Gene expression imputation across multiple brain regions provides insights into schizophrenia risk.

Authors:

Laura M. Huckins[1,2,3,4], Amanda Dobbyn[1,2], Douglas M. Ruderfer[5], Gabriel Hoffman[1,4], Weiqing Wang[1,2], Antonio F. Pardiñas[6] , Veera M Rajagopal[7,8,9], Thomas D. Als[7,8,9], Hoang Nguyen[1,2], Kiran Girdhar[1,2], James Boocock[10], Panos Roussos[1,2,3,4], Menachem Fromer[1,2], Robin Kramer[11], Enrico Domenici[12], Eric R. Gamazon[5,13], Shaun Purcell[1,2,4], CommonMind Consortium[14], the Schizophrenia Working Group of the Psychiatric Genomics Consortium[14], iPSYCH-GEMS Schizophrenia Working Group[14], Ditte Demontis[7,8,9], Anders D. Børglum[7,8,9], James T.R. Walters[6], Michael C. O'Donovan[6], Patrick Sullivan[15,16], Michael J. Owen[6], Bernie Devlin[17], Solveig K. Sieberts[18], Nancy J. Cox[5], Hae Kyung Im[19], Pamela Sklar[1,2,3,4], Eli A. Stahl[1,2,3,4]

Author Affiliations:

1. Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, NYC, NY;
2. Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, NYC, NY;
3. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, NYC, NY;
4. Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, NYC, NY, USA;
5. Vanderbilt University Medical Center, Nashville, TN;
6. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK;
7. Department of Biomedicine, Aarhus University, Aarhus, Denmark
8. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark.
9. Center for Integrative Sequencing, Aarhus University, Aarhus, Denmark.
10. Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA;
11. Human Brain Collection Core, National Institute of Mental Health, Bethesda, MD, USA;
12. Laboratory of Neurogenomic Biomarkers, Centre for Integrative Biology (CIBIO), University of Trento, Trento, Italy;
13. Clare Hall, University of Cambridge, Cambridge, United Kingdom
14. A list of members and affiliations appears at the end of the paper.
15. University of North Carolina at Chapel Hill, NC, USA
16. Karolinska Institutet, Stockholm, Sweden
17. Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA;
18. Systems Biology, Sage Bionetworks, Seattle, WA, USA;
19. Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA

Abstract

Transcriptomic imputation approaches combine eQTL reference panels with large-scale genotype data in order to test associations between disease and gene expression. These genic associations could elucidate signals in complex GWAS loci and may disentangle the role of different tissues in disease development. We use the largest eQTL reference panel for the dorso-lateral pre-frontal cortex (DLPFC) to create a set of gene expression predictors, and demonstrate their utility. We applied DLPFC and 12 GTEx-brain predictors to 40,299 schizophrenia cases and 65,264 matched controls for a large transcriptomic imputation study of schizophrenia. We identified 413 genic associations across 13 brain regions. Stepwise conditioning identified 67 non-MHC genes, of which 14 did not fall within previous GWAS loci. We identified 36 significantly enriched pathways, including hexosaminidase-A deficiency, and multiple porphyric disorder pathways. We investigated developmental expression patterns among the 67 non-MHC genes, and identified specific groups of pre- and post-natal expression.

Introduction

Genome-wide association studies (GWAS) have yielded large lists of disease-associated loci. Progress in identifying the causal variants driving these associations, particularly for complex psychiatric disorders such as schizophrenia, has lagged much further behind. Interpreting associated variants and loci is therefore vital to understanding how genetic variation contributes to disease pathology. Expression Quantitative Trait Loci (eQTLs), which are responsible for a substantial proportion of gene expression variance, have been posited as a link between associated loci and disease susceptibility[1–5], and have yielded results for a host of complex traits[6–9]. Consequently, numerous methods to identify and interpret co-localization of eQTLs and GWAS loci have been developed[10–13]. However, these methods require simplifying assumptions about genetic architecture (i.e., one causal variant per GWAS locus) and/or linkage disequilibrium, may be underpowered or overly conservative, especially in the presence of allelic heterogeneity, and have not yet yielded substantial insights into disease biology.

Biologically relevant transcriptomic information can be extracted through detailed RNA-sequencing, as recently described by the CommonMind Consortium[14] (CMC) in a large cohort of genotyped individuals with schizophrenia and bipolar disorder[14]. These analyses however are underpowered to detect statistically significant differential expression of genes mapping at schizophrenia (SCZ) risk loci, due to the small effects predicted by GWAS, combined with the difficulty of obtaining adequate sample sizes of neurological tissues[14], and do not necessarily identify all risk variation in GWAS loci. Transcriptomic imputation is an alternative approach that leverages large eQTL reference panels to bridge the gap between large-scale genotyping studies and biologically useful transcriptome studies[15,16]. Transcriptomic imputation approaches codify the relationships between genotype and gene expression in matched panels of individuals, then impute the genetic component of the transcriptome into large-scale genotype-only datasets, such as case-control GWAS cohorts, enabling investigation of disease-associated gene expression changes. This will allow us to study genes with modest effect sizes, likely representing a large proportion of genomic risk for psychiatric disorders[14,17].

