Cross-tissue transcriptome-wide association studies identify susceptibility genes shared between schizophrenia and inflammatory bowel disease

Florian Uellendahl-Werth, Carlo Maj, Oleg Borisov, Simonas Juzenas, Eike Matthias Wacker, Isabella Friis Jørgensen, Tim Alexander Steiert, Saptarshi Bej, Per Hoffmann, Christoph Schramm, Olaf Wolkenhauer, Karina Banasik, Søren Brunak, Stefan Schreiber, Tom Hemming Karlsen, Franziska Degenhardt, Markus Nöthen, Andre Franke, Trine Folseraas & David Ellinghaus

Genetic correlations and an increased incidence of psychiatric disorders in inflammatory-bowel disease have been reported, but shared molecular mechanisms are unknown. We performed cross-tissue and multiple-gene conditioned transcriptome-wide association studies for 23 tissues of the gut-brain-axis using genome-wide association studies data sets (total 180,592 patients) for Crohn’s disease, ulcerative colitis, primary sclerosing cholangitis, schizophrenia, bipolar disorder, major depressive disorder and attention-deficit/hyperactivity disorder. We identified NR5A2, SATB2, and PPP3CA (encoding a target for calcineurin inhibitors in refractory ulcerative colitis) as shared susceptibility genes with transcriptome-wide significance both for Crohn’s disease, ulcerative colitis and schizophrenia, largely explaining fine-mapped association signals at nearby genome-wide association study susceptibility loci. Analysis of bulk and single-cell RNA-sequencing data showed that PPP3CA expression was strongest in neurons and in enteroendocrine and Paneth-like cells of the ileum, colon, and rectum, indicating a possible link to the gut-brain-axis. PPP3CA together with three further suggestive loci can be linked to calcineurin-related signaling pathways such as NFAT activation or Wnt.

1 Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany. 2 Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany. 3 Institute of Biotechnology, Life Science Centre, Vilnius University, Vilnius, Lithuania. 4 Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 5 Department of Systems Biology & Bioinformatics, University of Rostock, Rostock, Germany. 6 Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. 7 Institute of Human Genetics, School of Medicine, University of Bonn & University Hospital Bonn, Bonn, Germany. 8 First Department of Internal Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. 9 Stellenbosch Institute of Advanced Study (STIAS), Stellenbosch University, South Africa. 10 Leibniz Institute for Food Systems Biology, Technical University Munich, Freising, Germany. 11 First Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany. 12 Research Institute for Internal Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway. 13 Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Duisburg, Germany. ✉ Email: d.ellinghaus@ikmb.uni-kiel.de

https://doi.org/10.1038/s42003-022-03031-6

Communications Biology | (2022) 5:80 | https://doi.org/10.1038/s42003-022-03031-6 | www.nature.com/commsbio
The direct link (gut-brain-axis [GBA]) of intestinal dysfunction and inflammation with brain development and mental illness is one of the most current topics in biomedical research. Gastrointestinal symptoms have been observed in patients with neurobehavioral, neurodevelopmental, and mental diseases, including attention-deficit/hyperactivity disorder (ADHD), schizophrenia (SCZ), and major depressive disorder (MDD). Vice versa, psychiatric comorbidity, including depression, anxiety, bipolar disorder (BD), and SCZ, is known to be more prevalent in inflammatory bowel disease (IBD) patients. Given the polygenic nature of most psychiatric and chronic inflammatory diseases, some of the (potentially shared) phenotypic variation in disease risk is expected to be explained by expression quantitative trait loci (eQTL) active in disease-relevant tissues of the GBA. A significant shared proportion of genomic correlation was detected between inflammatory (Crohn’s disease [CD], ulcerative colitis [UC]), and psychiatric traits (SCZ, BD) using LD score regression (LDSC). However, it remains unclear, whether the correlation is due to a few loci or is distributed across the genome.

In this study, we performed cross-tissue transcriptome-wide association studies (TWAS) for 23 tissues of the GBA using genome-wide association study (GWAS) summary statistics from ten case-control GWAS data sets (total 180,592 patients and 290,737 nonoverlapping controls, Supplementary Data 1; Methods) of three clinically related inflammatory diseases of the gastrointestinal tract (CD, UC and primary sclerosing cholangitis (PSC) because PSC patients suffer from a highly increased frequency (62–83%) of IBD called PSC-IBD) and four diseases of the mind and brain (SCZ, MDD, BD, and ADHD) for which to moderate strong genetic correlation ($r_g > 0.3$) at the genome-wide level has been described between chronic inflammatory diseases and psychiatric diseases, respectively. The TWAS approach can be viewed as a transcriptome-wide screening method to test for gene-disease associations from GWAS data sets. By using a recently published principled cross-tissue TWAS method (UTMOST)—simultaneously training expression imputation models across multiple disease-relevant tissues and performing cross-tissue gene-level conditioned association tests—in combination with GWAS/TWAS conditional analysis for cross-phenotype studies we were able to address several types of confounding factors that have previously compromised the validity of TWAS studies for complex diseases: (1) poor TWAS prediction accuracy due to imputation from reference expression panels containing tissues not relevant to disease, (2) a bias in the number of transcriptome-wide significant gene-disease associations due to different sample sizes of reference data sets in expression imputation models; (3) an excess of TWAS association signals at loci with multiple TWAS signals due to correlated expression among genes of the same locus, and (4) an inaccurate determination of association boundaries for TWAS signals based on SNP information.

The objectives of this cross-tissue, cross-phenotype TWAS were to (i) evaluate the effectiveness of cross-tissue imputation models using ten GWAS consortium studies of psychiatric and inflammatory disease; (ii) identify transcriptome-wide significant gene-disease associations in GBA tissues for each disease and localize the most strongly associated genes through cross-tissue and multiple-gene-conditioned fine-mapping analyses; (iii) quantify the overall component of (shared) genetically regulated expression changes ($\rho_e$) for immune-mediated and psychiatric diseases in relation to genome-wide genetic correlation ($\rho_g$); (iv) identify comorbidities and causal relationships using population-based administrative health data from the entire Danish population and by performing Mendelian Randomization (MR) analyses; (v) identify susceptibility genes and pathways shared between psychiatric disorders and inflammatory diseases of the gastrointestinal tract using bulk RNA-Seq data of different developmental stages and single-cell RNA-sequencing (scRNA-seq) expression data for tissues of the GBA.

### Results

**Cross-tissue transcriptome-wide association studies (TWAS) for seven diseases of the GBA.** The UTMOST framework was used to perform cross-tissue TWAS for 17,290 genes and 23 selected disease-relevant tissues, which had been identified as disease-relevant for the GBA in a previous study of Finucane and colleagues (Methods), using transcriptome-wide and genome-wide reference data of 4043 samples provided by consortia projects GTeX, STARNET, and BLUEPRINT (Fig. 1 and Supplementary Data 2). A schematic overview of the study workflow is shown in Supplementary Fig. 1. Gene expression imputation models were used using GWAS summary statistics data of ten consortium meta-analysis studies comprising 180,592 independent cases of European ancestry (CD: 21,771; UC: 18,621; PSC: 4,338; SCZ: 35,476; MDD: 59,851; BD: 20,352; ADHD: 20,183; Supplementary Data 1). We then tested for gene-disease associations for a maximum of 17,290 genes and 23 tissues for each of the ten GWAS meta-analysis studies. Single-tissue gene-wise (marginal; i.e. unconditioned) association $P$ values (analysis Single-tissue_marginal) (Supplementary Data 3) were combined into (marginal) joint-tissue gene-wise summary $P$ values (analysis GBJ_marginal, resulting in one $P_{\text{GBJ}_{\text{marginal}}}$ for each gene and disease; Supplementary Data 3) via a generalized Berk–Jones (GBJ) test. The cross-tissue GBJ gene test quantifies gene-disease associations across all tissues of the GBA (separately for each disease), adjusts the correlation between TWAS association statistics for individual tissues, and also provides (compared with standard univariate meta-analysis approaches) a substantial increase in statistical power in situations where eQTL effects are not unique to a single tissue (Methods). The GBJ test identified a total of 586 (277), 345 (179), 214 (79), 361 (104), 38 (21), 77 (10), and 42 (8) gene-disease associations (number of loci with size of ±1 Mb) with transcriptome-wide significance ($P_{\text{GBJ}_{\text{marginal}} < 0.05/15,587 = 3.20 \times 10^{-6}$; Supplementary Fig. 2; corrected for a maximum of 15,587 successfully tested and GBJ-meta-analyzed genes) for CD, UC, PSC, SCZ, MDD, BD, and ADHD, respectively (Supplementary Data 3). To evaluate a possible bias in the number of transcriptome-wide significant gene-disease associations due to different sample sizes of reference data sets, we further conducted TWAS analyses using expression imputation models provided by S-PrediXcan and FUSION (Methods). We observed on average only a moderate positive correlation between reference sample size and the number of significant genes per tissue in our cross-tissue TWAS results (average Spearman’s $r = 0.46$; Supplementary Data 4 and Supplementary Fig. 3). This suggests that cross-tissue TWAS imputation models can effectively select predictive cis eQTL variants (within ±1 Mb of the transcription start/end site of a gene) shared by multiple tissues, whereas tissue-specific eQTLs with strong effects were retained.

