Supplementary Methods

Data sources

Analyses were conducted using summary statistics from GWAS for ADHD, ASD, OCD, and TS as made available by the PGC. For TS, we combined results from the first GWAS on TS, conducted by Scharf et al. (1) and newly collected cases and controls. In total, 4,232 cases and 8,283 ancestry-matched controls were used for the analysis, which resulted in 8,868,895 variants overlapping in the meta-analysis. These summary statistics correspond to the GWAS carried out by Yu et al. (2), excluding samples from the Tic Genetic Consortium.

For ADHD, samples were collected by iPSYCH and PGC, with most of the samples genotyped using the Illumina PsychArray. Only samples of European ancestry were included in our analyses, comprising 19,099 cases and 34,194 ancestry-matched controls. In total, 8,047,421 variants overlapping across all cohorts after imputation were analyzed (3).

For ASD, we acquired the summary statistics of 18,382 cases and 27,969 ancestry-matched controls of European ancestry collected by iPSYCH and PGC. Most of the samples were genotyped with the Illumina PsychChip. After meta-analysis, 9,112,387 variants overlapping across sample sources were available (4).

For OCD, we used results from a meta-analysis of GWAS from two consortia: International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) (5) and OCD Collaborative Genetics Association Studies (OCGAS) (6), which led to a total of 2,688 affected samples and 7,037 ancestry-matched controls from Europe. Samples were genotyped with multiple different Illumina’s BeadChip arrays. After meta-analysis, 8,409,517 variants were found overlap and used for our study (7).

For all data obtained from the PGC, Ricopili pipeline (https://github.com/Ripkelab/ricopili/wiki) or comparable quality controls were carried out.

Linkage disequilibrium (LD)-score regression to estimate genetic correlation across disorders

LD-score regression analysis was carried out using the LDSC package (8). Only common SNPs (MAF > 0.01) with an imputation quality (INFO) score > 0.9 and matched with the provided HapMap3 SNPs reference were analyzed. LD scores estimated for the European samples from the 1000 Genomes phase 3 (9) were used as both the independent variable and the weight for the regression.
Investigating cross-disorder genetic architecture

In order to test for the presence of a common genetic factor that may underlie all traits of interest, we tested the common factor model using Genomic SEM for summary statistics of all disorders showing significant genetic correlation with TS. Prior to that, multivariable LDSC was carried out to obtain the covariance matrices using SNPs that survived the same quality controls for estimation of genetic correlation. Disease population prevalence used for the analysis were as follows: TS: 0.008, ADHD: 0.05, ASD: 0.01, OCD: 0.025. Fitness of model was evaluated using model chi-square, Akaike Information Criteria (AIC), Comparative Fit Index (CFI), and standardized root mean square residual (SRMR).

Causal risk factor inference

To estimate the causative association across traits, we carried out bidirectional generalized summary-data-based Mendelian randomization (GSMR) (10) across all disorders of interest. SNPs that are strongly associated with the exposure (p < 5 x 10^{-6}) were used as genetic instruments. This threshold was chosen so that all the diseases could have more than 10 near-independent genetic instruments (r^2 > 0.05) for analyses therefore test power could be granted. A heterogeneity in dependent instrument (HEIDI)-outlier approach was carried out to exclude pleiotropic SNPs (P_{HEIDI} < 0.01) that affect the outcome through pathways other than the exposure factor. We used each trait as the target and the other three as exposures respectively and ran 12 independent tests, which made the significant threshold for this analysis p < 4.17 x 10^{-3} under Bonferroni correction.

GWAS meta-analysis

To investigate the genetic variants underlying the observed overlap, cross-disorder meta-analysis was carried out for TS-ADHD-ASD jointly, as well as pairwise between TS and significantly correlated disorders. SNP-based GWAS meta-analyses was performed using ASSET (11), which takes into account dependency across studies due to sample overlap (12). For each study, the variants’ effect sizes were measured by the logarithm of the odds ratio (OR). The possibility of inflation of results was investigated through observed λ as well as the sample size -corrected value λ_{1000}. Variants with meta-analysis p-values below the genomewide significance threshold (p < 5 x 10^{-8}) were considered significant. To further highlight SNPs that contribute to risk across multiple phenotypes, we estimated the posterior probability of association (referred to as the m-value) with each disorder using a Bayesian statistical framework as implemented by MetaSoft (13). An m-value threshold of 0.9 has been recommended to predict with high confidence that a particular SNP is associated with a given disorder.