The large collection of dorso-lateral pre-frontal cortex (DLPFC) gene expression data collected by the CMC [14] affords us a unique opportunity to study and codify relationships between genotype and gene expression. Here, we present a novel set of gene expression predictor models, built using CMC DLPFC data[14]. We compare different regression approaches to building these models (including elastic net[15], Bayesian sparse linear mixed models and ridge regression[16], and using max eQTLs), and benchmark performance of these predictors against existing GTEx prediction models. We applied our CMC DLPFC predictors and 12 GTEx-derived neurological prediction models to predict gene expression in schizophrenia GWAS data, obtained through collaboration with the Psychiatric Genomics Consortium (PGC) schizophrenia working group, the "CLOZUK2" cohort, and the iPSYCH-GEMS schizophrenia working group. We identified 413 genome-wide significant genic associations with schizophrenia in our PGC+CLOZUK2 sample, constituting 67 independent associations outside the MHC region. We demonstrated the relevance of these associations to schizophrenia etiopathology using gene set enrichment analysis, and by examining the effects of manipulation of these genes in mouse models. Finally, we investigated spatio-temporal expression of these genes using a developmental transcriptome dataset, and identified distinct spatio-temporal patterns of expression across our associated genes.

Results

Prediction Models based on CMC DLPFC expression

Using matched CMC genotype and gene expression data, we developed DLPFC genetically regulated gene expression (GREX) predictor models. We systematically compared four approaches to building predictors[15,16] within a cross-validation framework. Elastic net regression had a higher distribution of cross-validation R^2 (R_{CV}^2) and higher mean R_{CV}^2 values (Supplementary Figure 1, 2a) than all other methods. We therefore used elastic net regression to build our prediction models. We compared prediction models created using elastic net regression on SVA-corrected and uncorrected data[14]. The distribution of R_{cv}^2 values for the SVA-based models was significantly higher than for the un-corrected data[14,18] (ks-test; $p < 2.2 \times 10^{-16}$; Supplementary figure 1b,c). In total, 10,929 genes were predicted with elastic net cross-validation $R_{cv}^2 > 0.01$ in the SVA-corrected data and were included in the final predictor database (mean $R_{cv}^2 = 0.076$).

To test the predictive accuracy of the CMC-derived DLPFC models, and to benchmark this against existing GTEx-derived prediction models, GREX was calculated in an independent DLPFC RNA-sequencing dataset (the Religious Orders Study Memory and Ageing Project, ROSMAP[19,20]). We compared predicted GREX to measured ROSMAP gene expression for each gene (Replication $R^{2,}$ or R_R^2) for the CMC-derived DLPFC models and twelve GTEx-derived brain tissue models[15,21] (Figure 1, Supplementary Figure 2b). CMC-derived DLPFC models had higher average R_R^2 values (Mean $R_R^2 = 0.056$), more genes with $R_R^2 > 0.01$, and significantly higher overall distributions of R_R^2 values than any of the twelve GTEx models (ks-test, $p < 2.2 \times 10^{-16}$ across all analyses; Figure 1). Median R_R^2 values were significantly correlated with sample size of the original tissue set (rho = 0.92, $p = 7.2 \times 10^{-6}$), the number of genes in the prediction model (rho = 0.9, $p = 2.6 \times 10^{-5}$), and the number of significant 'eGenes' in each tissue type (rho = 0.95, $p = 5.5 \times 10^{-7}$; Figure 1c). Notably, these correlations persist after removing obvious outliers (Figure 1c).

To estimate trans-ancestral prediction accuracy, GREX was calculated for 162 African-American individuals and 280 European individuals from the NIMH Human Brain Collection Core (HBCC) dataset (Supplementary Figure 2c). R_R^2 values were higher on average in Europeans than African-Americans (average $R_{R\_EUR}^2 = 0.048$, $R_{R\_AA}^2 = 0.040$), but were significantly correlated between African-Americans and Europeans (rho = 0.78, $p < 2.2 \times 10^{-16}$, Pearson test; Supplementary Figure 3).

Application of Transcriptomic Imputation to Schizophrenia

We used CMC DLPFC and 12 GTEx–derived brain tissue prediction models to impute GREX of 19,661 unique genes in cases and controls from the PGC-SCZ GWAS study[22]. Predicted expression levels were tested for association with schizophrenia. Additionally, we applied CMC and GTEx-derived prediction models to summary statistics from 11 PGC cohorts (for which raw genotypes were unavailable) and the CLOZUK2 cohort. Meta-analysis was carried out across all PGC-SCZ and CLOZUK2 cohorts using an inverse-variance based approach in METAL. Our final analysis included 40,299 cases and 65,264 controls (Supplementary Figure 4a).

We identified 413 genome-wide significant associations, representing 256 genes in 13 tissues (Figure 2a). The largest number of associations was detected in the CMC-DLPFC GREX data (Figure 2c; 49 genes outside the MHC, 69 genes overall). We sought replication of our CMC DLPFC SCZ-associations in an independent dataset of 4,133 cases and 24,788 controls in collaboration with the iPSYCH-GEMS SCZ working group (Supplementary Figure 4b). We tested for replication of all Bonferroni-significant genes identified in our CMC-DLPFC analysis. Twelve out of 100 genes replicated in the iPSYCH-GEMS data, significantly more than expected by chance (binomial test, p = 0.0043). Notably, 11/12 replicating loci are previous GWAS loci, compared to 38/88 non-replicating loci. There was significant concordance between our discovery (PGC+CLOZUK2) and replication (iPSYCH-GEMS) samples; 72/100 genes have consistent direction of effect, including all 12 replicating genes (binomial $p = 1.258 \times 10^{-5}$), and we found significant correlation of effect sizes ($p = 1.784 \times 10^{-4}$; rho = 0.036) and –log10 p-values ($p = 1.073 \times 10^{-5}$; rho = 0.043).

