HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2019 November 13.

Published in final edited form as:

Nat Genet. 2019 June ; 51(6): 933–940. doi:10.1038/s41588-019-0409-8.

Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits

Eric R. Gamazon[1,2,3,4,5,*], **Aeilko H. Zwinderman**[1], **Nancy J. Cox**[2,3], **Damiaan Denys**[1], and **Eske M. Derks**[1,6,*]

[1]Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [2]Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA. [3]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. [4]Data Science Institute, Vanderbilt University, Nashville, TN, USA. [5]Clare Hall, University of Cambridge, Cambridge, UK. [6]QIMR Berghofer, Translational Neurogenomics group, Brisbane, Australia.

Abstract

The genetic architecture of psychiatric disorders is characterized by a large number of small-effect variants[1] located primarily in non-coding regions, suggesting that the underlying causal effects may influence disease risk by modulating gene expression[2–4]. We provide comprehensive analyses using transcriptome data from an unprecedented collection of tissues to gain pathophysiological insights into the role of the brain, neuroendocrine factors (adrenal gland) and gastrointestinal systems (colon) in psychiatric disorders. In each tissue, we perform PrediXcan analysis and identify trait-associated genes for schizophrenia (*n*-associations = 499; *n*-unique genes = 275), bipolar disorder (*n*-associations = 17; *n*-unique genes = 13), attention deficit hyperactivity disorder (*n*-associations = 19; *n*-unique genes = 12), and broad depression (*n*-associations = 41; *n*-unique genes = 31). Importantly, both PrediXcan and Summary-data-based Mendelian Randomization/HEIDI analyses propose potentially causal genes in non-brain tissues, showing the utility of these tissues for mapping psychiatric disease genetic predisposition. Our analyses further highlight the importance of joint-tissue approaches as 76% of the genes were detected only in difficult-to-acquire tissues.

Editorial summary:

* egamazon@uchicago.edu, eske.derks@qimrberghofer.edu.au.

Author contributions

E.R.G.: Conception and design of study; analysis and interpretation of the data; drafting the manuscript; approval of the version of the manuscript to be published

A.H.Z.: Interpretation of data; approval of the version of the manuscript to be published

N.J.C.: Revising the manuscript critically for important intellectual content; approval of the version of the manuscript to be published

D.D.: Revising the manuscript critically for important intellectual content; approval of the version of the manuscript to be published

E.M.D.: Conception and design of study; analysis and interpretation of the data; drafting the manuscript; approval of the version of the manuscript to be published

Competing Interests Statement

The authors of this manuscript have no conflicts of interest to disclose.

Multi-tissue transcriptome analyses using PrediXcan identify numerous trait-associated genes for schizophrenia, bipolar disorder, attention deficit hyperactivity disorder, and broad depression, and highlight potentially causal genes in non-brain tissues.

A primary challenge in the field of psychiatric genetics is to understand the neurobiological mechanisms[5,6] underlying genetic predisposition to disease to facilitate detection of additional disease-associated loci as well as to elucidate the precise molecular mechanisms underlying known loci. Integrative transcriptome studies promise to address the functional gap, but to date, analyses of gene expression[7] for psychiatric conditions have been hampered by the relative inaccessibility of the relevant tissues. Here, up to 393 individuals were assayed in 10 distinct brain regions and 3 additional non-brain tissues from the GTEx tissue panel, including a total of 889 brain samples and 633 non-brain samples. We further included 538 dorsolateral prefrontal cortex (DLPFC) samples from the CommonMind Consortium (Supplementary Table 1a). We included both accessible (whole blood) and difficult-to-acquire tissues (i.e., brain, colon selected for its potential involvement in psychiatric disorders based on observed comorbidity with gastrointestinal disorders[8], and adrenal gland known to interact with the central nervous system through the hypothalamic-pituitary-adrenal axis[9]).

Using the largest available genome-wide association study (GWAS) meta-analysis results of five psychiatric traits[3,10–17] (Supplementary Table 1b), we investigated gene-level psychiatric disease associations from a comprehensive PrediXcan analysis[18] using imputation models derived from brain and non-brain transcriptomes. PrediXcan estimates the *Genetically Regulated eXpression* (GReX) for a gene in a given tissue for the GWAS samples and uses GReX to find genes associated with disease risk. We estimated the true positive rate ($\widehat{\pi_1}$) (see Methods) for association with disease among the gene-level PrediXcan associations in each tissue and identified genes that are genome-wide significant when Bonferroni correcting for the number of genes tested across tissues (see Methods). In subsequent analyses, we focused primarily on four out of five psychiatric traits, as we observed limited power for major depression (Supplementary Fig. 1a–e). We observed tissue- and disease-dependent levels of $\widehat{\pi_1}$ (Fig. 1a, Supplementary Table 2, and Supplementary Fig. 2a–d). The number of significant genes in each tissue (Table 1) was determined primarily by tissue sample size (Supplementary Fig. 3 and Supplementary Table 3; Spearman's $\rho = 0.91$, $P = 6.43 \times 10^{-6}$ for schizophrenia), showing the limitation of the number of significant genes as a metric for tissue prioritization. In contrast, $\widehat{\pi_1}$ was not associated with tissue sample size (Spearman's correlation $\rho = 0.27$, $P = 0.35$ for schizophrenia). Putamen basal ganglia ($\widehat{\pi_1} = 0.33$) and nucleus accumbens basal ganglia ($\widehat{\pi_1} = 0.32$) showed the highest $\widehat{\pi_1}$ for schizophrenia, indeed greater than the DLPFC ($\widehat{\pi_1} = 0.27$) despite the latter's larger sample size ($n = 538$). Our findings suggest that the relatively large number of schizophrenia associations in DLPFC from a recent study[19] is due to relatively large sample size and provides limited evidence for greater relevance of this brain structure. For broad depression, the non-brain tissues colon ($\widehat{\pi_1} = 0.21$) and whole blood ($\widehat{\pi_1} = 0.20$) showed the highest estimated true positive rate, suggesting the relevance of

the gastrointestinal or enteric nervous[20] and immune[21] systems for identifying genetic predispositions.

Gene discoveries are shown in Supplementary Table 4a–d and reveal 275 unique genes for schizophrenia, 13 for bipolar disorder, 12 for ADHD, and 31 for broad depression. Of these, 70 disease-associated genes are not in linkage disequilibrium (LD) ($r^2 > 0.10$) with a GWAS index SNP and are considered novel discoveries from the PrediXcan analysis (Table 1). No genes were significantly associated with major depression, which may be due to its relatively small GWAS sample size and its large genetic and phenotypic heterogeneity. Interestingly, the schizophrenia and broad depression results revealed substantial overlap, as 17 of the 31 broad depression genes, all located within the MHC region, were also found to be significantly associated with schizophrenia. In the PrediXcan analysis, the distribution of the number of cis-variant predictors for optimal modeling (see Methods) of GReX for each tissue-gene pair (Supplementary Fig. 4) and the number of contributing variants to the novel gene-level associations (Supplementary Table 4a–d) demonstrate the importance of multi-variant models of GReX to identify trait-associated genes.

We replicated the significant associations for schizophrenia and broad depression using UK Biobank[22] and BioVU[23], respectively (see Methods), but were unable to evaluate replication for the other two disorders due to lack of available replication datasets. Notably, we observed significantly greater replication rate for schizophrenia than expected by chance (Fig. 1b). We identified and replicated an association between increased GReX of *LRP8* in hypothalamus and schizophrenia risk ($P = 7.77 \times 10^{-6}$ in the discovery sample and $P = 6.62 \times 10^{-8}$ in the UKB replication sample). *LRP8* is a key component of the RELN pathway and was previously linked to psychosis risk[24]. Increased GReX for *FPKP* ($P = 2.4 \times 10^{-4}$) showed replicated association with depression in BioVU, but a more systematic validation for this phenotype would require greater sample size.