Partitioned heritability analysis

We carried out SNP partitioned heritability analysis and cell type specificity analysis for the GWAS meta-analysis results using the LDSC package as described by Finucane et al. (14). We investigated the possible enrichment of SNP heritability in 53 non-cell type specific annotation categories (baseline), including 24 main annotations and 29 extended annotations derived from the
main annotations as defined in (14). Even though these annotation categories were not mutually exclusive, we considered 53 as the number of hypothesis tested rather than 24, which gave us a more conservative significant threshold of $p < 9.43 \times 10^{-4}$ after multiple testing correction.

For the cell type specific annotations, we investigated the enrichment of SNP heritability in 13 brain relevant annotations from GTEx using reference created by Finucane et al. while controlling for the 53 baseline categories as defined in (14). After multiple testing correction, the significant threshold for this analysis was $p < 3.85 \times 10^{-3}$. We also looked into the heritability enrichment in cell type specific chromatin states using cell type specific annotation made publicly available by Finucane et al. The annotation reference included for 489 cell type specific chromatin states. The results were subject to a significant threshold of $p < 1.02 \times 10^{-4}$.

**SNP-based conditional analysis**

To further dissect the contribution of genetics on different groups of traits, we carried out multi-trait-based conditional and joint analysis (mtCOJO) (10) to adjust the summary statistics of TS-ADHD-ASD conditioning on TS-OCD and vice versa. Bidirectional causal effects between TS-ADHD-ASD and TS-OCD were first estimated using GSMR with strongly associated SNPs ($p < 1 \times 10^{-5}$). Genetic correlation, SNP-based heritability and potential covariance due to sample overlap were estimated through LD score regression, for which 1000Genomes phase 3 EUR subset was used as reference.

**Gene-based cross-disorder GWAS analysis**

Gene-based cross-disorder GWAS analysis was carried out using the MAGMA plug-in on the FUMA GWAS annotation platform (15,16). For this analysis, variants were mapped onto genes based on their exact physical positions without extended windows and aggregated association p-values were calculated for each gene. Analysis was carried out under a SNP-wise (mean) model. Considering the sample composition, a European ancestry reference from 1000 Genomes phase 3 was used as the reference panel. Analysis was done with the summary statistics of each disorder individually as well as all meta-analysis results obtained. Significance thresholds were set applying Bonferroni correction for each analysis, corresponding to the number of genes being tested.

**Gene-property analysis for tissue specificity**

To investigate phenotypic tissue specificity, a gene-property analysis testing for the relationship between tissue-specific gene expression and phenotype for associated genes was carried out using MAGMA for meta-analysis results with both GTEx v7 30 and 53 general tissue type expression atlas (17). Significant thresholds for these analyses were $p$-value $< 1.67 \times 10^{-3}$ and $p$-value $< 9.43 \times 10^{-4}$, respectively, under Bonferroni correction. The analysis was done for all meta-analyzed results.

**Gene-set analysis**

Gene-set analysis was also performed using MAGMA under a default competitive test. Gene sets and gene ontology (GO) terms tested were obtained from MsigDB v 6.1
(http://software.broadinstitute.org/gsea/msigdb), which contains 10,655 gene sets consistent across multiple sources. Bonferroni correction was applied to calculated association p-values to determine significance.

**Results annotation**

SNP-based annotation and gene mapping were carried out for significant SNPs with ANNOVAR (18), including functional predictions for all significant non-synonymous mutations using SIFT (19) and PolyPhen-2 (20) plug-ins of ANNOVAR. Regional plots for the top-variants were created for 400 kb windows using the LocusZoom platform (21). For all significant results from our SNP-based and gene-based meta-analyses, we looked up previously reported associations in the GWAS catalog (22). Aggregate functional information and tissue expression levels of the genes were acquired from the GeneCards database (23), the GTEx Portal (24), and the Expression Atlas (25). Annotation of independent genomic risk loci from the FUMA GWAS platform was also adopted under parameters LD $r^2 < 0.6$ for SNPs with association p $< 5 \times 10^{-5}$ and within 1000 kb away from the significant lead-SNP (p $< 5 \times 10^{-8}$). GO-annotation and the over-representation tests were performed using the R package ClusterProfiler v3.0.4 (26). Genes were mapped onto GO-terms based on org.Hs.eg.db (27). Enrichment of GO-terms was evaluated through a hypergeometric test (28). Network plotting was carried out using the built-in function of ClusterProfiler.