To identify the top independent associations within genomic regions, which include multiple associations for a single gene across tissues, or multiple nearby genes, we partitioned genic associations into 58 groups defined based on genomic proximity and applied stepwise forward conditional analysis within each group (Supplementary Table 1). In total, 67 non-MHC genes remained genome-wide significant after conditioning (Table 1; Figure 2a,b). The largest signal was identified in the CMC-DLPFC GREX data (24 genes; Figure 2c), followed by the Putamen (7 genes). Nineteen out of 67 genes did not lie within 1 Mb of a previously genome-wide significant GWAS locus[22] (shown in bold, Table 1); of these, 5/19 genes were within 1 Mb of a locus which approached genome-wide significance ($p < 5 \times 10^{-07}$). The remaining 14 genes all fall within nominally significant PGC-SCZ GWAS loci ($p < 8 \times 10^{-04}$), but did not reach genome-wide significance.

We compared our CMC-DLPFC prediXcan associations statistics to COLOC results from our recent study[23,24]. Briefly, COLOC tests for co-localization between GWAS loci and eQTL architecture. We calculated COLOC probabilities of no-colocalization ("PP3") and colocalization ("PP4"); we consider PP4 > 0.5 to be significant evidence of colocalization[25]. We found a significant correlation between prediXcan p-values and PP4 values; rho = 0.35, $p = 2.3 \times 10^{-311}$. Thirty-one genes had 'strong' evidence of co-localization between GWAS loci and lead or conditional eQTLs[23]; of these, 21 were genome-wide significant in our prediXcan analysis (significantly more than expected by chance, binomial p-value = 2.11×10^{-104}), and all had $p < 1 \times 10^{-4}$. We identified 40 GWAS loci with no significant prediXcan associations; all of these loci also had strong evidence for no co-localization in our COLOC analysis (median PP3 = 0.936, median PP4 = 0.0027).

Implicated genes highlight SCZ-associated molecular pathways

We tested for overlap between our non-MHC SCZ-associated genes and 8,657 genesets comprising 1) hypothesis-driven pathways and 2) general molecular database pathways. We corrected for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) correction[26].

We identified three significantly associated pathways in our hypothesis-driven analysis (Table 2). Targets of the fragile-X mental retardation protein formed the most enriched pathway (FMRP; $p = 1.96 \times 10^{-8}$). Loss of FMRP inhibits synaptic function, is comorbid with autism spectrum disorder, and causes intellectual disability, as well as psychiatric symptoms including anxiety, hyperactivity and social deficits[27]. Enrichment of this large group of genes has been observed frequently in studies of schizophrenia[28,29] and autism[27,30]. There was a significant enrichment among our SCZ-associated genes and genes that have been shown to be intolerant to loss-of-function mutations[31] ($p = 5.86 \times 10^{-5}$) as well as with copy number variants (CNVs) associated with bipolar disorder[32] ($p = 7.92 \times 10^{-8}$), in line with a recent GWAS study of the same individuals[29].

Next, we performed an agnostic search for overlap between our schizophrenia-associated genes and ~ 8,500 molecular pathways collated from large, publicly available databases. Thirty-three pathways were significantly enriched after FDR correction (Table 2, Suppl. Table 2), including a number of pathways with some prior literature in psychiatric disease. We identified an enrichment with porphyrin metabolism ($p = 1.03 \times 10^{-4}$). Deficiencies in porphyrin metabolism lead to "Porphyria", an adult-onset metabolic disorder with a host of associated psychiatric symptoms, in particular episodes of violence and psychosis[33–38]. Five pathways potentially related to porphyrin metabolism, regarding abnormal iron level in the spleen, liver and kidney are also significantly enriched, including 2/5 of the most highly enriched pathways ($p < 2.0 \times 10^{-4}$). The PANTHER and REACTOME pathways for Heme biosynthesis and the GO pathway for protoporphyrinogen IX metabolic process, which are implicated in the development of porphyric disorders, are also highly enriched ($p = 2.2 \times 10^{-4}$, 2.6×10^{-4}, 4.1×10^{-4}), although do not pass FDR-correction.

Hexosaminidase activity was enriched ($p = 3.47 \times 10^{-5}$) in our results; this enrichment is not driven by a single highly-associated gene; rather, every single gene in the HEX-A pathway is nominally significant in the SCZ association analysis (Supplementary Table 2). Deficiency of hexosaminidase A (HEX-A) results in serious neurological and mental problems, most commonly presenting in infants as "Tay-Sachs" disease[39]. Adult-onset HEX-A deficiency presents with neurological and psychiatric symptoms, notably including onset of psychosis and schizophrenia[40]. Five pathways corresponding to Ras- and Rab- signaling, protein regulation and GTPase activity were enriched ($p < 6 \times 10^{-5}$). These pathways have a crucial role in neuron cell differentiation[41] and migration[42], and have been implicated in the development of schizophrenia and autism[43–46]. We also find significant enrichment with protein phosphatase type 2A regulator activity ($p = 5.24 \times 10^{-5}$), which was associated with major depressive disorder (MDD) and across MDD, bipolar disorder (BPD) and SCZ in the same large integrative analysis[47], and has been implicated in antidepressant response and serotonergic neurotransmission[48].