**Cross-tissue and multiple-gene-conditioned fine-mapping analyses at loci with multiple-gene-disease association signals.** $P_{\text{GBJ}_{\text{marginal}}}$ values from cross-tissue TWAS analyses (Supplementary Data 3) showed weak to moderate inflation of GBJ association statistics ($\lambda_{1000}$ in the range [0.94; 1.05]; Supplementary Data 5). One reason for this may be over-regulation of multiple genes by the same eQTL, as well as LD between eQTLs with nonzero prediction weights, which may lead to correlated predicted expression between nearby genes and thus multiple correlated TWAS hits per locus, analogous to the situation in a GWAS study where an association
signal typically spans a large number of highly significant SNPs in high LD with the causal SNP. In order to distinguish between causal and marker gene-disease associations at significant TWAS loci, we performed (i) multiple-gene-conditioned single-tissue analysis (Single-tissueconditional) followed by (ii) multiple-gene-conditioned cross-tissue analysis (GBJconditional) (Methods) with nearby genes within range of ±1 Mb around transcriptome-wide significant genes from marginal results (genes highlighted as red dots in the Manhattan plots in Supplementary Fig. 2). This accounted for potential cross-tissue co-regulation and LD between nearby genes within range of ±1 Mb around transcriptome-wide significant SNPs, and furthermore, the marginal association result of nearby genes was conditioned based on all genes within the ±1 Mb region (avoiding specification of fixed locus boundaries based on LD blocks from genotype data). This analysis yielded a total of 215 significant gene-disease associations at genome-wide significance (Pgenetic < 5.0 × 10⁻⁸) for CD, UC, PSC, SCZ, BD, ADHD, respectively (Supplementary Data 3 and Supplementary Data 6, Methods), thus representing final transcriptome-wide significant cross-tissue multiple-gene-conditioned TWAS gene-disease association results (Table 1 and Fig. 2a) within and outside the boundaries of established GWAS susceptibility loci (Supplementary Data 7).

Comorbidity analysis and trait correlation at the genetic level and the level of predicted gene expressions. Using population-based administrative health records from the Danish National Patient Registry (DNPR), we conducted a retrospective cohort study for the period 1994 to 2018 by examining directed diagnosis pairs of 7,191,385 people from the entire Danish population (Methods). We confirmed statistically significant comorbidity (false discovery rate FDR < 0.05) between our four psychiatric and three immunological phenotypes for nine psychiatric-immunological pairs out of 42 disease pairs (Supplementary Data 6) with an increased incidence of SCZ (relative risk RR = 1.15), BP (RR = 1.28) and depression (RR = 1.52) in CD, depression in UC (RR = 1.34) and PSC (RR = 1.42), CD in SCZ (RR = 1.12), BD (RR = 1.29) and depression (RR = 1.57), UC (RR = 1.37) and PSC (RR = 1.57) in depression. We further applied pairwise genetic correlation analysis with LDSC as described in Tylee et al. (Methods) to complement TWAS analyses with up-to-date values of genetic correlations (Supplementary Data 9 and Supplementary Fig. 4).

Next, to examine pairwise trait correlations at the level of predicted (genetically regulated) gene expression (ρE) and to compare ρE with genome-wide genetic correlations (ρG), we calculated Spearman’s correlations between TWAS gene effect estimates for all pairs of diseases by using the gene-wise Z-scores from the analyses Single-tissueconditional marginal and GBImarginal and a correlation-pruned list of 1825 co-regulation- and eQTL-independent genes previously used in Mendelian randomization studies (to avoid measuring correlated predicted expression on the transcriptome-wide scale) (Methods). As an example (Single-tissueconditional marginal: Supplementary Data 10), the strongest positive correlation across psychiatric and immune phenotypes was observed for CD4+ T-cells (ρE = 0.11 for CD and BP and
Table 1 Transcriptome-wide significant (\(P_{\text{conditional}} < 3.20 \times 10^{-6}\)) genes from multiple-gene-conditioned fine-mapping analysis at loci with multiple-gene-disease association signals, for psychiatric and immune phenotypes in 23 gut-brain-axis (GBA) tissues.

| Disease    | UC       | SCZ      | CD       | Crohn    | UC       | CD       | CD       |
|------------|----------|----------|----------|----------|----------|----------|----------|
| TWAS discovery n unique genes across 23 tissues | 215 (81) | 150 (122) | 150 (122) | 150 (122) | 150 (122) | 150 (122) | 150 (122) |
| n unique genes in total from single tissues | 242 (240) | 197 (163) | 164 (164) | 164 (164) | 164 (164) | 164 (164) | 164 (164) |
| n unique genes in cis-GWAS loci | 168 (164) | 164 (164) | 164 (164) | 164 (164) | 164 (164) | 164 (164) | 164 (164) |
| n unique genes non-GWAS loci | 45 (46) | 46 (46) | 46 (46) | 46 (46) | 46 (46) | 46 (46) | 46 (46) |

For the selection of 23 tissues, we see Supplementary Data 2 and Methods. \(P_{\text{conditional}} \leq 3.20 \times 10^{-6}\) significantly associated with multiple-gene-disease association signals (analysis GB1conditional; see Methods). The number of associated loci varies from 1 to 3.6 Mb. The number of red dots in Fig. 2a corresponds to the number of significant genes (GB1marginal; Supplementary Data 3) from another brain or non-brain tissue (brain tissue if the expression heritability estimates \(h_{\text{med}}^2\)) of (13.4%, 18.2%) for NR5A2.

Mendelian randomization analysis for psychiatric and immune phenotypes. Mendelian randomization (MR) is a commonly used method to assess a causal influence of one trait (exposure) on another (outcome). To identify putative pairs of immune-mediated and psychiatric diseases that may be related in a causal manner (i.e., vertical pleiotropy\(^{29}\)), we conducted MR analysis for transcriptome-wide significant lead genes (i.e., our instruments with \(P_{\text{conditional}} < 3.20 \times 10^{-6}\) and filtered for the most-significant genes from Single-tissue\(_{\text{conditional}}\) within ±1 Mb regions) using multi-gene TWAS MR analysis (including HEIDI-outlier filtering that removes instruments with strong putative horizontal pleiotropic effects, where one gene has independent effects on multiple traits) as implemented in GSNMR\(^{20}\) (Methods). After Bonferroni correction for 171 pairwise tests (\(P_{\text{GSNMR}} < 2.92 \times 10^{-4}\)), we identified CD (\(P_{\text{GSNMR}} = 9.58 \times 10^{-5};\) OR = 1.10, 95% CI 1.05–1.15) and UC (\(P_{\text{GSNMR}} = 8.40 \times 10^{-6};\) OR = 1.17, 95% CI 1.09–1.26) as a putative risk factor (exposure) for SCZ in tissues “Colon_- Transverse” and “Esophagus_Gastroesophageal_Junction”, respectively (Supplementary Data 11). HEIDI-filtering identified CSNK1A1 and SGSM3 as potential horizontally pleiotropic genes, suggesting a pleiotropic function in both diseases.