**Transcriptome-wide association study**

Association between the studied disorders and gene expression levels in the brain was evaluated through summary-data-based Mendelian Randomization. The SMR software was used and analysis was performed for each individual disorder as well as using results from our GWAS meta-analyses (29). We used GWAS summary statistics for each studied disorder (as described above), the LD structure from from 1000 Genomes European reference panel and summary statistics from brain expression quantitative trait loci (eQTL) analysis (30), which quantified the effect of SNPs over gene expression levels in brain tissue (17,31). Only variants showing a consistent allele frequency (pairwise MAF difference between datasets no more than 0.20) across all three datasets (GWAS summary statistic, 1000 Genome reference, and eQTL summary statistic) were included in the analysis. All transcript probes with at least one cis-eQTL site showing peQTL $< 5 \times 10^{-8}$ were taken into consideration. SNPs affecting the same probe with LD $r^2 > 0.9$ or $< 0.05$ were pruned out from the analyses. Significance thresholds were based on Bonferroni correction for the number of probes tested.

To further verify that the effect of a probe on the trait was mediated by shared causal variants affecting both its expression and the trait rather than different variants in LD, we also carried out the HEIDI test to evaluate the heterogeneity in the effect sizes of SNPs over trait and expression for each probe, evaluated as $p_{\text{HEIDI}}$. As a default of the software, only SNPs with peQTL $< 1.5654 \times 10^{-3}$ were taken forward for this analysis. Up to top 20 independent SNPs in the cis-eQTL region were used for each tested probe to optimize the test power. A $p_{\text{HEIDI}} > 0.05$ indicates the existence
of a shared cause underlying the expression level of a transcript probe and the trait, suggesting dysregulation of the transcript is functionally relevant to the trait.

Supplemental References

1. Scharf JM, Yu D, Mathews CA, Neale BM, Stewart SE, Fagerness JA, et al. (2013): Genome-wide association study of Tourette’s syndrome. *Mol Psychiatry* 18: 721–728.

2. Yu D, Sul JH, Tsetsos F, Nawaz MS, Huang AY, Zelaya I, et al. (2019): Interrogating the Genetic Determinants of Tourette’s Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry* 176: 217–227.

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

4. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. (2019): Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* 51: 431–444.

5. Stewart SE, Yu D, Scharf JM, Neale BM, Fagerness JA, Mathews CA, et al. (2013): Genome-wide association study of obsessive-compulsive disorder. *Mol Psychiatry* 18: 788–798.

6. Mattheisen M, Samuels JF, Wang Y, Greenberg BD, Fyer AJ, McCracken JT, et al. (2015): Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. *Mol Psychiatry* 20: 337–44.

7. Arnold PD, Askland KD, Barlassina C, Bellodi L, Bienvenu OJ, Black D, et al. (2018): Revealing the complex genetic architecture of obsessive–compulsive disorder using meta-analysis. *Mol Psychiatry* 23: 1181–1188.

8. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. (2015): An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47: 1236–41.

9. 1000 Genomes Project Consortium (2015): A global reference for human genetic variation. *Nature*. Retrieved February 8, 2017, from http://www.nature.com/nature/journal/v526/n7571/abs/nature15393.html

10. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. (2018): Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 9. https://doi.org/10.1038/s41467-017-02317-2

11. Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, et al. (2012): A Subset-Based Approach Improves Power and Interpretation for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits. *Am J Hum Genet* 90: 821–835.

12. Lin D-Y, Sullivan PF (2009): Meta-Analysis of Genome-wide Association Studies with Overlapping Subjects. *Am J Hum Genet* 85: 862–872.