GREX associations are consistent with functional validation

To test the functional impact of our SCZ-associated predicted gene expression changes (GREX), we performed two *in-silico* analyses. First, we compared differentially expressed genes in the Fromer et al. CMC analysis to DLPFC prediXcan results; 76/460 are nominally significant in the DLPFC prediXcan analysis, significantly more than would be expected by chance (binomial test, $p = 8.75 \times 10^{-20}$). In particular, the Fromer et al. analysis highlighted six loci where expression levels of a single gene putatively affected schizophrenia risk. All six of these genes are nominally significant in our DLPFC analysis, and two (*CLCN3* and *FURIN*) reach genome-wide significance. In the conditional analysis across all brain regions, one additional gene (*SNX19*) reaches genome-wide significance. The direction of effect for all six genes matches the direction of gene expression changes observed in the original CMC paper, indicating that gene expression estimated in the imputed transcriptome reflects measured expression levels in brains of individuals with schizophrenia. Further, this observation is consistent with a model where the differential expression signature observed in CMC is caused by genetics rather than environment.

To understand the impact of altered expression of our 67 SCZ-associated genes, we performed an in-silico analysis of mouse mutants, by collating large, publicly available mouse databases[49–52]. We identified mutant mouse lines lacking expression of 37/67 of our SCZ-associated genes, and obtained 5,333 phenotypic data points relating to these lines, including 1,170 related to behavioral, neurological or craniofacial phenotypes. Twenty-five out of 37 genes were associated with at least one behavioral, neurological or related phenotype (Supplementary Table 3).

We carried out two tests to assess the rate of phenotypic abnormalities in SCZ-associated mouse lines. First, we compared the proportion of SCZ-gene lines with phenotypic abnormalities to the ‘baseline’ proportion across all mouse lines for which we had available data. SCZ-associated lines were significantly more likely to display any phenotype (paired t-test, $p = 0.009647$). Next, we repeated this analysis for genes identified in S-PrediXcan analyses of 66 publicly available

GWAS datasets. SCZ mouse lines had higher levels of nervous system (40.5% vs. 37.6%), behavioral (35.1% vs. 32.0%), and eye/vision phenotypes (29.7% vs. 17.0%) compared to these ‘baseline’ GWAS comparisons. SCZ mouse lines also had higher rates of embryonic phenotypes, usually indicative of homozygous lethality or mutations incompatible with life (27.0% vs. 21.1%).

Distinct pattern of SCZ risk throughout development

We assessed expression of our SCZ-associated genes throughout development using BRAINSPAN[53]. Data were partitioned into eight developmental stages (four pre-natal, four post-natal), and four brain regions[32,53] (Figure 3a). SCZ-associated genes were significantly co-expressed, in both pre-natal and post-natal development and in all four brain regions, based on local connectedness[54] (Figure 3b), global connectedness[54] (i.e., average path length between genes, Supplementary Figure 5), and network density (i.e., number of edges, Supplementary Figure 6). Examining pairwise gene expression correlation (Supplementary Figure 7) and gene co-expression networks (Supplementary Figure 8) for each spatiotemporal point indicated that the same genes do not drive this co-expression pattern throughout development; rather, it appears that separate groups of genes drive early pre-natal, late pre-natal and post-natal clustering.

To visualize this, we calculated Z scores measuring the spatio-temporal specificity of gene expression for each SCZ-associated gene, across all 32 time-points (Figure 4). Genes clustered into four groups (Supplementary Figure 9), with distinct spatio-temporal expression signatures. The largest cluster (Cluster A, Figure 4a; 29 genes) spanned early to late-mid pre-natal development (4-24 weeks post conception (pcw)), either across the whole brain (22 genes) or in regions 1-3 only (7 genes). Twelve genes were expressed in late pre-natal development (Figure 4d; 25-38 pcw); 10 genes were expressed in regions 1-3, post-natally and in the late pre-natal period (Figure 4c), and 15 genes were expressed throughout development (Figure 4b), either specifically in region four (nine genes) or throughout the brain (six genes).

In order to probe the biological relevance of our four BRAINSPAN clusters, we compared these gene lists to known and candidate gene sets with relevance to schizophrenia[55]. Genes in clusters A and B, (i.e., clusters with pre-natal expression) were involved in brain morphology and development, nervous system development, neuron development and morphology and synaptic development, function, and morphology (Supplementary Table 4). These associations were not seen in clusters C and D (i.e., genes with late pre-natal and post-natal expression).

We noticed a relationship between patterns of gene expression and the likelihood of behavioral, neurological or related phenotypes in our mutant mouse model database. Mutant mice lacking genes expressed exclusively pre-natally in humans, or genes expressed pre- and post-natally, were more likely to have any behavioral or neurological phenotypes than mutant mice lacking expression of genes expressed primarily in the third trimester or post-natally ($p = 1.7 \times 10^{-4}$) (Supplementary Figure 10).

Discussion

In this study, we present DLPFC gene expression prediction models, constructed using CommonMind Consortium genotype and gene expression data. These prediction models may be applied to either raw data or summary statistics, in order to yield tissue-specific gene expression information in large data sets. This allows researchers to access transcriptome data for non-peripheral tissues, at scales currently prohibited by the high cost of RNA sequencing, and circumventing distortions in measures of gene expression stemming from errors of measurement or environmental influences. As disease status may alter gene expression but not the germline profile, analyzing genetically regulated expression ensures that we identify only the causal direction of effect between gene expression and disease[15]. Large, imputed transcriptomic datasets represent the first opportunity to study the role of subtle gene expression changes (and therefore modest effect sizes) in disease development.