Identification of gene-disease associations shared between psychiatric and immune phenotypes. After identifying pairs of immune-mediated and psychiatric diseases as putative correlated traits on the transcriptome-wide level, we sought to identify susceptibility genes shared between psychiatric and immune phenotypes from the list of all transcriptome-wide significant genes from both analyses Single-tissue\(_{\text{conditional}}\) and GB1\(_{\text{conditional}}\) (Table 1). Out of 288 possible pairwise tissue-specific combinations of psychiatric and immune phenotypes (Supplementary Data 12), gene overlap analysis revealed 31 (of those 4 GB1\(_{\text{conditional}}\)) pairs of psychiatric and immune phenotypes that share at least one susceptibility gene. After excluding genes within the major histocompatibility complex (MHC; chr6: 25–34 Mb), which was particularly relevant to pairs PSC-MDD and PSC-SCZ in various tissues, we identified three non-MHC TWAS susceptibility genes (NR5A2, SATB2, PPP3CA) from analyses Single-tissue\(_{\text{conditional}}\) to be shared between brain (SCZ, CD, UC: Hypothalamus; SCZ, UC: Frontal cortex BA9; SCZ, CD: Putamen basal ganglia) and intestinal tissues (CD, UC: Colon transversus; UC: Colon sigmoidum; CD: Colon transversus) (Table 2; more detailed results in Supplementary Data 3). In summary, these three genes met the transcriptome-wide significance threshold of \(3.20 \times 10^{-6}\) in the same brain or non-brain tissue for at least one immune disease and one psychiatric disease, where at least one of the diseases must also show another transcriptome-wide significant signal for another brain or non-brain tissue (brain tissue if the first signal occurred in a non-brain tissue or vice versa), with substantial gene expression heritability estimates \(h_{\text{med}}^2\) (i.e., heritability mediated by the cis genetic component of gene expression levels) across the 23 tissues (\(h_{\text{med-min}}^2\), \(h_{\text{med-max}}^2\)) of (13.4%, 18.2%) for NR5A2.
(6.9%, 24.8%) for SATB2, and (5.0%, 10.9%) for PPP3CA (Table 2; Methods). Thus, these genes met the most stringent criteria for shared susceptibility genes, as they were shared between psychiatric and immune phenotypes in the same tissues and across the brain and non-brain tissues and were therefore prioritized below. We observed that an increase (decrease) in predicted expression for

For the selection of 23 tissues, see Fig. 1, Supplementary Data 2, and Methods. Results from unconditioned cross-tissue TWAS analysis (GBJ marginal; before multiple-gene-conditioned corresponding hierarchical single linkage clustering across diseases. Genetic trait correlations for single-tissue results (analysis Single-tissue conditional) are available in Supplementary Data 9, 10. For the selection of 23 tissues, see Fig. 1, Supplementary Data 2, and Methods. Results from unconditioned cross-tissue TWAS analysis (GBJ marginal; before multiple-gene-conditioned fine-mapping analysis) are shown in Supplementary Fig. 2.

GWAS/TWAS conditional analysis to describe the GWAS contribution to the TWAS result. To estimate the contribution of the three shared same-tissue susceptibility genes NR5A2, SATB2, and PPP3CA relative to the established GWAS signals as defined by the recent consortia fine-mapping studies for SCZ [31] and CD/UC [32,33], we performed a joint GWAS/TWAS fine-mapping analysis across the 23 tissues of the GBA using the GWAS consortium summary statistics listed in Supplementary Data 1. For NR5A2, SATB2, and PPP3CA, respectively, we performed a multi-SNP-based conditional and joint association analysis using GCTA COJO [34] to condition GWAS summary statistics on independent eQTL effects of the three genes (Methods). We examined how much of the GWAS fine-mapping signal remained after the signal of the TWAS gene eQTLs was removed. We demonstrated the applicability of our approach by examining and replicating established eQTL associations of GWAS signals as reported in the fine-mapping GWAS studies for CD/UC (Supplementary Note 1 and Supplementary Figs. 6, 7; detailed results showing all genes from the regions around NR5A2, SATB2, PPP3CA, INO80E, SF3B1, SGSM3, and ZC3H7B are likely to be proxies for other gene-disease associations at the respective loci.)

(6.9%, 24.8%) for SATB2, and (5.0%, 10.9%) for PPP3CA (Table 2; Methods). Thus, these genes met the most stringent criteria for shared susceptibility genes, as they were shared between psychiatric and immune phenotypes in the same tissues and across the brain and non-brain tissues and were therefore prioritized below. We observed that an increase (decrease) in predicted expression for NR5A2 and SATB2 (PPP3CA) was associated with an increased risk of SCZ, CD, and/or UC (Table 2). In addition, four other gene candidates (INO80E, SF3B1, SGSM3, and ZC3H7B) from the secondary analysis (GBJ marginal) met the transcriptome-wide significance threshold and showed associations across both brain and intestinal tissues, but each in different tissues for psychiatric and immune phenotypes, implying that for these four genes the most significant tissues differ between psychiatric and immune phenotypes (Supplementary Data 12–14). This suggests that those four genes may be involved in SCZ and CD/UC, but not necessarily in the same tissue. We will show below by gene set enrichment analysis that INO80E, SGSM3, and ZC3H7B are likely to be proxies for other gene-disease associations at the respective loci.)

GWAS/TWAS conditional analysis to describe the GWAS contribution to the TWAS result. To estimate the contribution of the three shared same-tissue susceptibility genes NR5A2, SATB2, and PPP3CA relative to the established GWAS signals as defined by the recent consortia fine-mapping studies for SCZ [31] and CD/UC [32,33], we performed a joint GWAS/TWAS fine-mapping analysis across the 23 tissues of the GBA using the GWAS consortium summary statistics listed in Supplementary Data 1. For NR5A2, SATB2, and PPP3CA, respectively, we performed a multi-SNP-based conditional and joint association analysis using GCTA COJO [34] to condition GWAS summary statistics on independent eQTL effects of the three genes (Methods). We examined how much of the GWAS fine-mapping signal remained after the signal of the TWAS gene eQTLs was removed. We demonstrated the applicability of our approach by examining and replicating established eQTL associations of GWAS signals as reported in the fine-mapping GWAS studies for CD/UC (Supplementary Note 1 and Supplementary Figs. 6, 7; detailed results showing all genes from the regions around NR5A2, SATB2, PPP3CA, INO80E, SF3B1, SGSM3, and ZC3H7B are likely to be proxies for other gene-disease associations at the respective loci.)

Fig. 2 Manhattan plots showing transcriptome-wide significant (\(P_{\text{conditional}} < 3.20 \times 10^{-6}\) from GBJ conditional) gene discoveries for psychiatric and immune phenotypes in 23 gut-brain-axis (GBA) tissues, and proportion of shared predicted transcriptome-wide disease-associated gene expression (\(\rho_G\)) for pairs of immune-mediated and psychiatric diseases averaged across 23 tissues of the GBA. a Gray dotted line, transcriptome-wide significance threshold; red dots, lead gene-disease associations at TWAS loci (for regions of size \(\pm 1\) Mb) and conditioned on all genes within \(\pm 1\) Mb regions. CD Crohn’s disease, UC ulcerative colitis, PSC primary sclerosing cholangitis, SCZ schizophrenia, MDD major depressive disorder, BIP bipolar disease, ADHD attention-deficit/hyperactivity disorder. b Spearman’s trait correlation (\(\rho\)) values at the transcriptome-wide level of gene expression for each disease pair and averaged across 23 tissues of the GBA show that effects of genetic risk variants shared between psychiatric and immune phenotypes are largely mediated by gene expression. Strongest positive correlations across psychiatric and immune phenotypes were observed for pairs UC and BP (\(\rho_G = 0.15\), CD and BP (\(\rho_G = 0.14\), UC and SCZ (\(\rho_G = 0.13\), and CD and SCZ (\(\rho_G = 0.11\), which is in line with genome-wide genetic correlations (\(\rho_{GE}\)) for pairs UC and BP (\(\rho_{GE} = 0.23\), CD and BP (\(\rho_{GE} = 0.22\), UC and SCZ (\(\rho_{GE} = 0.14\), and CD and SCZ (\(\rho_{GE} = 0.12\)) reported by Tylee et al. [11]. The dendrogram reflects the corresponding hierarchical single linkage clustering across diseases. Genetic trait correlations for single-tissue results (analysis Single-tissue conditional) are given in Supplementary Data 9, 10. For the selection of 23 tissues, see Fig. 1, Supplementary Data 2, and Methods. Results from unconditioned cross-tissue TWAS analysis (GBJ marginal; before multiple-gene-conditioned fine-mapping analysis) are shown in Supplementary Fig. 2.
and all significant tissues are described in the Supplementary Note 2 and Supplementary Figs. 8–37). Joint GWAS/TWAS fine-mapping results for the additional gene candidates (IN080E, SF3B1, SGSM3, and ZC3H7B) from analyses GBJconditional are discussed in Supplementary Fig. 38–41 and Supplementary Note 4.