Applying a single-SNP analysis, we also identified the target genes for expression quantitative trait loci (eQTLs) using the “best eQTL per eGene”[25,26] (b-eQTL, FDR < 0.05) in brain and non-brain transcriptomes to gain additional insights into the role of gene regulation in conferring disease predisposition. As in the PrediXcan analysis, we observed tissue- and disease-dependent levels of $\widehat{\pi_1}$ (Fig. 1c and Supplementary Table 5). The substantial true positive rate for schizophrenia in anterior cingulate cortex ($\widehat{\pi_1} = 0.45$, number of independent true associations $n_{ta} = 580$) suggests that b-eQTLs from this specific brain region may be used to improve the search for novel trait associations.

From the joint-tissue eQTL analysis (see Methods), which aims to improve power for eQTL discovery through joint-tissue mapping that captures heterogeneity in effect sizes between tissues while accounting for differences in sample size, non-brain b-eQTLs continued to show a substantial true positive rate of association after excluding eQTLs also found in brain (Fig. 1d and Supplementary Table 6). For example, 41% (i.e., 355 variants) of the whole blood b-eQTLs that are not brain eQTLs are estimated to be true associations with schizophrenia (compared to 30%; i.e., 1,836 variants prior to the filtering).

We aimed to improve the functional characterization of known disease-associated loci, focusing primarily on the 145 schizophrenia loci[13] by considering the target genes from the single-SNP joint-tissue eQTL analysis. The 145 schizophrenia loci, in general, contained multiple genes (up to 131 genes per locus, defined as the cis-region, +/− 1 Mb, of the index SNP; average number of genes is 10; Fig. 2a), complicating the search for the causal gene mechanisms and highlighting the importance of incorporating eQTL and other functional information to identify the relevant genes. We found that the 145 schizophrenia loci on average implicated 3.7 associated genes. This implies a substantial reduction in the number of proposed genes. The joint-tissue analysis implicated 290 genes at known GWAS loci to be significantly associated with schizophrenia in a tissue-specific manner (Fig. 2b).

We aimed to further reduce the number of candidate causal genes of the schizophrenia risk loci by distinguishing pleiotropic or causal genes (i.e., genetic variants influence both gene expression and disease risk) from genes that are associated indirectly[27] (e.g., genes with causal eQTLs in LD with disease causal variants). We performed Summary-based Mendelian Randomization (SMR) analyses (see Methods) in conjunction with the Heterogeneity in Dependent Instruments (HEIDI) test using brain eQTLs from a large-scale meta-analysis[28]. Despite the limitations of these methods[29], they may be used as a starting point for further functional studies of proposed genes. We found that 65 genes passed the default significance threshold of 8.4×10^{-6} under SMR. Nineteen of these genes showed a $P_{\mathrm{HEIDI}} \geq 0.05$ (Supplementary Table 7a) from the subsequent HEIDI test, indicating that the null hypothesis of a single variant affecting both gene expression and disease risk could not be rejected. Of the 65 genes, 42 showed significant association in at least one tissue from the PrediXcan analysis. The extended MHC region is characterized by a disproportionately large number of OMIM genes and NHGRI catalog SNPs[30]; the large number of proxy SNPs and the high gene density at this locus illustrate the challenges in fine-mapping the causal variant(s) and in finding the causal gene mechanism(s), respectively. Notably, PrediXcan analysis, which assumes multiple-variant imputation models, implicated *C4A*[31] to be associated with schizophrenia in seven GTEx brain tissues, colon, adrenal gland, and whole blood, but not in DLPFC (Fig. 2c).

We explored whether a non-brain tissue reveals potentially pleiotropic or causal genes and performed a secondary analysis using eQTL information from GTEx whole blood (see Methods). Significant SMR association was observed for 37 genes, of which for seven the null hypothesis of a single variant affecting both gene expression and disease risk could not be rejected ($P_{\mathrm{HEIDI}} > 0.05$). Notably, all seven genes showed $P_{\mathrm{HEIDI}} < 0.05$ based on brain eQTLs, consistent with a model of linkage in which two distinct causal variants (for expression and disease risk) in the locus exist, implying that non-brain tissues can reveal important additional insights into the biological mechanisms of schizophrenia.

There is substantial sharing of eQTL SNPs (as estimated by $\widehat{\pi_1}$) among the tissues (Fig. 3a), with whole blood a clear outlier for the level of eQTL sharing with the other tissues. Among the brain regions, cerebellum and cerebellar hemisphere are outliers (Fig. 3a). We compared results of the single-SNP METASOFT joint-tissue analysis across tissues and found that different tissues may propose different regulatory mechanisms within the same locus (Fig.

3b). For example, *GATA2DA* shows strong tissue-specific effects in the METASOFT analysis with strong evidence for association in whole blood, while other tissues do not indicate association. In contrast, *PBX4* is implicated in hippocampus and cerebellum, while associations are weaker in the other brain tissues and absent in non-brain tissues.

We observed significant correlations in GReX of disease-associated genes for all tissues (Fig. 4a for pairwise GReX correlations in putamen basal ganglia, which had the largest $\widehat{\pi_1}$ for schizophrenia and showed significantly greater correlations than random GReX-derived co-expression networks in the tissue (empirical $P < 0.001$; Fig. 4b; see Methods), raising the possibility that they belong to a small number of genetically determined co-expression networks. Hierarchical clustering (see Methods) highlighted several clusters among these genes that appear to have coordinated GReX (Fig. 4c for putamen basal ganglia). Underscoring the challenges of precisely mapping trait causal effects (e.g., LD contamination), these coordinated expression patterns may also highlight "convergent" function[32] and pathophysiology[33] for the genes and, possibly, trans-regulatory mechanisms[34] (e.g., involving the genes on distinct chromosomes) in the brain underlying schizophrenia genetic predisposition. Thus, genetic susceptibility loci may act through disease-associated genes within co-expression networks to influence a limited number of core disease-related biological processes.

We performed Gene Ontology (GO) enrichment analysis (see Methods) of the PrediXcan implicated genes for schizophrenia and identified "biological processes" that are concordant between tissues (Fig. 5a, for example, for the comparison between putamen basal ganglia and whole blood; Spearman $\rho = 0.92$, $P < 2.2 \times 10^{-16}$). However, we also identified biological processes specific to one tissue of a tissue pair (Fig. 5b for whole blood) and uncovered enriched biological processes (Bonferroni-adjusted $P < 0.05$; Fig. 5c in the tissue with the highest estimated $\widehat{\pi_1}$). These findings suggest either the tissue specificity of disease-associated genetically determined expression or that tissue-shared genetically determined expression may exert their tissue-level or pleiotropic function by implicating different processes in the different tissues.

Our study suggests that there is enormous value in both brain and non-brain eQTL annotations for improved discovery of novel susceptibility genes for psychiatric disorders. PrediXcan analyses across tissues revealed 331 unique disease-associated genes, 76% of which were detected only in difficult-to-acquire tissues. Furthermore, brain and non-brain tissues highlight different molecular mechanisms underlying psychiatric traits. Non-brain tissues identified 113 of the 331 (34%) of the disease-associated PrediXcan genes. Despite the challenges of distinguishing relevant pathogenic tissue(s) from tissue-shared gene regulation[35], our study highlights the importance of joint-tissue analyses, in particular using non-brain tissues. Notably, our findings indicate that disease-associated genes may involve multiple variant-predictors, implying that, consistent with the recommendations of Gusev *et al.*[36] and our own previous work[18,35], gene expression imputation may substantially extend single-eQTL-based analysis. We further showed that eQTL and transcriptome-based annotation of schizophrenia loci can substantially reduce the number of genes to be investigated in future functional studies. For example, using PrediXcan and SMR analyses

and taking into account potential LD confounding[27], we proposed multiple gene mechanisms, which should be further explored by functional studies. Co-expression network analysis of schizophrenia-associated GReX revealed clustering into a limited number of genetically defined clusters. Consistent with this, GO enrichment analysis of schizophrenia-associated GReX implicated a limited number of biological processes, highlighting shared molecular mechanisms, including immune function and antigen presentation, in line with reports on the role of immune dysregulation in schizophrenia[37,38]. In conclusion, characterizing genes based on genetic regulation in brain and non-brain transcriptomes improves both the identification of novel neuropsychiatric genes and the functional elucidation of genetic risk factors of neuropsychiatric diseases.