There are some inherent limitations to this approach. The accuracy of transcriptomic imputation is reliant on access to large eQTL reference panels, and it is therefore vital that efforts to collect and analyze these samples continue. Transcriptomic inputation has exciting advantages for gene discovery as well as downstream applications[15,56,57]; however, the relative merits of existing methodologies are as yet under-explored. Here, sparser elastic net models better captured gene expression regulation than BSLMM; at the same time, the improved performance of elastic net over max-eQTL models suggests that a single eQTL model is over-simplified[2,15]. Fundamentally, transcriptomic imputation methods model only the genetically regulated portion of gene expression, and so cannot capture or interpret variance of expression induced by environment or lifestyle factors, which may be of particular importance in psychiatric disorders. Given the right study design, analyzing genetic components of expression together with observed expression could open doors to better study the role of gene expression in disease.

Sample size and tissue matching contribute to accuracy of transcriptomic imputation results. Our CMC-derived DLPFC prediction models had higher average validation R^2 values in external DLPFC data than GTEx-derived brain tissue models. Notably, the model with the second highest percent of genes passing the R^2 threshold is the Thyroid, which has the largest sample size among the GTEx brain prediction models. When looking at mean R^2 values, the second highest value comes from the GTEx Frontal Cortex, despite the associated small sample size, implying at least some degree of tissue specificity of eQTLs architecture.

We compared transcriptomic imputation accuracy in European and African-American individuals, and found that our models were applicable to either ethnicity with only a small decrease in accuracy. Common SNPs shared across ethnicities have important effects on gene expression, and as such we expect GREX to have consistency across populations. There is a well-documented dearth of exploration of genetic associations in non-European cohorts[58,59]. We believe that these analyses should be carried out in non-European cohorts.

We applied the CMC-DLPFC and GTEx-derived prediction models to schizophrenia cases and controls from the PGC2 and CLOZUK2 collections, constituting a large transcriptomic analysis of schizophrenia. Predicted gene expression levels were calculated for 19,661 unique genes across brain regions (Figure 1c) and tested for association with SCZ case-control status. We identified 413 significant associations, constituting 67 independent associations. We found significant replication of our CMC DLPFC associations in a large independent replication cohort, in collaboration with the iPSYCH-GEMS consortium. Our prediXcan results were significantly correlated with co-localization estimates (“PP4”) from COLOC. Importantly, GWAS loci with no significant prediXcan associations also had no evidence for co-localization with eQTLs. Together, these results imply that our prediXcan associations identify genes with good evidence for colocalization between GWAS and eQTL architecture, and are not contaminated by linkage disequilibrium (LD). One caveat is that four of our associations (*SNX19*, *NAGA*, *TYW5*, and *GNL3*) have no evidence for colocalization in COLOC results, or following visual inspection of local GWAS and eQTL architecture, and may be false positives.

We compared our CMC DLPFC associations to results using a single-eQTL based method, SMR[12], in the PGC+CLOZUK schizophrenia GWAS[60], which identified 12 genome-wide significant associations. All significant SMR associations were also significant in our DLPFC prediXcan analysis, and all directions of effect were concordant between the two studies. A recent TWAS study of 30 GWAS summary statistic traits[56] identified 38 non-MHC genes associated at tissue-level significance with SCZ in CMC- and GTEx-derived brain tissues (i.e., matching those used in our study). Of these, 26 also reach genome-wide significance in our study, although in many instances these genes are not identified as the lead independent associated gene following our conditional analysis. Among our 67 SCZ-associated genes, 19 were novel, i.e. did not fall within 1 Mb of a previous GWAS locus (including 5/7 of the novel brain genes identified in the recent TWAS analysis).

We used conditional analyses to identify independent associations within loci. These analyses clarify the most strongly associated genes and tissues (Table 1), while we note that nearly co-linear gene-tissue pairs could also represent causal associations. The tissues highlighted allowed us to tabulate apparently independent contributions to SCZ risk from different brain regions, even though their transcriptomes are highly correlated generally. We find DLPFC and Cerebellum effects, as well as from Putamen, Caudate and Nucleus Accumbens Basal Ganglia. One caveat here is that tissue-associations are likely driven by sample size of the eQTL reference panel, as well as biology. It is likely that the large sample size of the DLPFC reference panel contributes partially to the greater signal identified in the DLPFC.

We used these genic associations to search for enrichments with molecular pathways and gene sets, and identified 36 significant enriched pathways. Among novel pathways, we identified a significant association with HEX-A deficiency. Despite the well-studied and documented symptomatic overlap between adult-onset HEX-A deficiency and schizophrenia, we believe that this is the first demonstration of shared genetics between the disorders. Notably, this overlap is not driven by a single highly-associated gene which is shared by both disorders; rather, every single gene in the HEX-A pathway is nominally significant in the SCZ association analysis, and five genes have $p < 1 \times 10^{-3}$, indicating that there may be substantial shared genetic etiology between the two disorders that warrants further investigation. Additionally, we identified a significant overlap between our SCZ-associated genes and a number of pathways associated with porphyrin metabolism. Porphyric disorders have been well characterized and are among early descriptions of "schizophrenic" and psychotic presentations of schizophrenia, as described in the likely eponymous mid-19th century poem "Porphyria's Lover", by Robert Browning[61], and have been cited as a likely diagnosis for the various psychiatric and metabolic ailments of Vincent van Gogh[62–67] and King George III[68].