Increased genetically regulated expression of NR5A2 in the hypothalamus and colon transversum was associated with increased risk for SCZ, CD, and UC. We found that the TWAS association of NR5A2 at 1q32.1 in the hypothalamus and colon transversum (Fig. 3a) was driven by the nearby (nonoverlapping) established SCZ and CD/UC GWAS signals located 104 and 654 kilobases (kb) upstream of NR5A2, respectively. This GWAS locus was not assigned a gene in the CD/UC consortium fine-mapping study35 but was assigned an overlap with an epigenetic mark for immune cells (Immune_H3K4me1). Given the functional evidence for NR5A2 in attenuating inflammatory damage in murine and human intestinal organoids36 (see also Supplementary Note 3) we propose NR5A2 as the best candidate gene for the GWAS locus 1q32.1 with a possible effect on the hypothalamic–pituitary–adrenal axis (HPA axis)10.

Increased genetically regulated expression of SATB2 in frontal cortex BA9 and colon sigmoideum was associated with increased risk of SCZ and UC. The TWAS association of SATB2 at 2q33.1 (Fig. 3b) is driven by the nearby (nonoverlapping) established SCZ and UC GWAS signals located 67 and 425 kilobases (kb) downstream and upstream of SATB2, respectively. While multiple TWAS eQTLs were identified for SATB2 in the 95% credible set of the UC GWAS peak (consisting of intergenic variants between LOC101927619 and SATB235), it is convincing that the GWAS/TWAS fine-mapping analysis revealed a possible explanation of the GWAS signal by eQTLs at this locus. SATB2 as a potential susceptibility gene for UC and SCZ has already been implicated by other studies (Supplementary Note 3).

Decreased genetically regulated expression of PPP3CA in the putamen basal ganglia and colon transversum was associated with increased risk of SCZ and CD. We found that the TWAS association of PPP3CA at 4q24 (Fig. 3c) is driven by the nearby (nonoverlapping) established SCZ and CD GWAS signals located 280 and 384 kilobases (kb) upstream of PPP3CA, respectively. GWAS fine-mapping analysis33 assigned the 95% credible set of fine-mapped SNP variants (n = 170 SNPs) to the gene BANK1, which is closest to the lead GWAS SNP variant at 4q24. Because PPP3CA encodes a subunit of calcineurin (Supplementary Note 3) that has been associated with SCZ and is also a known drug target (due to a direct physical interaction with FKBP Prolyl Isomerase 1 A (FKBP1A)38, also known as FKBP12) for the treatment of UC via calcineurin inhibitors, we propose PPP3CA as the best candidate gene for the GWAS association signal at 4q24.

Analysis of bulk RNA-Seq data of different developmental stages and scRNA-seq expression data for tissues of the GBA. Having tested the three gene-disease associations for plausibility at the GWAS level in the previous section, we hypothesized that high, tissue- or cell-type-specific expression would be indicative of a particular function in the tissue of interest, and that expression of a gene could also occur only in a small subset of cells of only a particular tissue or at a particular human developmental stage, which is hypothesized for SCZ37. We used publicly available non-diseased bulk tissue38,39 and scRNA-Seq data sets40,41 (Methods) to look up in which tissues and at which developmental stage of human tissues and cells the shared susceptibility genes are generally expressed, and found that the observed patterns matched the existing knowledge on the genes (Supplementary Note 3). It is also possible that an adverse event
in utero can interfere with normal brain development and create a brain vulnerability that, in an individual already at risk (genetic predisposition), can lead to SCZ later in life\(^3\). We, therefore, examined the expression of our three candidate genes in developmental data from brain\(^3\) and intestine\(^3\), since it is conceivable that a susceptibility gene is expressed during fetal development but is no longer expressed in adulthood. Again, NR5A2 was almost absent in bulk brain samples from different developmental stages, whereas SATB2 was strongly upregulated specifically in the fetal forebrain during midfetal development, and PPP3CA was weakly expressed in all developmental stages of the brain\(^3\) (Fig. 4a). In the intestine, it was the opposite: NR5A2 was strongly expressed in the ileum and colon during the fetal and juvenile stages, whereas SATB2 was weakly expressed in the colon sigmoidum and not ileum, and PPP3CA almost not at all\(^3\) (Fig. 4a). To achieve resolution at the single-cell level, we...
analyzed human single-nucleus and scRNA-seq data from different cell types of the adult brain and gastrointestinal tract, respectively (Fig. 4b). In these data sets, NR5A2 is expressed in the ileal progenitor, stem, and transient amplifying (TA) cells, less strongly in the colon and rectum, and not in the brain, consistent with the bulk expression data. Consistent with the bulk expression from developmental stages of the intestine (Fig. 4a), SATB2 was strongly expressed in cell types of the colon and rectum and barely expressed in brain cells. PPP3CA expression was highly enriched in enterendocrine and Paneth-like cells of the ileum, colon, and rectum despite weak expression in the bulk RNA-Seq data of ileum or colon, which may imply a specific role of PPP3CA in immune modulation in the gut for CD and UC; in the brain, PPP3CA is also expressed in neurons (Fig. 4b). Bulk tissue and scRNA-seq results for genes INO80E, SF3B1, SGSM3, and ZC3H7B are shown in Supplementary Fig. 42 and discussed in Supplementary Note 5.

Calcineurin-dependent NFAT and Wnt signaling as shared signaling pathways for IBD and SCZ. Based on the statistically unambiguous gene prioritization from multiple-gene-conditioned fine-mapping analyses, it is likely that NR5A2, SATB2, and PPP3CA are shared susceptibility genes for both IBD and SCZ. The genes INO80E, SGSM3, and ZC3H7B from the secondary and not so stringent gene overlap analysis (GBJconditional) might be proxies for other genes with strongly correlating gene expression at the same locus (except for SF3B1, Supplementary Figs. 38–42), where the gene-disease association signal of the eleven nearby genes of SF3B1 clearly disappeared after conditioning, see Supplementary Figs. 8–37, because these three loci showed an extremely high gene density with 45 (INO80E), 16 (SGSM3), and 29 (ZC3H7B) genes (Supplementary Figs. 8–37) making the multiple-gene-conditioned fine-mapping analysis difficult. We assumed that these genes, which according to previous findings have nothing to do with IBD or SCZ (Supplementary Note 4), may be masking the causative gene of the locus. To test whether any of the secondary gene candidates could be linked to NR5A2, SATB2, and PPP3CA, we performed gene set enrichment analysis as implemented in EnrichR using the NCATS BioPlanet pathway resource of 1658 curated human pathways and 27 input genes including our seven candidate genes (NR5A2, SATB2, PPP3CA, SF3B1, INO80E, SGSM3, and ZC3H7B; Supplementary Data 13), 19 correlated transcriptome-wide significant candidate genes after multiple-gene conditioned analysis for INO80E, SGSM3 and ZC3H7B, and CNK1A1 (a pleiotropic HEIDI gene outlier for SCZ and UC from MR analysis (Supplementary Data 11) representing a case of horizontal pleiotropy). The strongest associated term was “calcineurin-dependent NFAT signaling role in lymphocytes” (enrichment $P = 8.57 \times 10^{-7}$; Supplementary Data 15) including genes CNK1A1, EP300 (same locus as SGSM3), and MAPK3 (same locus as INO80E) in addition to our main candidate gene PPP3CA. In our view, EP300 and MAPK3 are the more likely candidates instead of SGSM3 and INO80E, respectively. Both genes are important signal transducers and are known to be part of or intersect with the Wnt pathway (enrichment $P = 1.23 \times 10^{-3}$; Supplementary Data 15), which activates NFAT via calcineurin signaling. Interestingly, LRH-1, the gene product of NR5A2 (Supplementary Note 3), has been shown to bind to β-catenin, which is also part of the Wnt pathway. Recent findings also suggest that SCZ and BD are characterized by abnormal Wnt gene expression and plasma protein levels, suggesting that drugs targeting the Wnt signaling pathway may play a role in the treatment of severe mental disorders. Thus, the Wnt signaling pathway and NFAT activation are the most likely candidate pathways for IBD and SCZ, where risk is mediated by gene expression.