Online Methods

Please see the Life Sciences Reporting Summary for additional details.

Tissues and eQTL data.

We used GTEx v6p dataset in the analyses presented here, including the (cis) eQTLs and "best eQTL per eGene"[25,26] (b-eQTLs) in each of the brain regions ($n = 10$) and in selected non-brain tissues (whole blood, adrenal gland and colon). We present results on 889 brain RNA-seq samples from 10 brain regions and 633 non-brain samples. Up to 393 distinct individuals are included in this analysis. We used the CommonMind Consortium transcriptome data from 538 samples in dorsolateral prefrontal cortex (DLPFC). See Supplementary Table 1a for the sample size for each tissue.

GWAS data psychiatric disorders.

We used genome-wide association study (GWAS) meta-analysis results of five psychiatric traits[3,10–17] (Supplementary Table 1b). These studies have been conducted in compliance with all ethical regulations. All studies have been approved by the ethical board of the relevant institution. Since each study comprises many different study cohorts, please see original publications for details.

PrediXcan as expression-based risk score analysis.

Proposing a gene mechanism from a disease-associated eQTL assumes that a single-eQTL model suffices to implicate the gene. In this case, a disease-associated eGene may be viewed as a single-variant-determined gene expression phenotype that shows an association with disease. Indeed, if x is the genotype (expressed as allelic dosage) for a GWAS individual at the eQTL variant S and $\widehat{\beta}$ its estimated effect size on the expression of the gene g as determined by a reference eQTL resource such as GTEx, then $\widehat{\beta}x$ is the imputed eQTL-determined expression of the eGene g for the GWAS sample; the imputed (genetically determined) gene expression can be tested for association with disease risk Y using GWAS summary statistics (effect size $\widehat{\alpha}$ and corresponding standard error $SE(\widehat{\alpha})$ of S on Y).

Some genes, however, may require multiple variants to adequately impute the genetically determined component of the gene expression phenotype or the "**G**enetically **R**egulated **eX**pression" (GReX)[18]. Thus, we applied PrediXcan[18] to the GWAS results for five

psychiatric traits to discover genes with differential GReX ($P_{\text{Bonferroni-adjusted}} < 0.05$ after adjustment for the total number of genes across all tissues tested) between patients and controls. PrediXcan implements a tissue-dependent gene expression imputation model[18] (using elastic net), which is evaluated, for optimal modeling, through the 10-fold cross-validation R^2 within the reference transcriptome panel (GTEx). Genes implicated by PrediXcan for a given disease may be viewed as a *generalized* disease-associated eGene (which is conventionally the target gene of *one* source eQTL variant, namely the b-eQTL). An observed association between GReX and a disease phenotype from a causal link would suggest that the causal direction of effect is from GReX to disease risk since the trait is not likely to alter the germline genetic profile. In this generalized version, we can, using GWAS summary statistics, estimate the effect size $\hat{\alpha}$ and corresponding standard error $SE(\widehat{\alpha})$ of the imputed genetically determined component of gene expression on disease risk as follows (by noting that the imputed gene expression is *a genetic risk score*[39] with weights derived from the reference transcriptome panel):

$$\hat{\alpha} = \frac{\sum_{j=1}^{m} \hat{\theta}_j \hat{\beta}_j s_j^{-2}}{\sum_{j=1}^{m} \hat{\beta}_j^{2} s_j^{-2}}$$

$$SE(\hat{\alpha}) = 2\sqrt{\frac{1}{\sum_{j=1}^{m} \hat{\beta}_j^{2} s_j^{-2}}}$$

where $\hat{\beta}_j$ is the effect size of the j-th variant from the gene expression model (derived from the reference panel) consisting of uncorrelated m cis-variant predictors (i.e., located within +/− 1 Mb of the gene), while $\hat{\theta}_j$ and s_j are the effect size and the corresponding standard error of this effect size from the GWAS trait summary data. We note that, under the null hypothesis (of zero contribution $\hat{\theta}_j$ to trait variation from all predictors), we obtain the distribution of $\hat{\alpha}$:

$$\hat{\alpha} \sim N\left(0, (SE(\hat{\alpha}))^2\right)$$

which allows us to assign a P-value to the observed effect size of the imputed genetically determined component of gene expression on disease risk. When $m = 1$, this reduces to the special case of testing an eGene for association with phenotype.

In the general case in which some LD is present in the set of m cis-variant predictors for a given gene, we estimate the correlation between the “genetic risk score” generated from the gene expression prediction model and the phenotype using only GWAS summary statistics. Let X be the n-by-m standardized genotype matrix and y be the standardized phenotype n-by-1 matrix in the test GWAS. (In general, the genotype data X and phenotype data y are not publicly available.) Then the genetic risk score $GRS(n)$ for the n GWAS individuals is given by:

$$GRS(n) := X\beta$$

where β is the m-vector of model-derived effect sizes. The correlation between $GRS(n)$ and y is thus given by the following expression:

$$cor(GRS(n), y) = \beta^T X^T y \Big/ \sqrt{\beta^T X^T X \beta y^T y} \quad (*)$$

We note that since X and y are standardized, the term $X^T y$ in the numerator of (*) is a correlation m-vector (equal to the trait effect size vector (because of the standardization), easily obtained using the GWAS summary statistics for the m variants). Furthermore, $X^T X$ in the denominator of (*) yields an m-by-m matrix (LD) matrix, which can be estimated from an ancestry-matched reference panel (such as from the 1000 Genomes project) with or without LD regularization[40,41].

TWAS[36] and S-PrediXcan[42] have implemented approaches based on summary statistics.

We used the S-PrediXcan[18,42] implementation to test genetic associations with five psychiatric traits (Supplementary Table 1). The P-value threshold was corrected for the total number of genes tested across all tissues. Since there are two pairs of brain tissues that are replicates and therefore do not constitute independent tissues (cortex vs. frontal cortex and cerebellum vs. cerebellar hemisphere), we only corrected for one of the tissues for each replicate pair. We removed the tissue that included the lowest number of imputed genes. The P-value thresholds are slightly different across traits and are reported in Supplementary Table 1. We further used PrediXcan[18] in the replication analyses of schizophrenia in UKB samples (n = 1,561 cases and 267,494 psychiatrically healthy controls) and of broad depression in the BioVU samples[23] (n = 3,570 cases and 14,456 controls). To evaluate replication for schizophrenia, we applied S-PrediXcan to GWAS summary statistics derived from an application of BOLT-LMM[43]. We explored potential sample overlap between schizophrenia GWAS data and UKB samples, using GEAR[44] and LD score regression[45] analyses. The results of these analyses did not suggest a significant amount of overlap (LD score regression covariate intercept = 0.0038 (s.e. = 0.006) and LambdaMeta = 1.54). The BioVU samples included in our analyses were of European genetic ancestry, as confirmed by principal components analysis.

PrediXcan software can be downloaded from the github repository (https://github.com/hakyim/PrediXcan).

Estimating the true rate of association (π_1).