Finally, we assessed patterns of expression for the 67 SCZ-associated genes throughout development using spatio-temporal transcriptomic data obtained from BRAINSPAN. We identified four clusters of genes, with expression in four distinct spatiotemporal regions, ranging from early pre-natal to strictly post-natal expression. There are plausible hypotheses and genetic evidence for SCZ disease development in adolescence, given the correlation with age of onset, as well as prenatally, supported by genetic overlap with neurodevelopmental disorders[69–71] as well as the earlier onset of cognitive impairments[72–75]. Understanding the temporal expression patterns of SCZ-associated genes can help to elucidate gene development and trajectory, and inform research and analysis design. Identification of SCZ-associated genes primarily expressed prenatally is notable given our adult eQTL reference panels, and may reflect common eQTL architecture across development, which is known to be partial[76–78]; therefore, our results should spur interest in extending transcriptomic imputation data and/or methods to early development[76]. Identification of SCZ-associated genes primarily expressed in adolescence and adulthood is of particular interest for direct analysis of the brain transcriptome in adult psychiatric cases.

eQTL data have been recognized for nearly a decade as potentially important for understanding complex genetic variation. Nicolae et al.[1] showed that common variant-common disease associations are strongly enriched for genetic regulation of gene expression. Therefore, integrative approaches combining transcriptomic and genetic association data have great potential. Current transcriptomic imputation association analyses increase power for genetic discovery, with great potential for further development, including leveraging additional data types such as chromatin modifications[79] (e.g. methylation, histone modification), imputing different tissues or different exposures (e.g. age, smoking, trauma) and modeling trans/coexpression effects. It remains critical to leverage transcriptomic impuation associations to provide insights into specific disease mechanisms. Here, the accelerated identification of disease-associated genes allows the detection of novel pathways and distinct spatiotemporal patterns of expression in schizophrenia risk.

URLs

https://github.com/theboocock/coco/ "CoCo", an R implementation of GCTA-COJO.

gene2pheno.org Publicly-available whole-blood-derived S-PrediXcan results (as of March 2018,)

https://github.com/laurahuckins/CMC_DLPFC_prediXcan Our CMC-derived DLPFC prediction models are publicly available at

Acknowledgements

We dedicate this manuscript to the memory of Pamela Sklar, whose guidance and wisdom we miss daily. We strive to continue her legacy of thoughtful, innovative, and collaborative science.

Data were generated as part of the CommonMind Consortium supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881 and R37MH057881S1, HHSN271201300031C, AG02219, AG05138 and MH06692.

Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories and the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner, Barbara Lipska (NIMH).

ROSMAP study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute.

The iPSYCH-GEMS team would like to acknowledge funding from the Lundbeck Foundation (grant no R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, an Advanced Grant from the European Research Council (project no: 294838), the Danish Strategic Research Council the Novo Nordisk Foundation for supporting the Danish National Biobank resource, and grants from Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 09/05/16. BrainSpan: Atlas of the Developing Human Brain [Internet]. Funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01.

Hae Kyung Im was supported by R01 MH107666-01.

Author Contributions

LMH designed the study and specific sub-analyses, ran analyses, and wrote the manuscript. AD designed and ran analyses, and contributed to the writing group. DR contributed to study and analytical design, and writing. GEH contributed to analytical design and writing. WW, NTH, JB, designed and ran specific analyses. AFP, VR, TDA, KG, MF all ran specific analyses. SKS designed the study and analyses, and contributed to the writing group. PR, RK designed the study and contributed data. ED designed the study, contributed data, and contributed to the writing group. EG designed specific analyses and contributed to the writing group. SP designed the study. All three consortia (CMC, PGC-SCZ, iPSYCH-GEMS) contributed data. DD, AB, JW, MO'D, MO contributed data, advised on analyses, and contributed to the writing group. PFS advised on analyses and contributed to the writing group. BD designed the study, contributed data, advised on analyses, and contributed to the writing group. NC and HKI designed the study, advised on analyses, and contributed to the writing group. PS and EAS designed the study and specific analyses, ran analyses, and contributed to the writing group.

Potential Conflicts of Interest

ED has received research support from Roche in the period 2016-2018.

All other authors declare no conflicts of interest.

CommonMind Consortium Working Group

Jessica S Johnson 1, Hardik R Shah 2,4, Lambertus L Klein 17, Kristen K Dang 18, Benjamin A Logsdon 18, Milind C Mahajan 2,4, Lara M Mangravite 18, Hiroyoshi Toyoshiba 20, Raquel E Gur 21, Chang-Gyu Hahn 22, Eric Schadt 2,4, David A Lewis 17, Vahram Haroutunian 1,18,23,24, Mette A Peters 18, Barbara K Lipska 11, Joseph D Buxbaum 25, 26, Keisuke Hirai 27, Thanneer M Perumal 18, Laurent Essioux 28

20. Integrated Technology Research Laboratories, Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, Fujisawa, Kanagawa, Japan
21. Neuropsychiatry Section, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
22. Neuropsychiatric Signaling Program, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
23. Psychiatry, JJ Peters Virginia Medical Center, Bronx, New York, USA.
24. Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, USA
25. Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.
26. Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA
27. CNS Drug Discovery Unit, Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, Fujisawa, Kanagawa, Japan
28. F. Hoffman-La Roche Ltd

iPSYCH-GEMS SCZ working group

Anders D. Børglum 7,8,9, Ditte Demontis 7,8,9, Veera Manikandan Rajagopal 7,8,9, Thomas D. Als 7,8,9, Manuel Mattheisen 7,8,9, Jakob Grove 7,8,9,29, Thomas Werge 8,30,31, Preben Bo Mortensen 8,7,32,33, Carsten Bøcker Pedersen 8,32,33, Esben Agerbo 8, 32,33, Marianne Giørtz Pedersen 8, 32,33, Ole Mors 8,34, Merete Nordentoft 8, 35, David M. Hougaard 8,36, Jonas Bybjerg-Grauholm 8,36, Marie Bækvad-Hansen 8,36, Christine Søholm Hansen 8,36

29. Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark
30. Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark
31. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
32. National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark
33. Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark
34. Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark
35. Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark
36. Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

Potential Conflicts of Interest: TW has acted as advisor and lecturer to H. Lundbeck A/S

Schizophrenia Working Group of the Psychiatric Genomics Consortium

Stephan Ripke 37,38, Benjamin M. Neale 37,38,39,40, Aiden Corvin 41, James T. R. Walters 6, Kai-How Farh 37, Peter A. Holmans 6,42, Phil Lee 37,38,40, Brendan Bulik-Sullivan 37,38, David A. Collier 43,44, Hailiang Huang 37,39, Tune H. Pers 39,45,46, Ingrid Agartz 47,48,49, Esben Agerbo 8,32,33, Margot Albus 50, Madeline Alexander 51, Farooq Amin 52,53, Silviu A. Bacanu 54, Martin Begemann 55, Richard A. Belliveau Jr 38, Judit Bene 56,57, Sarah E. Bergen 38,58, Elizabeth Bevilacqua 38, Tim B. Bigdeli 54, Donald W. Black 59, Richard Bruggeman 60, Nancy G. Buccola 61, Randy L. Buckner 62,63,64, William Byerley 65, Wiepke Cahn 66, Guiqing Cai 2,3, Dominique Campion 67, Rita M. Cantor 10, Vaughan J. Carr 68,69, Noa Carrera 6, Stanley V. Catts 68,70, Kimberly D. Chambert 38, Raymond C. K. Chan 71, Ronald Y. L. Chen 72, Eric Y. H. Chen 72,73, Wei Cheng 15, Eric F. C. Cheung 74, Siow Ann Chong75, C. Robert Cloninger 76, David Cohen 77, Nadine Cohen 78, Paul Cormican 41, Nick Craddock 6,42, James J. Crowley 79, David Curtis 80,81, Michael Davidson 82, Kenneth L. Davis 3, Franziska Degenhardt 83,84, Jurgen Del Favero 85, Ditte Demontis 7,8,9, Dimitris Dikeos 86, Timothy Dinan 87, Srdjan Djurovic 49,88, Gary Donohoe 41,89, Elodie Drapeau 3, Jubao Duan 90,91, Frank Dudbridge 92, Naser Durmishi 93, Peter Eichhammer94, Johan Eriksson 95,96,97, Valentina Escott-Price 6, Laurent Essioux 98, Ayman H. Fanous 99,100,101,102, Martilias S. Farrell 79, Josef Frank 103, Lude Franke 104, Robert Freedman 105, Nelson B. Freimer 106, Marion Friedl 107, Joseph I. Friedman 3, Menachem Fromer 1,37,38,40, Giulio Genovese 38, Lyudmila Georgieva 6, Ina Giegling 107,108, Paola Giusti-Rodrı´guez 79, Stephanie Godard 109, Jacqueline I. Goldstein 37,39, Vera Golimbet 110, Srihari Gopal 111, Jacob Gratten 112, Lieuwe de Haan 113, Christian Hammer 55, Marian L. Hamshere 6, Mark Hansen 114, Thomas Hansen 8,30, Vahram Haroutunian 3,25,23, Annette M. Hartmann 107, Frans A. Henskens 68,115,116, Stefan Herms 83,84,117, Joel N. Hirschhorn 39,46,118, Per Hoffmann 83,84,117, Andrea Hofman 83,84, Mads V. Hollegaard 36, David M. Hougaard 36, Masashi Ikeda 119, Inge Joa 120, Antonio Julia 121, Rene S. Kahn 66, Luba Kalaydjieva 122,123, Sena Karachanak-Yankova 124, Juha Karjalainen 104, David Kavanagh 6, Matthew C. Keller 125, James L. Kennedy 126,127,128, Andrey Khrunin 129, Yunjung Kim 79, Janis Klovins 130, James A. Knowles 131, Bettina Konte 107, Vaidutis Kucinskas 132, Zita Ausrele Kucinskiene 132, Hana Kuzelova-Ptackova 133, Anna K. Kahler 58, Claudine Laurent 51,134, Jimmy Lee Chee Keong 75,135, S. Hong Lee 112, Sophie E. Legge 6, Bernard Lerer 136, Miaoxin Li 72,73,137, Tao Li 138, Kung-Yee Liang 139, Jeffrey Lieberman 140, Svetlana Limborska 129, Carmel M. Loughland 68,141, Jan Lubinski 142, Jouko Lonnqvist 143, Milan Macek Jr 133, Patrik K. E. Magnusson 58, Brion S. Maher 144, Wolfgang Maier 145, Jacques Mallet 146, Sara Marsal 121, Manuel Mattheisen 7,8,9,147, Morten Mattingsdal 49,148, Robert W. McCarley 149,150, ColmMcDonald 151, Andrew M. McIntosh 152,153, Sandra Meier 103, Carin J. Meijer 113, Bela Melegh 56,57, Ingrid Melle 49,154, Raquelle I. Mesholam-Gately 149,155, Andres Metspalu 156, Patricia T. Michie 68,157, Lili Milani 156, Vihra Milanova 158, Younes Mokrab 43, Derek W. Morris 41,89, Ole Mors 8,9,159, Kieran C. Murphy160, Robin M. Murray 161, Inez Myin-Germeys 162, Bertram Muller-Myhsok 163,164,165, Mari Nelis 156,

