10.1038/s41588-024-01707-9.
9. Strom, N. I. *et al.* Genome-wide association study of major anxiety disorders in 122,341 European-ancestry cases identifies 58 loci and highlights GABAergic signaling. *medRxiv* (2024) doi:10.1101/2024.07.03.24309466.
10. Strom, N. I. *et al.* Genome-wide association study identifies 30 obsessive-compulsive disorder associated loci. *Nat. Genet. 57, 1389-1401* (2025).
11. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
12. Staples, J. *et al.* PRIMUS: Rapid Reconstruction of Pedigrees from Genome-wide Estimates of Identity by Descent. *Am. J. Hum. Genet.* **95**, 553–564 (2014).
13. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, btaa1029– (2020).
14. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
15. Fu, J. M. *et al.* Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).
16. Ruzzo, E. K. *et al.* Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850–866.e26 (2019).
17. Satterstrom, F. K. *et al.* Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584.e23 (2020).
18. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).
19. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).
20. Jiao, J. *et al.* Identification of De Novo JAK2 and MAPK7 Mutations Related to Autism Spectrum Disorder Using Whole-Exome Sequencing in a Chinese Child and Adolescent Trio-Based Sample. *J. Mol. Neurosci.* **70**, 219–229 (2020).
21. Grotzinger, A. D. *et al.* Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat. Genet.* **54**, 548–559 (2022).
22. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785–801.e8 (2019).
23. Ziffra, R. S. *et al.* Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).

24. Wamsley, B. *et al.* Molecular cascades and cell type-specific signatures in ASD revealed by single-cell genomics. *Science* **384**, eadh2602 (2024).
25. Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
26. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).