13. Han B, Eskin E (2012): Interpreting Meta-Analyses of Genome-Wide Association Studies ((K. Kerr, editor)). *PLoS Genet* 8: e1002555.

14. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. (2015): Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 47: 1228–1235.
15. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: Generalized Gene-Set Analysis of GWAS Data ((H. Tang, editor)). *PLOS Comput Biol* 11: e1004219.

16. Watanabe K, Taskesen E, van Bochooven A, Posthuma D (2017): Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8: 1826.

17. GTEx Consortium (2017): Genetic effects on gene expression across human tissues. *Nature* 550: 204–213.

18. Wang K, Li M, Hakonarson H (2010): ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164–e164.

19. Kumar P, Henikoff S, Ng PC (2009): Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4: 1073–1081.

20. Adzhubei I, Jordan DM, Sunyaev SR (2013): Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet* 76: 7.20.1-7.20.41.

21. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, *et al.* (2010): LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26: 2336–2337.

22. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, *et al.* (2017): The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 45: D896–D901.

23. Safran M, Dalah I, Alexander J, Rosen N (2010): GeneCards Version 3: the human gene integrator. Retrieved February 8, 2017, from http://database.oxfordjournals.org/content/2010/baq020.long

24. Carithers L, Moore H (2015): The genotype-tissue expression (GTEX) project. Retrieved February 8, 2017, from http://online.liebertpub.com/doi/pdf/10.1089/bio.2015.29031.hmm

25. Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, *et al.* (2018): Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* 46: D246–D251.

26. Yu G, Wang L-G, Han Y, He Q-Y (2012): clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi A J Integr Biol* 16: 284–287.

27. Carlson M (2018): org.Hs.eg.db: Genome wide annotation for Human. *R Packag version 370*. Retrieved December 20, 2018, from https://bioconductor.org/packages/release/data/annotation/manuals/org.Hs.eg.db/man/org.Hs.eg.db.pdf

28. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G (2004): GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20: 3710–3715.

29. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, *et al.* (2016): Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 48: 481–487.

30. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, *et al.* (2018): Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* 9: 2282.

31. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, *et al.* (2017): An xQTL map
integrates the genetic architecture of the human brain’s transcriptome and epigenome. *Nat Neurosci* 20: 1418–1426.

**Supplementary Table Legends**

**See Excel file for tables**

**Table S1.** Full results for joint genetic architecture analysis using GenomicSEM, including model fitness, standardized and unstandardized results. Also see figure 1.

**Table S2.** Full results for the causality inference using GSMR. #snps denotes the number of SNPs used for the analysis. Also see Figure 2.

**Table S3.** Summary statistics for all significant results from SNP-based GWAS meta-analyses across TS, ADHD, ASD and OCD. *m-value* = Posterior probability for association for each individual disorder; *SIFT/Poly1/Poly2* = functional prediction for nonsynonymous exonic SNPs; *HetISq* = heterozygosity I² statistic; *HetChiSq* = heterozygosity chi-square statistic; *HetPVal* = heterozygosity test p-value; *disorder-OR/P* = odds ratio statistic and p-value in the original individual disorder GWAS study.

**Table S4.** Full annotation of top genomic risk regions from SNP-based GWAS meta-analyses. An asterisk (*) indicates novel LD regions not been reported associated with corresponding traits in published GWAS. *rsID* = rsID of the leading SNP of the region; *p* = p-value of the leading SNP from the meta-analysis; *Study* = Previous studies reporting significant association at this locus; *trait* = trait reported associated with the locus by the study; *reported gene* = gene reported by the study; *mapped gene* = gene mapped onto the reported region.

**Table S5.** Comparison of statistics for matching SNPs from our TS-ADHD-ASD and TS-OCD SNP-based GWAS meta-analysis results with PGC eight-disorder GWAS meta-analysis. Table includes leading SNPs in regions with genomewide significant pleiotropic SNPs identified by the SNP-based analysis or the eight-disorder cross-disorder analysis from PGC (corresponds to selected light blue and clear rows in table 2). Red font denotes SNPs found genomewide significant and pleiotropic, and test statistics (p-value, OR and m-values for disorders analyzed) for the same SNP are reported for both studies if available.