We estimated the true positive rate ($\widehat{\pi_1} := 1 - \widehat{\pi}_0$) for association with disease among the eQTLs from the “best eQTL per eGene” (b-eQTL, FDR < 0.05) set or, separately, among the gene-level results from the PrediXcan analysis in each tissue, using the method of Storey *et al.*[25]. The false discovery rate (FDR) at a given P-value p was estimated using the following equation:

$$FDR(p) = \frac{\widehat{\pi_0} K p}{Count\left(p_{i,D} \leq p\right)}$$

In the case of b-eQTLs, K is the number of eQTLs used, the $p_{i,D}$ is the set of P-values indexed by the eQTL SNP i for association with the disease trait D, and $\widehat{\pi_0}$ is the proportion of null SNPs:

$$\widehat{\pi_0} = Count\left(p_{i,D} > \tau\right)/K\left(1-\tau\right)$$

which was defined with the tuning parameter τ (selected to be 0.50, as in Storey *et al.*). In the case of the gene-level results from the PrediXcan analysis, the input set is that of genes tested (i.e., with imputation models).

Joint-tissue eQTL mapping.

In addition to single-tissue eQTL analysis, the joint analysis of the brain regions and the additional tissues may facilitate identification of additional eQTLs and quantification of their tissue specificity. We used METASOFT for joint-tissue eQTL analysis, across the 10 brain regions, whole blood, adrenal gland, and colon. METASOFT had been previously used in the second-phase GTEx Consortium paper[46]. Joint-tissue eQTL mapping extends single-tissue eQTL mapping results by borrowing information across tissues. This approach allowed us to quantify the tissue specificity or tissue-sharedness of an eQTL-gene pair while taking into account the differential eQTL discovery power between tissues.

We used joint-tissue eQTL mapping to identify eQTLs in non-brain tissues that are not active in any of the brain regions. This analysis provided an approach to testing the importance of regulatory variation in non-brain tissues for psychiatric disease predisposition through the estimated true positive rate $\widehat{\pi_1}$ for disease associations among the non-brain eQTLs that are not also regulatory in brain.

SMR and HEIDI analysis of schizophrenia summary statistics.

We sought to further refine the set of candidate causal genes by distinguishing pleiotropy or causality (i.e., a single variant affecting both gene expression and disease risk) from a model of linkage (i.e., two distinct causal variants, for gene expression and disease risk). We applied Summary-data based Mendelian Randomization (SMR) and HEterogeneity In Dependent Instruments (HEIDI) tests[47], which were implemented in SMR software (http://cnsgenomics.com/software/smr/), to the schizophrenia GWAS results. For the input GWAS results, effect sizes were calculated as log(*odds ratio*) with its corresponding standard error. The top associated eQTL for the SMR test was selected using the default value of $P_{\text{eQTL}} < 5.0 \times 10^{-8}$. For the HEIDI test, up to 20 SNPs (eQTL $P < 1.57 \times 10^{-3}$; LD r^2 with top SNP < 0.9) in the cis-eQTL region (including the top cis-eQTL) were included per recommendation of the software developers.

The SMR analyses were based on eQTL information provided by Qi *et al.*[28] who meta-analyzed GTEx-brain, CMC, and ROSMAP eQTL data (estimated effective sample size: n = 1,194). The Brain-eMeta eQTL summary data were downloaded in binary format (see http://cnsgenomics.com/software/smr/#DataResource). In addition, we explored the value of a non-brain tissue by using eQTL summary data based on GTEx whole blood (version 7) for SMR analysis (http://cnsgenomics.com/software/smr/#DataResource). SMR locus plots and plots of GWAS and eQTL effect sizes were created with R code (http://cnsgenomics.com/software/smr/#SMRlocusplot).

Co-expression analysis, hierarchical clustering, and pathway analysis.

We hypothesized that genetically defined co-expression networks (versus networks based on directly measured gene expression, as typically done) would improve our ability to highlight novel biology underlying psychiatric disease predisposition. Using Weighted Gene Co-expression Network Analysis (WGCNA v1.61) with default settings, we performed co-expression network analysis and hierarchical clustering of the GReX levels estimated in the 1000 Genomes European samples (using gene expression imputation, as implemented in PrediXcan[18]) for the disease-associated genes. In general, the availability of a reference population panel (e.g., 1000 Genomes) in a population of interest enables network analysis of the genetically determined expression traits.

We also applied multiscale bootstrap resampling[48] (n = 1,000 bootstrap replicates) to hierarchical clustering in order to quantify the uncertainty in the clusters, with “average” as the agglomerative method and “correlation” as distance method, generating an “Approximately Unbiased” P-value and a “Bootstrap Probability” for each cluster. This analysis allowed us to identify the clusters for which the null hypothesis of the non-existence of the cluster is rejected at $\alpha = 0.05$.

In each tissue, we determined the significance of the set of pairwise (Spearman) correlations in GReX levels (of the 1000 Genomes European samples). We evaluated whether the resulting network was more highly correlated than expected based on 1,000 co-expression networks derived from the GReX of random sets of genes (each comprising the same number of genes as the actual network). We further explored whether some disease-associated genes may be regulatory for other such genes. The GReX correlation matrix for the disease-associated genes may be used to identify novel trans-regulatory mechanisms driving genetic susceptibility. For example, a cis eQTL (schizophrenia-associated) SNP for a gene may serve as a trans eQTL for another distal gene (that is correlated with the first).

Using clusterProfiler[49], we compared the sets of “biological processes” (as defined by Gene Ontology) implicated by the disease-associated genes from each pair of tissues. We calculated the Spearman correlation for the proportion of disease-associated genes mapping to each biological process to quantify the level of concordance between tissues. For each tissue pair, we identified the biological processes present only in one tissue and with at least five genes. Furthermore, we performed a two-proportions z-test using prop.test in R and identified those processes that significantly differed between tissues (q-value < 0.05) between tissues.

Software for statistical analysis and visualization.

The software package R (https://www.r-project.org/) was used for the statistical analysis. We used the libraries “ggplot2” and “corrplot” for visualization.

Data Availability Statement

The protected data for the GTEx project (e.g., genotype and RNA-sequence data) are available via access request to dbGaP accession number phs000424.v6.p1. Processed GTEx data (e.g., gene expression and eQTLs) are available on the GTEx portal (https://gtexportal.org). The URLs of the summary statistics datasets of all the GWAS meta-analyses analyzed in the paper can be found in Supplementary Table 1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. SAIC-F subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplements to University of Miami grants DA006227 & DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina - Chapel Hill (MH090936 & MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the University of Pennsylvania (MH101822). E.M.D. is supported by the Foundation Volksbond Rotterdam. E.R.G. is grateful to the President and Fellows of Clare Hall, University of Cambridge for providing a stimulating intellectual home and benefited from a Fellowship in the college. E.R.G. also acknowledges support from the Academic Medical Center, University of Amsterdam during the early stages of the study. E.R.G. and N.J.C. acknowledge support from R01MH113362. We acknowledge the Psychiatric Genomics Consortium (PGC) groups for sharing the published meta-analysis data.