Discussion

By performing cross-tissue and multiple-gene conditioned transcriptome-wide association studies (TWAS) for 23 tissues of the GBA for three clinically related inflammatory diseases of the gastrointestinal tract (CD, UC, and PSC) and four diseases of the mind and brain (SCZ, MDD, BD, and ADHD) we identified susceptibility genes NR5A2, SATB2, and PPP3CA shared between IBD (CD/UC) and SCZ that are (genetically) expressed both in intestinal and brain tissues. We estimated the proportion of the genetically regulated component of expression ($h_{\text{med-max2}}^2$) with respect to total gene expression, measured across a maximum of 23 different tissues, to be (13.4%, 18.2%) for NR5A2, (6.9%, 24.8%) for SATB2, and (5.0%, 10.9%) for PPP3CA. These three genes could represent a molecular link between psychiatric and inflammatory traits and may contribute to the observed genetic correlations and increased prevalence of SCZ in CD/UC patients at the population level. The correlations between BD and CD/UC are ematic. Tylet et al. find BD and CD/UC correlated on the genetic level, but they cite a study in which BD-UC was not significant and we find...
no significance, too. Contrarily, we find BD and CD/UC correlated on the genetic expression level, but no single gene overlap hits. A possible explanation is that Tylee et al used a different dataset for BD. BD is closely related to SCZ and depending on the inclusion criteria, might appear more or less similar to SCZ.

Our bioinformatic cross-tissue TWAS pipeline for selected tissues of GBA addressed typical technical drawbacks of previous TWAS studies and enabled conditional analysis of multiple disease associations at TWAS loci. In our opinion, TWAS analyses for CD, UC, and SCZ had the highest statistical power to detect shared TWAS susceptibility genes, as specifically for CD, UC, and SCZ a much higher proportion of expression-mediated heritability ($h_{med}^2/h_{g}^2$) has recently been estimated across GTEx tissues than for other psychiatric disorders such as BD\(^4\), so our TWAS analyses for BD had lower statistical power. Since $h_{med}^2/h_{g}^2$ can vary considerably for different traits or diseases, we recommend that future TWAS studies additionally estimate $h_{med}^2/h_{g}^2$ (e.g., with the MESC software\(^4\)) to assess whether a TWAS can be expected to have a low or rather high number of gene-disease associations for the disease under study. Another key outcome of our GWAS/TWAS conditional analysis for these three genes is that the expression-disease associations largely explain the previously fine-mapped GWAS association signals at nearby GWAS susceptibility loci for SCZ\(^31\), CD, and UC\(^32,33\), suggesting that risk is mediated by genetically regulated expression at those loci. Together with the results of our gene set enrichment pathway analysis, we showed it...
is likely that genetic variation mediated by gene expression in the Wnt signaling pathway and NFAT activation by calcineurin is associated with both SCZ and IBD. One limitation is that we could only use candidate genes from six TWAS loci for pathway analysis, so future, more powerful TWAS are expected to provide a more detailed picture here. Another limiting factor of our study is that the heritability of our three (conservatively) selected gene-disease associations accounts for only a small fraction of the estimated $h^2$ and may present only the tip of the iceberg. We avoided further multiple testing correction for all seven traits, as this would be counterproductive for TWAS screening after Bonferroni correction for the number of all gene-disease associations (although not statistically independent) and conditional analysis for nearby genes at suggestive loci. In addition, the proportion of shared $h^2$ between diseases is difficult to quantify and requires the development of bivariate (for pairs of diseases) methods to estimate the shared $h^2$.

A very interesting observation is the gene-disease association with PPP3CA. PPP3CA encodes a subunit of calcineurin (the catalytic subunit Calcineurin A). Calcineurin is a Ca$^{2+}$-calmodulin-regulated serine/threonine protein phosphatase, which activates the transcription factor NFAT (nuclear factor of activated T-cells) and thus plays a key role in the regulation of the signal transduction for immune modulation in Paneth and enteroid cells. NR5A2 (encoding human LRH-1) regulates glucocorticoid synthesis and is known to protect epithelial integrity and attenuate inflammatory damage in murine and human intestinal organoids, including those derived from IBD patients (Supplementary Note 3). Although NR5A2 appears to be nearly absent from the brain and blood based on the present data, there is evidence of its function in brain tissues (Supplementary Note 3), so we suspect that NR5A2 expression is restricted to less-studied regions of the brain and that it has separate functions in the gut and brain. SATB2 is a promising candidate for a GBA susceptibility gene because it is predominantly expressed in the colon sigmoideum and prefrontal cortex, consistent with our TWAS results in which SATB2 expression is associated with SCZ and UC but not with CD. SATB2 is also a prognostic marker in UC-associated colorectal cancer (Supplementary Note 3). In summary, gene expression analyses for NR5A2, SATB2, and PPP3CA using intestinal and brain tissue data from different developmental stages and from scRNA-seq studies support the findings from previous human, mouse, and organoid studies and suggest a pleiotropic role for these genes as part of the GBA.

**Methods**

**Consortium GWAS summary statistics.** Crohn's disease (CD) and ulcerative colitis (UC) case and control cohorts (Supplementary Data 1) from 15 countries...
across Europe, North America, and Australia have previously been described, quality controlled (QCed), and genotype imputed\(^1\). Primary sclerosing cholangitis (PSC) case and control cohorts from 14 countries in Europe and North America have previously been described, QCed and imputed\(^2\). The number of overlapping (duplicate) GWAS samples for CD, UC, and PSC GWAS data sets were determined by identity-by-descent (IBD) analysis\(^3\). Because of a partial sample overlap in cases and controls between Immunochip and GWAS data sets for CD, UC, and PSC (Supplementary Data 1), the smaller sample sized GWAS data sets for CD, UC, and PSC (data sets #2, #4, #6 in Supplementary Data 1) were used only for validation of TWAS association results from the TWAS discovery phase (see text below). Schizophrenia (SCZ), major depressive disorder (MDD), bipolar disorder (BD), and attention-deficit/hyperactivity disorder (ADHD) imputed GWAS summary statistics were used as described elsewhere\(^4\) and were downloaded from https://www.med.unc.edu/pgc/download-results/.

Written, informed consent was obtained from all study participants, and the institutional ethical review committees of the participating centers approved all protocols.