**Table S6.** Detailed results comparison between TS-ADHD-ASD, TS-OCD SNP and gene-based GWAS meta-analysis and PGC eight-disorder GWAS meta-analysis (Psychiatric Genomics Consortium et al., 2019). Regions satisfying at least one of the three following criteria are included: 1. hosting genomewide significant SNPs in the TS-ADHD-ASD, TS-OCD SNP-based analysis; 2. genomewide significant genes from the TS-ADHD-ASD, TS-OCD gene-based analysis; 3. hosting genomewide significant SNPs that also have m-value > 0.9 across the disorders of interest here (ie TS-ADHD-ASD, TS-OCD) in the 8-disorder analysis. For each region, the following are shown: region basepair position, number of genomewide significant SNPs with m-value>0.9 in all
disorders analyzed, leading SNP, leading SNP p-value, OR and m-values for all disorders of interest from SNP based analysis and Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2019; significant genes from the gene-based analysis, p-values and whether the gene is still significant when analyzing using only SNPs with m-value > 0.9 in all disorders of interest; leading SNP, p-value and OR from the original individual GWAS analyses. Asterisk (*) indicates the region is also highlighted by the TWAS analysis.

Table S7. Partitioned heritability analysis. Baseline results included 53 non-cell type specific annotations; brain cell types included 13 brain relevant cell type specific annotations; chromatin included results for 489 cell type specific annotation of chromatin states, as described by Finucane et al. (14). Asterisk (*) in the significant column denotes the annotation categories significantly enriched under multiple testing correction.

Table S8. Significant results from conditional analyses (TS-ADHD-ASD conditioned on TS-OCD and vice versa), compared with original meta-analyses results. Including SNPs that are genomewide significant in either the original meta-analysis results or the conditioned. \( b, se, pval \) correspond to beta, standard error and p-value in the original meta-analyses results respectively; \( bC, bC\_se, bC\_pval \) correspond to conditioned beta, standard error and p-value. \( diff \) = increment (+) or decrement (-) of effect (in terms of z-score) after conditioned.

Table S9. Significant genes from gene-based GWAS analyses. P-values from individual disorder gene-based analyses are also shown.

Table S10. Significant results from cross-disorder tissue specificity analysis, testing 53 tissue types from GTEx v7 tissue expression atlas. The significance threshold is set following Bonferroni correction (p < 9.43 x 10^{-4})

Supplementary Tables

Table S11. Transcriptome-wide analysis. Significant results from transcriptome-wide analysis, using SMR.

| CHR | Gene       | TS-ADHD-ASD |
|-----|------------|-------------|
|     |            | Beta | SE   | \( p_{SMR} \) | \( P_{HEIDI} \) |
| 17  | LRRC37A4P  | -0.0353 | 0.0073 | 1.38E-06 | 9.57E-02 |
| 17  | RP11-707O23.5 | 0.0335 | 0.0069 | 1.26E-06 | 6.93E-02 |
**Supplementary Figures**

**Figure S1. Gene networks plot (gene-based).** Top ten gene networks from top 200 genes from gene-based analysis results. A. TS-ADHD-ASD gene-based network plot; B. TS-OCD gene-based network plot.
Figure S2. Tissue specificity analyses (30 tissue types). Cross-disorder tissue specificity analysis testing 30 general tissue types from GTEx v7 tissue expression atlas. Red bar indicates significant enrichment of gene expression in corresponding tissue under Bonferroni correction ($p < 1.67 \times 10^{-3}$). Panel on top right corner of each figure shows detailed statistics for significantly enriched tissue. A. TS-ADHD-ASD cross-disorder tissue specific expression enrichment; B. TS-OCD cross-disorder tissue specific expression enrichment. See also Table S9 and Figure S3.
Figure S3. Tissue specificity analysis (53 tissue types). Tissue specificity analysis, testing 53 tissue types from GTEx v7 tissue expression atlas. Red bar indicates significant enrichment of gene expression in corresponding tissue under Bonferroni correction ($p < 9.43 \times 10^{-4}$). Panel on top right corner of each figure shows detailed statistics for significantly enriched tissue. A. TS-ADHD-ASD cross-disorder tissue specific expression enrichment; B. TS-OCD cross-disorder tissue specific expression enrichment.