References

1. Gratten J, Wray NR, Keller MC & Visscher PM Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nature neuroscience 17, 782–790, doi:10.1038/nn.3708 (2014). [PubMed: 24866044]
2. Nicolae DL et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS genetics 6, e1000888, doi:10.1371/journal.pgen.1000888 (2010). [PubMed: 20369019]
3. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427, doi:10.1038/nature13595 (2014). [PubMed: 25056061]
4. Gamazon ER et al. The convergence of eQTL mapping, heritability estimation and polygenic modeling: emerging spectrum of risk variation in bipolar disorder. arxiv (2013).
5. Geschwind DH & Flint J Genetics and genomics of psychiatric disease. Science 349, 1489–1494, doi:10.1126/science.aaa8954 (2015). [PubMed: 26404826]
6. Ozomaro U, Wahlestedt C & Nemeroff CB Personalized medicine in psychiatry: problems and promises. BMC medicine 11, 132, doi:10.1186/1741-7015-11-132 (2013). [PubMed: 23680237]

7. Jansen R et al. Gene expression in major depressive disorder. Molecular psychiatry 21, 444, doi: 10.1038/mp.2015.94 (2016). [PubMed: 26100536]
8. Bercik P et al. The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. Gastroenterology 141, 599–609, 609 e591–593, doi:10.1053/j.gastro.2011.04.052 (2011). [PubMed: 21683077]
9. Kato TA, Hayakawa K, Monji A & Kanba S Missing and Possible Link between Neuroendocrine Factors, Neuropsychiatric Disorders, and Microglia. Frontiers in integrative neuroscience 7, 53, doi: 10.3389/fnint.2013.00053 (2013). [PubMed: 23874274]
10. Neale BM et al. Meta-analysis of genome-wide association studies of attention-deficit/ hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry 49, 884–897, doi:10.1016/j.jaac.2010.06.008 (2010). [PubMed: 20732625]
11. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature genetics 43, 977–983, doi:10.1038/ng.943 (2011). [PubMed: 21926972]
12. Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. Molecular psychiatry 18, 497–511, doi:10.1038/mp.2012.21 (2013). [PubMed: 22472876]
13. Pardinas AF et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nature genetics 50, 381–389, doi:10.1038/ s41588-018-0059-2 (2018). [PubMed: 29483656]
14. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. Cell 173, 1705–1715 e1716, doi:10.1016/j.cell.2018.05.046 (2018). [PubMed: 29906448]
15. Wray NR et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature genetics 50, 668–681, doi:10.1038/s41588-018-0090-3 (2018). [PubMed: 29700475]
16. Howard DM et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nature communications 9, 1470, doi:10.1038/ s41467-018-03819-3 (2018).
17. Demontis D et al. Discovery of the first genome-wide significant risk loci for attention deficit/ hyperactivity disorder. Nat. Genet 51, 63–75 (2019). [PubMed: 30478444]
18. Gamazon ER et al. A gene-based association method for mapping traits using reference transcriptome data. Nature genetics 47, 1091–1098, doi:10.1038/ng.3367 (2015). [PubMed: 26258848]
19. Huckins LM et al. Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. Nat. Genet 51, 659–674 (2017).
20. Foster JA & McVey Neufeld KA Gut-brain axis: how the microbiome influences anxiety and depression. Trends in neurosciences 36, 305–312, doi:10.1016/j.tins.2013.01.005 (2013). [PubMed: 23384445]
21. Leonard BE The concept of depression as a dysfunction of the immune system. Current immunology reviews 6, 205–212, doi:10.2174/157339510791823835 (2010). [PubMed: 21170282]
22. Sudlow C et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine 12, e1001779, doi:10.1371/journal.pmed. 1001779 (2015). [PubMed: 25826379]
23. Denny JC et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nature biotechnology 31, 1102–1110, doi: 10.1038/nbt.2749 (2013).
24. Xiao X et al. Further evidence for the association between LRP8 and schizophrenia. Schizophr Res, doi:10.1016/j.schres.2017.05.002 (2017).
25. Storey JD & Tibshirani R Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America 100, 9440–9445, doi:10.1073/ pnas.1530509100 (2003). [PubMed: 12883005]

26. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660, doi:10.1126/science.1262110 (2015). [PubMed: 25954001]
27. Wainberg M et al. Opportunities and challenges for transcriptome-wide association studies. Nat. Genet 51, 592–599 (2019). [PubMed: 30926968]
28. Qi T et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nature communications 9, 2282, doi:10.1038/s41467-018-04558-1 (2018).
29. Hauberg ME et al. Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression. American journal of human genetics 101, 157, doi:10.1016/j.ajhg.2017.06.003 (2017). [PubMed: 28686855]
30. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. Nature genetics 43, 969–976, doi: 10.1038/ng.940 (2011). [PubMed: 21926974]
31. Sekar A et al. Schizophrenia risk from complex variation of complement component 4. Nature 530, 177–183, doi:10.1038/nature16549 (2016). [PubMed: 26814963]
32. Zhang J et al. Spatial clustering and common regulatory elements correlate with coordinated gene expression. PLoS computational biology 15, e1006786, doi:10.1371/journal.pcbi.1006786 (2019). [PubMed: 30822341]
33. Voineagu I et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474, 380–384, doi:10.1038/nature10110 (2011). [PubMed: 21614001]
34. Westra HJ et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature genetics 45, 1238–1243, doi:10.1038/ng.2756 (2013). [PubMed: 24013639]
35. Gamazon ER et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nature genetics 50, 956–967, doi:10.1038/s41588-018-0154-4 (2018). [PubMed: 29955180]
36. Gusev A et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat. Genet 50, 538–548 (2018). [PubMed: 29632383]
37. Jones AL, Mowry BJ, Pender MP & Greer JM Immune dysregulation and self-reactivity in schizophrenia: do some cases of schizophrenia have an autoimmune basis? Immunology and cell biology 83, 9–17, doi:10.1111/j.1440-1711.2005.01305.x (2005). [PubMed: 15661036]
38. Stringer S, Kahn RS, de Witte LD, Ophoff RA & Derks EM Genetic liability for schizophrenia predicts risk of immune disorders. Schizophrenia research 159, 347–352, doi:10.1016/j.schres.2014.09.004 (2014). [PubMed: 25266548]

Methods-only references

39. International Consortium for Blood Pressure Genome-Wide Association Studies et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478, 103–109, doi:10.1038/nature10405 (2011). [PubMed: 21909115]
40. Shi H, Kichaev G & Pasaniuc B Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. American journal of human genetics 99, 139–153, doi:10.1016/j.ajhg.2016.05.013 (2016). [PubMed: 27346688]
41. Gamazon ER, Cox NJ & Davis LK Structural architecture of SNP effects on complex traits. American journal of human genetics 95, 477–489, doi:10.1016/j.ajhg.2014.09.009 (2014). [PubMed: 25307299]
42. Barbeira A et al. Integrating tissue specific mechanisms into GWAS summary results. Biorxiv, doi: 10.1101/045260 (2017).
43. Loh PR et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature genetics 47, 284–290, doi:10.1038/ng.3190 (2015). [PubMed: 25642633]
44. Chen GB et al. Across-cohort QC analyses of GWAS summary statistics from complex traits. Eur J Hum Genet 25, 137–146, doi:10.1038/ejhg.2016.106 (2016). [PubMed: 27552965]
45. Bulik-Sullivan BK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics 47, 291–295, doi:10.1038/ng.3211 (2015). [PubMed: 25642630]

46. The GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 550, 204–213, doi:10.1038/nature24277 (2017). [PubMed: 29022597]

47. Zhu Z et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature genetics 48, 481–487, doi:10.1038/ng.3538 (2016). [PubMed: 27019110]

48. Shimodaira H Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. Ann Stat 32, 2616–2641, doi:10.1214/009053604000000823 (2004).