Selection of tissues for TWAS of the GBA. Recently, a heritability enrichment analysis of specifically expressed genes across 205 tissues and cell types in combined with GWAS summary statistics of CD, UC, SCZ, MDD, BD, and ADHD and subsequent validation of enrichments results with chromatin data yielded statistically significant enrichment results for tissues of the central nervous system (CNS) for psychiatric diseases and for tissues from blood/immune cell types and the gastrointestinal tract for inflammatory bowel diseases\(^5\). Based on these results, we restricted our TWAS analyses using genome-wide and transcriptome-wide reference data of 4043 samples from brain, immune cell and the gastrointestinal tissues (Supplementary Data 2) available from consortia projects GTEx\(^6\) (version V7, dbGaP accession code phs000424.v7.p2), STARNET\(^7\) (dbGaP accession code phs000120.v1.p1), and BLUEPRINT\(^8\) ftp://ftp.ebi.ac.uk/pub/databases/blueprint/). As described by Hu et al.\(^9\), biallelic SNPs with minor allele frequency (MAF) ≥ 0.01 were retained and normalized gene expression information was adjusted to correct for potentially confounding effects of sex, sequencing platform, and the three main principal components from principal component analysis (PCA) of the genome-wide genotype data, and with an estimation of expression residuals (PEER) factors\(^9\).

UTMOST expression imputation models. To estimate the genetically regulated expression of each gene in the genome across 23 tissues of the GBA, we applied multivariate-response penalized regression models as implemented in UTMOST (https://github.com/laker-jerome/UTMOST) to predict cross-tissue gene expression from reference data sets. Briefly, a standard least-squares loss function was minimized, with an l1 Lasso penalty adaptively set on sample sizes to select predictive SNP variables in a range of ±1 Mb around each gene and force shrinkage in effect size estimation (within-tissue effects), and further with a group-Lasso penalty set across multiple tissues on the effect sizes of each SNP to select eQTLs shared between tissues (cross-tissue effects)\(^10\). Such cross-tissue models for expression imputation favor eQTLs that are shared across tissues while preserving tissue-specific effects, and UTMOST’s approximation procedure proposed for coordinate descent iteration handles incomplete data in the reference expression matrix.

Association testing and GBJ test. UTMOST was further used to test for gene-based association at the level of gene expression regulation, taking into account cross-tissue regulatory effects. A univariate regression model was applied to test the association between predicted gene expression (i.e., the genetically regulated component of gene expression) and disease status for each tissue separately (although previously trained over several tissues at the same time). To summarize the tissue-specific results and test whether a gene is significant for at least one tissue, a generalized Berk–Jones (GBJ) test\(^11\) was used to combine associations across individual tissues into a unified test of association. The GBJ is based on the absolute values of gene trait Z-score, allowing for directions of eQTL effects in the different tissues, and uses tissue covariance to correctly weight results across tissues and correct for multiple tissue testing. We used the Bonferroni correction on the number of genes tested to determine transcriptome-wide significance (P < 0.05/15,587 = 3.20 × 10^{-5}). For the CD, UC, and PSC phenotypes, it had to hold additionally that transcriptome-wide significant genes replicate with P < 0.05 in the smaller GWAS data sets #2, #4, #6 in Supplementary Data 1.

Comparison with other pre-trained single-tissue expression imputation models. The elastic net approach from S-PrediXcan\(^12\) was used for expression imputation of 19 GTEx tissues (Supplementary Data 2) followed by gene-disease association testing using logistic regression for our ten study data sets (Supplementary Data 1). The precomputed reference based on GTEx v7 data was available at http://predictdb.org/. In addition, FUSION\(^13\) Bayesian linear mixed model was used for expression imputation and gene-disease association testing with precomputed weights derived from the same 19 GTEx tissues v7 (http://gusevlab.org/projects/fusion/). The number of Single-tissue marginal significant genes (P < 3.20 × 10^{-5}) from UTMOST, FUSION, and S-PrediXcan for the 19 GTEx tissues and the 10 GWAS studies included in this study were compared with the number of reference samples used for gene expression imputation training. The number of reference samples used in expression imputation training was taken from the supplementary information or the databases of the respective publications; it is possible that the number of GTEx v7 reference samples differs for the different tools due to different pre-filtering analyses. For all tools, only the marginal, uncorrected, P values were considered for benchmark purposes. Spearman correlation and a simple linear regression model were computed separately for each of the ten studies. A paired Mann–Whitney U-test was applied to test for differences in slopes (β) between the three tools.

Cross-tissue multiple-locus-genes conditional analysis and GBJ test. We used the UTMOST multiple regression approach to perform multiple-locus-genes conditional cross-tissue analysis for each transcriptome-wide significant gene (from unconditioned analysis) in each of the ten studies separately to prioritize gene-disease associations at the same locus within the TWAS locus boundaries of the
±1 Mb range. The conditional analysis accounts for one of the main problems of the TWAS gene-based association test, namely the potential co- 

regulation of multiple genes by the same eQTLs or the presence of independent eQTLs across genes that are in linkage disequilibrium (LD)15. Again, the association statistics of the conditional analysis were combined across tissues using the generalized Berk–Jones score (GBJ) test described above. Again, we used the Bonferroni cor- 

rection on the number of genes tested to determine transcriptome-wide sig- 

nificance (P < 0.05/15,587 = 3.20 × 10−4).

Overlap of disease associations with the boundaries of known sus- 
ceptibility loci from genome-wide association studies (GWAS) for CD, UC, 

PSC, SCZ, MDD, BD, ADHD. For genomic position comparison, locus boundaries of established GWAS loci were adopted from the last 

f-mapping GWAS studies for SCZ27 and CD143. A gene was labeled “Overlap with GWAS locus” in Table 2 if the gene had an Locuszoom overlap with GWAS loci. Gene coor- 

dinates (GRCh37, Jan 2020) were obtained using the Ensembl database and 

BioMart59.

Comorbidity analysis in the Danish National Patient Registry (DNPR). To determine significant co-occurrences (disease pairs) for the seven diseases under study, we screened an independent dataset covering ICD-10 diagnosis codes from 7,191,385 people of the entire Danish population in the period from 1994 to 201871. The DNPR includes primary and secondary diagnoses coded according to the International Statistical Classification of Diseases 10th Revision (ICD-10) from all Danish hospitals. We identified patients diagnosed with one of the seven dis- 

theses, for example, we identified a random control set of 20 controls, matched on same-sex and year of birth for each patient diagnosed with CD. Since some of the psychiatric disorders we studied have an average prevalence of more than 1% in the general population, which is especially true for depression and ADHD, we needed a sufficiently high number of controls per case to avoid a significant number of cases occurring by chance in the control group. However, the gain in statistical power, in general, decreases rapidly beyond 4 to 20 controls per 

case2. Therefore, we have selected 20 controls per case so that the calculation does not take too long and because additional controls would not result in a significant gain of statistical power.

The incidence of the six other diseases (X) was calculated for the patients with CD and the randomly matched control group and the strength of the association between the diseases was assessed by relative risk (RR) using equation 1. Here, A is the number of CD patients also diagnosed with disease X, while A is the total number of CD patients. Likewise, C is the number of controls diagnosed with disease X and C is the total number of control patients.

\[
\text{Relative risk} = \frac{A}{C} / \frac{A}{C}
\]

P values for all disease pairs were calculated using a two-sided Fisher exact test and corrected for multiple testing using the false discovery rate (FDR). This analysis is repeated for all seven diseases under study and A as RR as well as FDR adjusted P values were reported in Supplementary Data 8.

Trait correlation at the level of GWAS summary statistics (LDSC). We applied LDSC57 genome-wide correlation as described in Tylee et al.13. Shortly, we excluded the extended MHC region and filtered for HapMap3 SNPs with an INFO score of at least 0.9 and a MAF of at least 0.05. The summary stats were subse- 

quently processed by the munge_sumstats.py program of the LDSC software package (https://github.com/bulik/lodsc), and heritabilities and genetic correlation were calculated on the liability scale using population prevalences (Supplementary Data 9).

Trait correlation at the level of predicted gene expressions. Genes are often co- 

regulated, i.e. they either share the same eQTLs or eQTLs exist in strong LD. For this reason, we used a subset of 1825 co-regulated and eQTL-independent genes recently provided for Mendelian randomization studies38, and further excluded genes from the extended MHC region (chr6:25-34 Mb) to avoid inflated correlation measures. Spearman’s correlation of all traits with each other was calculated separately for each of the 23 tissues based on the single-tissue marginal Z-scores (Supplementary Data 10) and additionally based on the GBJ test marginal test- 

statistics (Supplementary Data 10 and Fig. 2).

Mendelian randomization (MR) analysis at the level of predicted gene expressions. We performed multi-gene MR analysis at the level of predicted gene expressions using GSMR30 (Generalized Summary-data-based Mendelian Rando- 

mization) to test for a causal association between one trait (exposure) on another (outcome) using summary-level TWAS gene association statistics. We used only standardized transcriptome-wide significant marginal UTMOST gene-disease associations as instruments (Pmarginal conditional < 3.2e-6) that were likely indepen- 

dent of neighboring genes by enforcing a minimum distance of 1 Mb between the instrumental genes (see section Cross-tissue multiple-locus-genomes conditional anal- 

ysis). MR analysis was performed on our tissue-specific TWAS results. At least 

10 significant, independent gene-disease associations were necessary as instruments to obtain reliable regression results56. Because the genetically regulated expression is (by definition) heritable and inferred from SNP disease association effects, the MR approach from GSMR was performed with (statistically independent) TWAS genes. This satisfies the requirements of MR analysis according to Bowden et al.74. We further used the HEIDI (heterogeneity in dependent instruments)-outlier test60 to detect and eliminate pleiotropic gene disease associations (genetic instruments) that have apparent pleiotropic effects (horizontal pleiotropy, \(P_{\text{HEIDI}} < 0.01\)) on both traits (exposure and outcome) under investigation.

Gene overlap analysis. To find potential genes mediating the gut-brain axis in the same or different tissues, we considered as candidates all genes that were transcriptome-wide significant for both one inflammatory and one psychiatric trait (i) in the same tissue or (ii) in the GBJ test (both, marginally \((P_{\text{marginal}} = 3.20 × 10^{-4})\) and \((P_{\text{conditional}} < 3.20 × 10^{-10})\) significant) in the same or different tissues, we considered as candidates all genes that were transcriptome-wide significant for both one inflammatory and one psychiatric trait (i) in the same tissue or (ii) in the GBJ test (both, marginally \((P_{\text{marginal}} = 3.20 × 10^{-4})\) and \((P_{\text{conditional}} < 3.20 × 10^{-10})\) significant) in the same or different tissues, we considered as candidates all genes that were transcriptome-wide significant for both one inflammatory and one psychiatric trait (i) in the same tissue or (ii) in the GBJ test (both, marginally \((P_{\text{marginal}} = 3.20 × 10^{-4})\) and \((P_{\text{conditional}} < 3.20 × 10^{-10})\) significant) in the same or different tissues, we considered as candidates all genes that were transcriptome-wide significant for both one inflammatory and one psychiatric trait (i) in the same tissue or (ii) in the GBJ test (both, marginally \((P_{\text{marginal}} = 3.20 × 10^{-4})\) and \((P_{\text{conditional}} < 3.20 × 10^{-10})\) significant) in the same or different tissues, we considered as candidates all genes that were transcriptome-wide significant for both one inflammatory and one psychiatric trait (i) in the same tissue or (ii) in the GBJ test (both, marginally \((P_{\text{marginal}} = 3.20 × 10^{-4})\) and \((P_{\text{conditional}} < 3.20 × 10^{-10})\) significant) in the same or different tissues, we considered as candida...
BLUEPRINT blood cell reference data from 77 mRNA libraries of 27 cell types from infant to adult donors (E-MTAB-3827, https://www.ebi.ac.uk/gxa/experiments/E-MTAB-3827/Downloads). The obtained gene expression values were gene-wise centered and Z-score normalized for visualization in R.

Single-cell RNA-seq (scRNA-seq) data of healthy human small intestine, colon, and rectum biopsies, from two donors each, with a total of 14,537 cells were retrieved from GSE125907 and single-cell libraries from five male postmortem donors (three prefrontal cortex samples, four hippocampus) from the GTEx (Habib et al.; 28846088) were retrieved via https://www.covid19cellatlas.org. Data were processed and log-normalized as described by Sunngak et al.27 and visualized using the scanny v1.6.0 package29 (https://scanny.readthedocs.io/en/stable/api/scanny.pl.dotplot.html).

**NFAT pathway visualization.** The list of 55 genes of the term “calcineurin-dependent NFAT” signaling role in lymphocytes from the database BioPlanet 201936 was downloaded from https://maayyanlab.cloud/Enrichr/#stats36 (May 25, 2021) and used as input for GeneMania31 with no genes added and all available physical interaction edges, pathway-based edges and colocalization edges plotted. The resulting network was exported and imported into Cytoscape 3.8.252 to adjust the visual style. We chose both marginal and conditional nominal significance ($P_{LHS}$ conditional < 0.05) as the threshold for highlighting genes.

**Data availability**

The source data used for this study can be downloaded from the following repositories: The GWAS data can be found as PGC GWAS summary statistics (https://www.med.yale.edu/pgc/download-results) and iBDDGC GWAS summary statistics (ftp://ftp.sanger.ac.uk/pub/consortium/mendelian-genes-to-study-traits/mendelian-genedb-iBDDGC/). Gene ontology enrichment reference data can be found at BLUEPRINT (ftp://ftp.ebi.ac.uk/pub/databases/blueprint/). GTeX, (https://www.gtexportal.org/home/datasets), gene expression atlas, accession numbers E-MTAB-6814 (https://www.ebi.ac.uk/gxa/experiments/E-MTAB-6814/Downloads), E-MTAB-5015, (https://www.ebi.ac.uk/gxa/experiments/E-MTAB-5015/Downloads), E-MTAB-3827, (https://www.ebi.ac.uk/gxa/experiments/E-MTAB-3827/Downloads), and at the COVID-19 cell atlas https://www.covid19cellatlas.org. Gene ontology enrichment reference data of BioPlanet 2019 were downloaded from the Enrichr website https://maayyanlab.cloud/Enrichr/#stats. The data availability of these data is not subject to any restrictions.

**Code availability**

Codes used in this paper are available from websites of the ScanPy package (https://scanny.readthedocs.io/en/stable/api/scanny.pl.dotplot.html), the UTMOST software (https://github.com/Joker-Jerome/UTMOST), the MetaXcan software, (https://github.com/hakynlab/MetaXcan), predictdb/db/, the Fusion software (http://gusevlab.org/projects/fusion/), and the LDXC software (https://github.com/bulk/ldxc). The codes used to perform the analysis of the 1,426,196 significantly associated loci with phenotype traits in the previous GWAS studies, were gene-wise centered and Z-score normalized for visualization in R.