49. Yu G, Wang LG, Han Y & He QY clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology 16, 284–287, doi:10.1089/omi.2011.0118 (2012). [PubMed: 22455463]

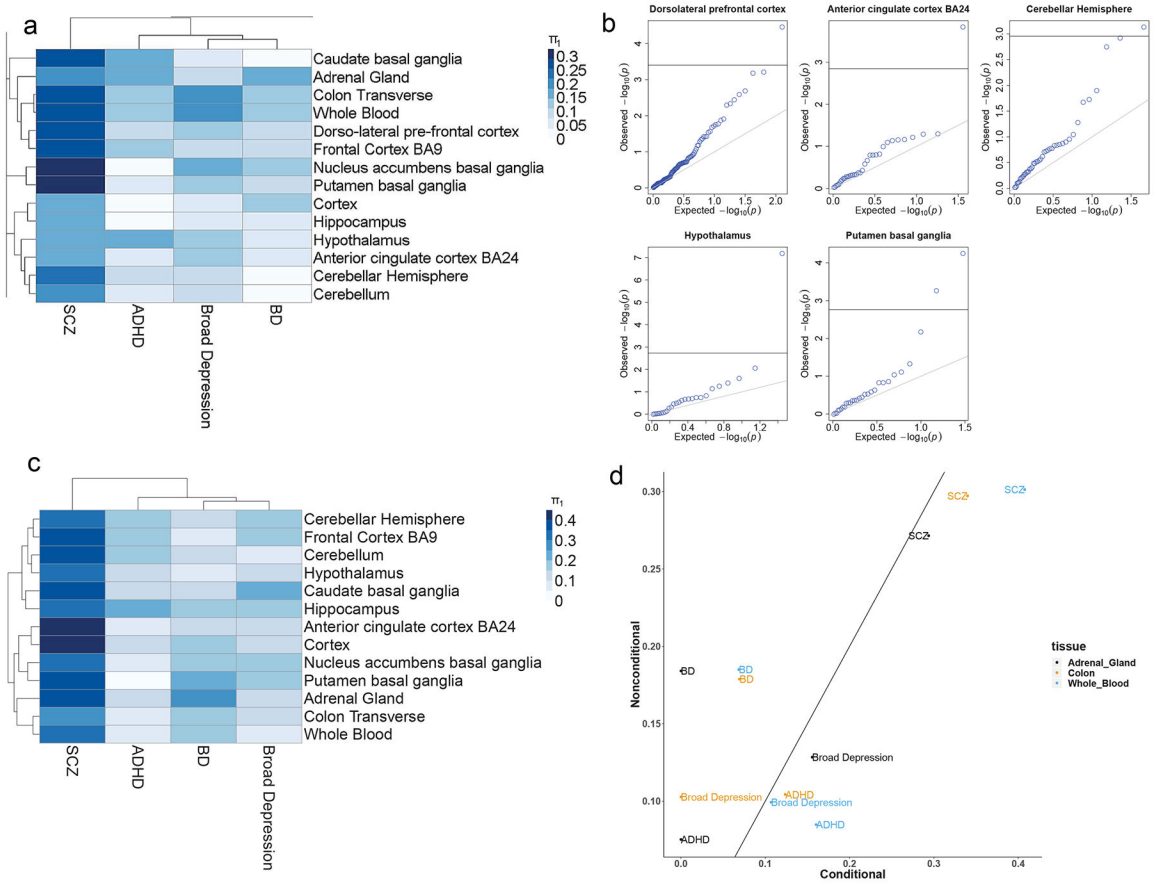

Figure 1 |. PrediXcan and eQTL analysis of GWAS of psychiatric traits.

a, Heatmap and hierarchical clustering based on true positive rate (π_1) for trait associations among PrediXcan associations in each tissue. Schizophrenia and bipolar disorder were most highly enriched in putamen basal ganglia and adrenal gland, respectively. Broad depression showed relatively strong enrichment in colon. **b**, Significant replication of schizophrenia PrediXcan associations (Bonferroni-adjusted $P < 0.05$, indicated by horizontal line) was observed within the UK Biobank sample (1,561 cases; 267,494 controls). Replication *P*-value for a gene was from the application of S-PrediXcan to GWAS summary statistics derived from BOLT-LMM. **c**, Heatmap and hierarchical clustering based on true positive rate (π_1) for trait associations among b-eQTL associations in each tissue. **d**, Non-brain eQTL enrichment was observed even after conditioning on brain eQTLs from the joint-tissue eQTL analysis. The joint-tissue analysis allowed us to quantify the tissue specificity and tissue-sharedness of an eQTL-gene pair while taking into account differential power for eQTL discovery between tissues. The *x*-axis shows the π_1 after including only those b-eQTLs active in the non-brain tissue but not regulatory in any of the 10 brain regions. The *y*-axis shows the π_1 without the filtering. Points below the diagonal line show disease and tissue pairs for which a higher true positive rate was observed after filtering the brain eQTLs. Sample sizes: schizophrenia (40,675 cases, 64,643 controls); ADHD (20,183 cases, 35,191 controls); bipolar disorder (11,974 cases, 51,792 controls); broad depression (113,769 cases, 208,811 controls).

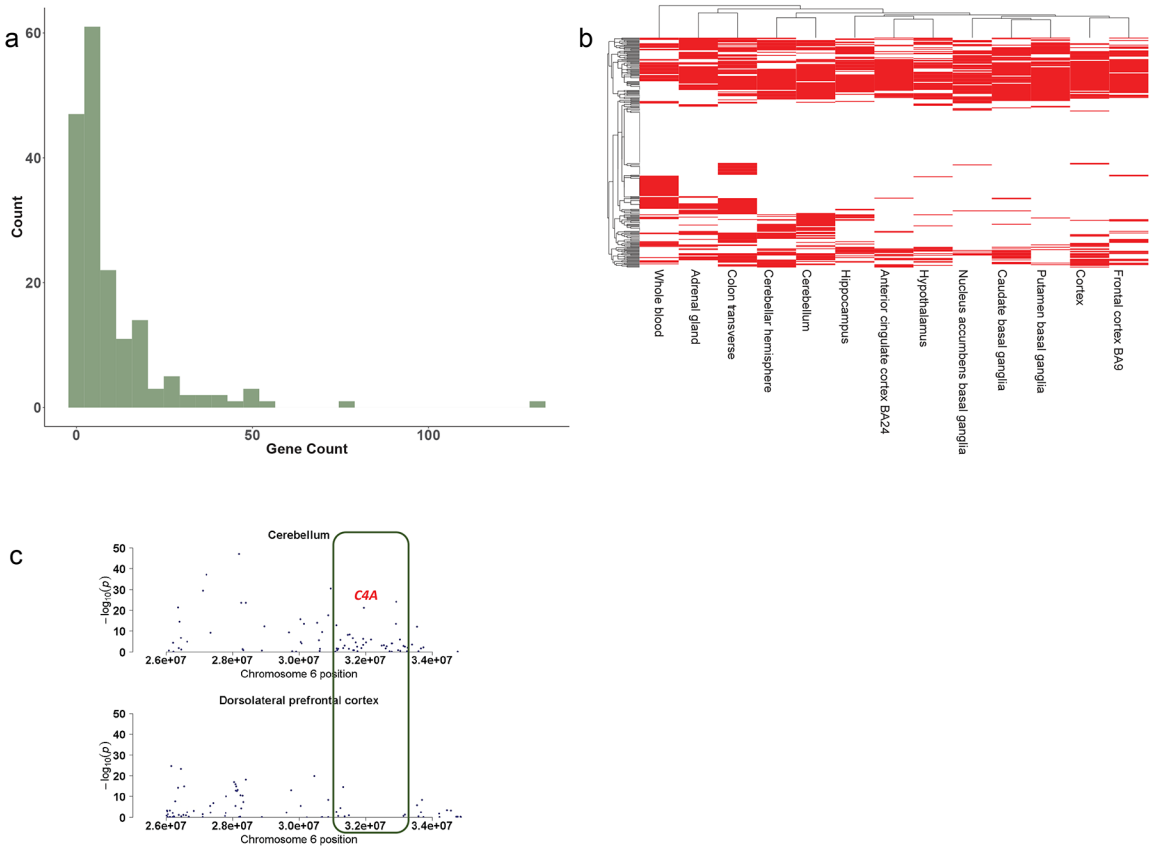

Figure 2 |. Proposing causal variant and causal gene mechanism at known schizophrenia-associated loci (*n* = 145 loci; 179 index SNPs).

a, Distribution of number of nearby genes (defined as +/− 1 Mb of index SNPs) at known loci. **b**, From the joint-tissue (METASOFT) analysis, 290 genes at known loci were significantly associated with schizophrenia (40,675 cases, 64,643 controls) in a tissue-dependent manner. Significant associations (METASOFT *m*-value ≥ 0.90) of these 290 genes are shown as red bars, while non-significant associations are shown as white bars. The figure shows clustering of genes that are shared across tissues, but also indicates genes that have tissue-specific effects. **c**, LocusZoom plot of the region surrounding the *C4A* gene indicates that *C4A* is significantly associated in GTEx cerebellum but not in dorsolateral prefrontal cortex. *y*-axis is the schizophrenia association *P*-value (in log scale). The large number of SNPs and genes in the locus illustrate the challenges in fine-mapping the causal variant and in determining the gene driver(s).

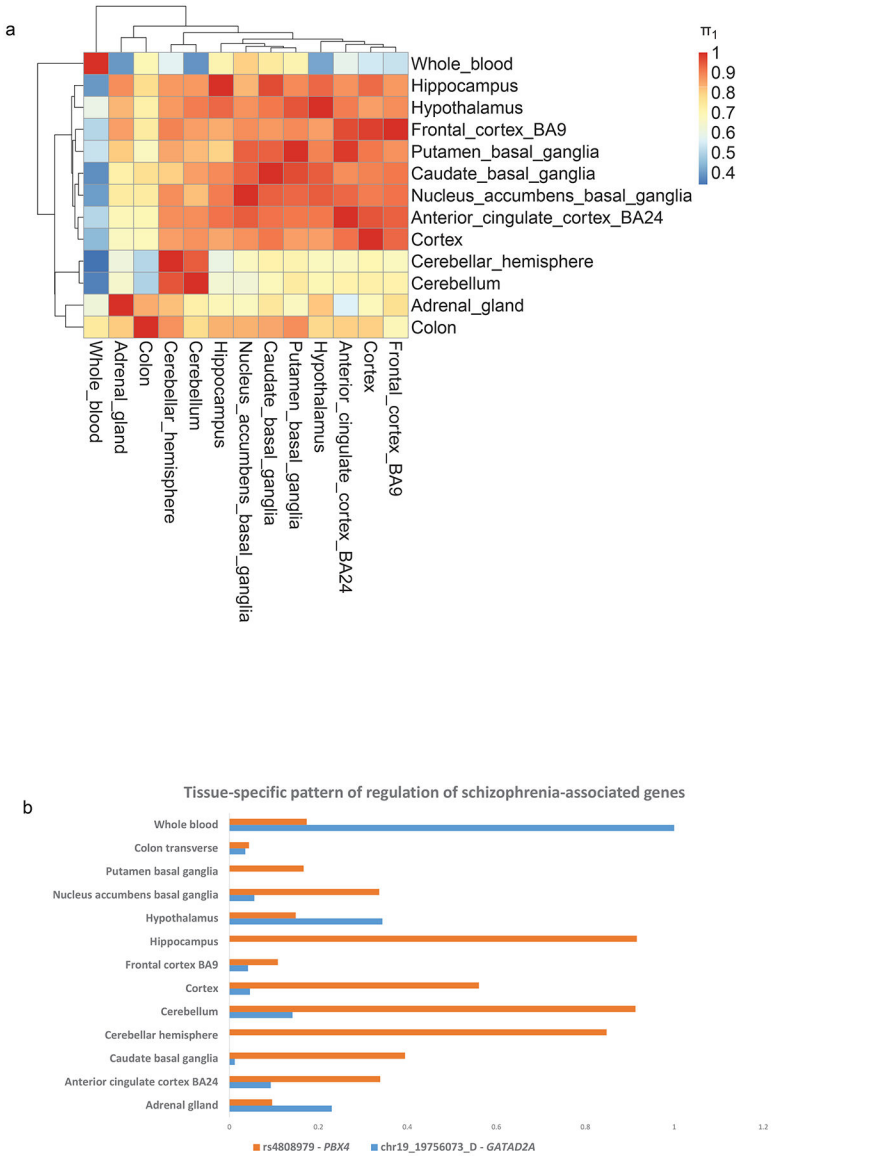

Figure 3 |. Complexity of identifying the relevant gene mechanism in tissue of pathology.

a, There is extensive eQTL sharing among tissues, as estimated by $\widehat{\pi_1}$. Note that the easily accessible tissue whole blood is a clear outlier, showing the least amount of sharing with the other tissues. Among the brain regions, cerebellum and cerebellar hemisphere are outliers. Sample sizes for the tissues can be found in Supplementary Table 1. **b**, Illustration of tissue-specific regulation from the METASOFT analysis. The b-eQTL of *GATA2DA* (rs2905432) impacts its expression only in whole blood, while the b-eQTL of *PBX4* (chr19_19756073_D) influences its expression in cerebellum/cerebellar hemisphere and in hippocampus, but not in the remaining tissues. These examples indicate that delineating the biological mechanisms of schizophrenia based on eQTL information requires a tissue-specific approach for at least some of the eQTL-gene associations.

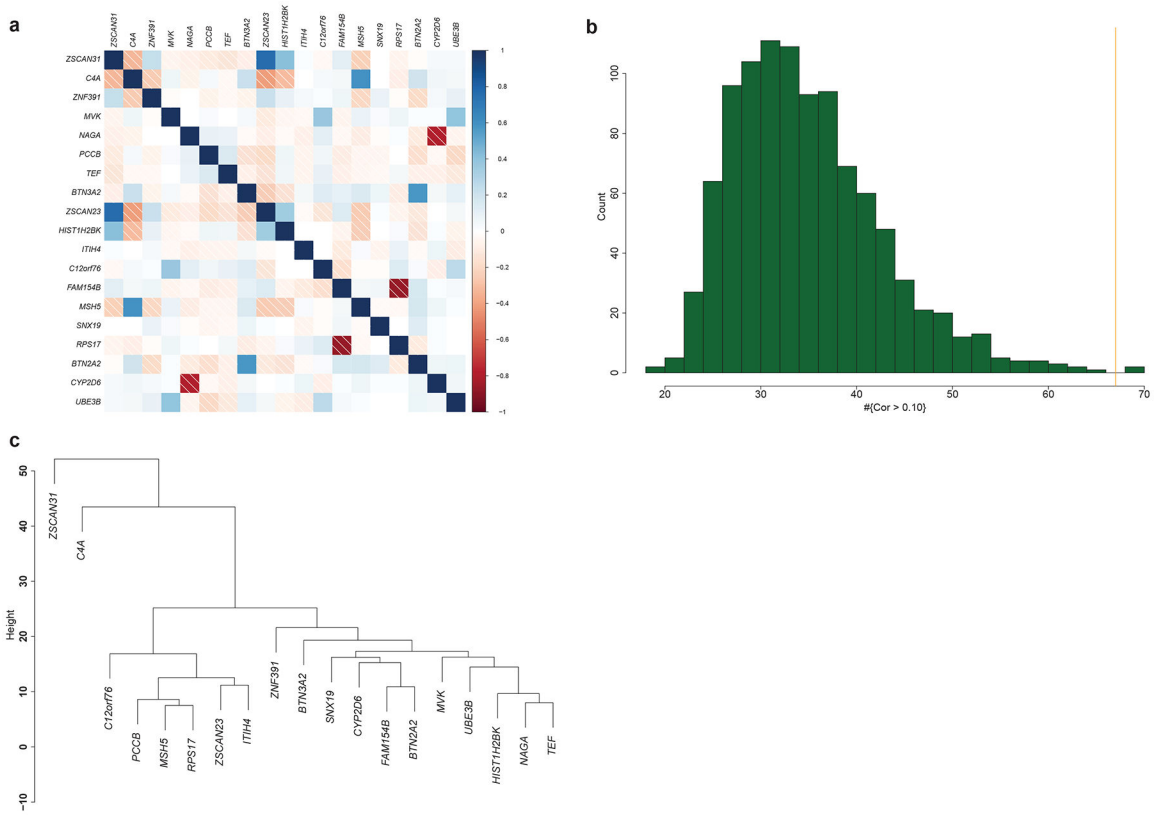

Figure 4 |. Genetically determined co-expression networks of disease-associated GReX.