**Received:** 9 July 2021; **Accepted:** 23 December 2021; **Published online:** 20 January 2022

**References**

1. Heiss, C. N. & Olofsson, L. E. The role of the gut microbiota in development, function and disorders of the central nervous system and the enteric nervous system. *J. Neuroendocrinol.* 31, e12684 (2019).
2. Skonieczna-Zydecka, K., Marlicz, W., Misera, A., Koulaozidis, A. & Loniewski, I. Microbiome-The missing link in the gut-brain axis: focus on its role in gastrointestinal and mental health. *J. Clin. Med.* 7, 521 (2018).
3. Kan, J. M., Cowan, C. S. M., Ooi, C. Y. & Kasparian, N. A. What can the gut microbiome teach us about the connections between child physical and mental health? A systematic review. *Dev. Psychobiol.* 61, 700–713 (2019).
4. Ming, X. et al. A gut feeling: a hypothesis of the role of the microbiome in health? A systematic review. *J. Psychosom. Res.* 70, 365–370 (2015).
5. Korchynskyi, O., Lin, B., Wang, T., Huang, M. & Deng, Y. Effect of gut microbiota on host metabolic health. *Front. Pharmacol.* 10, 544 (2019).
6. Hung Di, D. et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* 48, 510–518 (2016).
7. Cross-Disorder Group of the Psychiatric Genomics C, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* 45, 984–989 (2013).
8. Vainberg, M. et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 51, 592–599 (2019).
9. Hu, Y. et al. A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat. Genet.* 51, 568–576 (2019).
10. Korn, E. et al. Investigating the neurodevelopment hypothesis of neuropsychiatric traits. *Nat. Genet.* 51, 933–940 (2019).
11. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* 50, 621–629 (2018).
12. Consortium, G. T. et al. Genetic effects on gene expression across human tissues. *Nature* 550, 204–213 (2017).
13. Franzen, O. et al. Cardiometabolic risk loci share downstream cis- and trans-acting regulation across tissues and diseases. *Science* 335, 827–830 (2016).
14. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* 167, 1398–1414 e1324 (2016).
15. Sun, R., Hsu, S., Bader, G. D., Lin, X. & Kraft, P. Powerful gene set analysis in GWAS with the Generalized Berk-Jones statistic. *PLoS Genet.* 15, e1007530 (2019).
16. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* 9, 1825 (2018).
17. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48, 245–252 (2016).
18. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* 6, 95–108 (2005).
19. Consortium, E. P. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 474, 799–816 (2007).
20. Porci, E. et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat. Commun.* 10, 3300 (2019).
21. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89–R98 (2014).
22. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* 9, 224 (2018).
23. Pardinas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* 50, 381–389 (2018).
24. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* 47, 979–986 (2015).
25. Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* 547, 173–178 (2017).
26. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* 44, 369–375 (2012). *S361–363.
27. Breyer, J. R. et al. LHRH regulates intestinal inflammatory disease by maintaining epithelial homeostasis and cell survival. *Nat. Genet.* 40, 455 (2018).
28. Liu, J. et al. Calcineurin is a common target of cyclophillin-cyclosporin A and FKBP-FK506 complexes. *Cell* 66, 807–815 (1991).
29. Rehn, A. E. & Rees, S. M. Investigating the neurodevelopmental hypothesis of schizophrenia. *Clin. Exp. Pharmacol. Physiol.* 32, 687–696 (2005).
30. Cardoso-Moreira, M. et al. Gene expression across mammalian organ development. *Nature* 571, 505–509 (2019).
31. Kraiczy, J. et al. DNA methylation defines regional identity of human intestinal epithelial organoids and undergoes dynamic changes during differentiation. *Gut* 68, e997 (2019).
32. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* 14, 955–958 (2017).
41. Wang, Y. et al. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. J. Exp. Med. 217, e20191130 (2020).

42. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97 (2016).

43. Huang, R. et al. The NCATs Bio-Platform - An integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems biology, and chemical genomics. Front. Pharmacol. 10, 445 (2019).

44. Racine, M. E. et al. Beta/Catenin interaction regulates adult progenitor cell differentiation downstream of WNT5a/protein kinase C (PKC). J. Biol. Chem. 291, 6559–6582 (2016).

45. Yumoto, F. et al. Structural basis of coactivation of liver receptor homolog-1 by beta-catenin. Proc. Natl Acad. Sci. USA 109, 143–148 (2012).

46. Hoosei, E. Z. et al. Exploring the Wnt signaling pathway in schizophrenia and bipolar disorder. Transl. Psychiatry 8, 55 (2018).

47. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. 48, 709–717 (2016).

48. Yao, D. W., O’Connor, L. J., Price, A. L. & Gusave, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. 52, 626–633 (2020).

49. Hubbard, J. M. & Klee, C. B. Functional domain structure of calcineurin A: mapping by limited proteolysis. Biochemistry 28, 1868–1874 (1989).

50. Yeh, H. & Markmann, J. F. Transplantation: are calcineurin inhibitors safer than mTOR inhibitors? Nat. Rev. Nephrol. 9, 11–13 (2013).

51. Bendickova, K., Tidu, F. & Fric, J. Calcineurin-NFAT signalling in myeloid leukocytes: new prospects and pitfalls in immunosuppressive therapy. EMBO Mol. Med. 9, 990–999 (2017).

52. Matsuoka, K. et al. Tacrolimus for the treatment of ulcerative colitis. Intest. Res. 13, 219–226 (2015).

53. Smith, R. W., M., Morgan, L., Parkes, M. & Lee, J. C. Tacrolimus suppositories: a safe and effective treatment for treatment-refractory proctitis. J. Crohn’s Colitis 13, S408–S409 (2019).

54. Jaeger, S. U. et al. Tacrolimus suppositories in therapy-resistant ulcerative colitis. Inflamm. Intest. Dis. 3, 116–124 (2019).

55. Takao, K. & Miyakawa, T. Investigating gene-to-behavior pathways in schizophrenic and bipolar disorders. Trends Pharmacol. Sci. 36, 279–290 (2015).

56. Niu, Y., Sun, I. W., Liu, S. I., Loh el, W. & Lin, Y. C. Calcineurin ointment-induced relapse of schizophrenia: a case report. Int. J. Neuropsychopharmacol. 10, 851–854 (2007).

57. Heinrich, T. W. & Marcangelo, M. Psychiatric issues in solid organ transplantation. Harv. Rev. Psychiatry 17, 398–406 (2009).

58. Gribble, F. M. & Reimann, F. Mechanisms of enteroendocrine cells and gut hormones in metabolism. Nat. Rev. Endocrinol. 15, 226–237 (2019).

59. Keshav, S. Paneth cells: leukocyte-like mediators of innate immunity in the intestine. J. Leukoc. Biol. 80, 500–508 (2006).

60. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations - Data sets. ftp://ftp.sanger.ac.uk/pub/consortia/ibdgenetics/ibdgc-trans-ancestry-filtered-summary-stats.tgz (2015).

61. Ji, S. G. et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat. Genet. 49, 269–273 (2017).

62. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427 (2014).

63. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681 (2018).

64. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. 51, 793–803 (2019).

65. Duan, X. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. 51, 63–75 (2019).

66. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput. Biol. 6, e1000770 (2010).

67. Siggaard, T. et al. Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million Danish patients. Nat. Commun. 11, 4952 (2020).

68. Pang, D. A relative power table for nested matched case-control studies. Occup. Environ. Med. 56, 67–69 (1999).

69. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 1286–1292 (2015).

70. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int. J. Epidemiol. 44, 512–525 (2015).

71. Witt, J. et al. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 563–569 (2010).

72. Tan, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011).

73. Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330 (2020).

74. Sunnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat. Med. 26, 681–687 (2020).

75. Wolf, F. A., Angerer, P. & Hoech, F. J. SCANDY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018).

76. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128 (2013).

77. Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Nucleic Acids Res. 38, W214–W220 (2010).

78. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the eMed Vernetzungsfonds funding concept (GB-XMAP grants 01ZIK709A, 01ZIK709B, and 01ZIK709C). The study received infrastructure support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Cluster of Excellence 2167 ”Precision Medicine in Chronic Inflammation (PMD)” (EXC 2167-39084018). Access to the DNPR was approved by the Danish Health Data Authority (FSEID-00003092 and FSEID-00004491 and supported by Novo Nordisk Foundation (grants NN1F4CC0001 and NN1F70025794). We thank the PGC for making the GWAS summary statistics publicly available. We thank the IBDDG and IPSCG for permission to use the GWAS/Immunochip summary statistics; a complete list of members and affiliations of the IBDDG and the IPSCG can be found in the Supplementary Information. Figure 5b was generated using BioRender.com and a template from Akiko Iwasaki, Ph.D. and Ruslan Medzhitov, Ph.D.

Author contributions

DE was responsible for the concept and the design of the study. TF, AF, MN, FD, THK, SS, CS, PH, and PK assisted with recruitment or phenotyping of patients in the inflammatory and psychiatric disease clinical cohorts from which the GWAS summary statistics were derived or provided GWAS summary statistics. FU-W and DE implemented statistical models and performed computation analyses. CM, OB, EMW, IFJ, S Be, OW, KB, and S Br. helped with bioinformatic analyses. SJ and FU-W performed RNA-sequencing data analysis. TAS designed Fig. 1. DE and FU-W curated and interpreted results and wrote the manuscript. All authors reviewed, edited, and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03031-6.

Correspondence and requests for materials should be addressed to David Ellinghaus.

Peer review information Communications Biology thanks Eske Derks and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editor: George Inglis. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