a, We observed significant correlations in GReX of disease-associated genes (Bonferroni-adjusted P <0.05 from S-PrediXcan analysis of schizophrenia) for all tissues. Shown here are the pairwise GReX correlations in putamen basal ganglia, which had the largest $\widehat{\pi_1}$ for schizophrenia (40,675 cases, 64,643 controls). **b**, The observed GReX correlations were significantly greater (empirical P < 0.001) than expected based on 1,000 randomly generated, genetically defined co-expression networks. **c**, Hierarchical clustering of GReX of disease-associated genes in putamen basal ganglia was performed and identified several genetically determined clusters that appear to have coordinated expression including on distinct chromosomes.

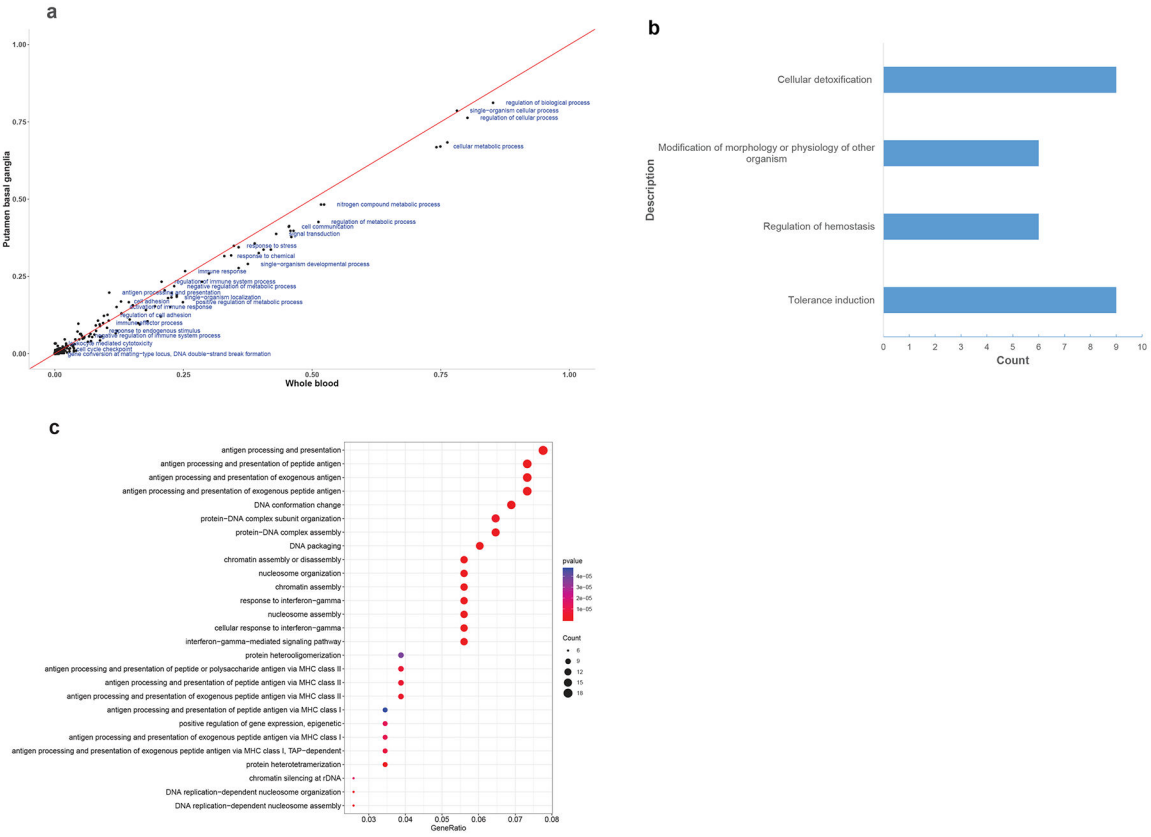

Figure 5 |. Tissue-specific GReX, joint-tissue eQTL mapping, and tissue-specific or tissue-shared functional categories for schizophrenia.

a, Substantial concordance between tissues (for example, between whole blood ($n = 338$) and putamen basal ganglia ($n = 82$), Spearman $\rho = 0.92$, $P = 3.86 \times 10^{-164}$) was observed in the proportion of disease-associated GReX levels (Bonferroni-adjusted $P < 0.05$ from S-PrediXcan analysis) that map to a biological process (as defined by Gene Ontology). **b**, List of biological processes present only in whole blood and represented by five or more disease-associated GReX levels in the tissue. **c**, Enriched biological processes (hypergeometric test, Bonferroni-adjusted $P < 0.05$) in putamen basal ganglia ($n = 82$), which had the largest $\widehat{\pi_1}$ for schizophrenia, and the gene ratio for the genes that map to the processes.

Table 1 |

Number of gene discoveries from PrediXcan analyses and an overview of the number of significant associations in brain, difficult-to-acquire, and non-brain tissues

	Schizophrenia	Bipolar disorder	ADHD	Broad depression
Tissue*	***n***	***n***	***n***	***n***
Dorsolateral prefrontal cortex	126	8	4	14
Adrenal gland	60	2	4	4
Anterior cingulate cortex BA24	35	2	2	3
Caudate basal ganglia	38	3	3	6
Cerebellar hemisphere	45	4	4	4
Cerebellum	68	5	4	8
Cortex	40	2	3	4
Frontal cortex BA9	35	2	4	5
Hippocampus	21	0	2	3
Hypothalamus	27	2	2	3
Nucleus accumbens basal ganglia	37	0	3	1
Putamen basal ganglia	29	1	3	2
Colon transverse	74	4	4	6
Whole blood	94	5	4	13
n unique genes “total”	275	13	12	31
n unique genes “novel PrediXcan”	46	5	8	11
n genes “difficult-to-acquire” (%)	242 (88)	12 (92)	12 (100)	24 (77)
n genes “brain” (%)	204 (74)	7 (54)	11 (92)	21 (66)
n genes “non-brain” (%)	91 (33)	7 (54)	3 (25)	12 (39)
n genes “difficult-to-acquire-only” (%)	207 (75)	10 (77)	11 (92)	23 (74)
n genes “brain-only” (%)	113 (41)	6 (46)	9 (75)	9 (29)
n genes “non-brain only” (%)	71 (26)	6 (46)	1 (8)	10 (32)
n genes “whole blood” (%)	68 (25)	3 (23)	1 (8)	8 (26)

n unique genes “total” = number of unique genes that are significant in ≥1 tissue; *n* unique genes “novel PrediXcan” = number of unique genes that are not in LD ($r^2 > 0.10$) with a GWAS index SNP; *n* genes “difficult-to-acquire tissue” = number of genes that are significant in ≥1 difficult-to-acquire tissue; *n* genes “brain” = number of genes that are significant in ≥1 brain tissue; *n* genes “non-brain” = number of genes that are significant in whole blood, adrenal gland or colon; *n* genes “brain-only” = number of genes that are significant in ≥1 brain tissue and not significant in any non-brain tissue; *n* unique genes “non-brain only” = number of genes that are significant in whole blood, adrenal gland or colon, but not significant in any of the brain tissues; *n* unique genes “whole blood” = number of genes that are significant in whole blood.

*Sample sizes for the tissues can be found in Supplementary Table 1.