Igor Nenadic 166, Deborah A. Nertney 167, Gerald Nestadt 168, Kristin K. Nicodemus 169, Liene Nikitina-Zake 130, Laura Nisenbaum 170, Annelie Nordin 171, Eadbhard O’Callaghan172, ColmO’Dushlaine 38,F. Anthony O’Neill 173, Sang-YunOh 174, Ann Olincy 126, Line Olsen 8,64, Jim Van Os 162,175, Psychosis Endophenotypes International Consortium, Christos Pantelis 68,176, George N. Papadimitriou 86, Sergi Papiol 55, Elena Parkhomenko 3, Michele T. Pato 131, Tiina Paunio177,178, Milica Pejovic-Milovancevic179, Diana O. Perkins 180, Olli Pietiläinen 178,181, Jonathan Pimm 81, Andrew J. Pocklington 6, John Powell 161, Alkes Price 39,182, Ann E. Pulver 168, Shaun M. Purcell 1, Digby Quested 183, Henrik B. Rasmussen 30,43, Abraham Reichenberg 3, Mark A. Reimers 184, Alexander L. Richards6, Joshua L. Roffman 62,64, Panos Roussos 1,4, Douglas M. Ruderfer 1,5,6, Veikko Salomaa 97, Alan R. Sanders 90,91, Ulrich Schall 68,141, Christian R. Schubert 185, Thomas G. Schulze 103,186, Sibylle G. Schwab 187, Edward M. Scolnick 38, Rodney J. Scott 68,188,189, Larry J. Seidman 144,155, Jianxin Shi 190, Engilbert Sigurdsson 191, Teimuraz Silagadze 192, Jeremy M. Silverman 3,193, Kang Sim 75, Petr Slominsky 129, Jordan W. Smoller 38,40, Hon-Cheong So 72, Chris C. A. Spencer 194, Eli A. Stahl 1,2,3,4, Hreinn Stefansson 195, Stacy Steinberg 195, Elisabeth Stogmann 196, Richard E. Straub 197, Eric Strengman 66,198, Jana Strohmaier 103, T. Scott Stroup 140, Mythily Subramaniam 75, Jaana Suvisaari 143, Dragan M. Svrakic 76, Jin P. Szatkiewicz 79, Erik Soderman 47, Srinivas Thirumalai 199, Draga Toncheva 124, Sarah Tosato 200, JuhaVeijola 201,202, John Waddington 203, Dermot Walsh 204, Dai Wang 111, Qiang Wang 138, Bradley T. Webb 54, Mark Weiser 82, Dieter B. Wildenauer 205, Nigel M. Williams 6, Stephanie Williams 79, Stephanie H. Witt 103, Aaron R. Wolen 184, Emily H. M. Wong 72, Brandon K. Wormley 54, Hualin Simon Xi 206, Clement C. Zai 126,127, Xuebin Zheng 207, Fritz Zimprich 196, Naomi R. Wray 112, Kari Stefansson 195, Peter M. Visscher 112, Wellcome Trust Case-Control Consortium, Rolf Adolfsson 171, Ole A. Andreassen 49,154, Douglas H. R. Blackwood 153, Elvira Bramon 208, Joseph D. Buxbaum 2,3,24,25, Anders D. Børglum 7,8,9,159, Sven Cichon 83,84,117,209, Ariel Darvasi 210, Enrico Domenici 211, Hannelore Ehrenreich 55,Tonu Esko 39,46,118,156, Pablo V. Gejman 90,91, Michael Gill 41, Hugh Gurling 81, Christina M. Hultman 58, Nakao Iwata 119, Assen V. Jablensky 68,123,205,212, Erik G. Jonsson 47,49, Kenneth S. Kendler 213, George Kirov 6, Jo Knight 125,127,128, ToddLencz 214,215,216, Douglas F. Levinson 51, QingqinS. Li 111, Jianjun Liu 207,217, Anil K. Malhotra 214,215,216, Steven A. McCarroll 38,118, Andrew McQuillin 81, Jennifer L. Moran 38, Preben B. Mortensen 8,32,33, Bryan J. Mowry 112,218, Markus M. Nothen 83,84, Roel A. Ophoff 10,66,105, Michael J. Owen 6,42, Aarno Palotie 38,40,181, Carlos N. Pato 131, Tracey L. Petryshen 38,149,219, Danielle Posthuma 220,221,222, Marcella Rietschel 103, Brien P. Riley 213, Dan Rujescu 107,108, Pak C. Sham 72,73,137, Pamela Sklar 1,2,3,4,25 , David St Clair 223, Daniel R. Weinberger 197,224, Jens R. Wendland 185, Thomas Werge 8,30,225, Mark J. Daly 37,38,39, Patrick F. Sullivan 58,79,180 & Michael C. O’Donovan 6,42

37. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
38. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.
39. Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.
40. Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
41. Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Dublin 8, Ireland.
42. NationalCentre for Mental Health, Cardiff University, Cardiff CF244HQ,UK.
43. Eli Lilly and Company Limited, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey GU20 6PH, UK.
44. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, London SE58AF,UK.
45. Center for BiologicalSequence Analysis, Department of Systems Biology, Technical University of Denmark, DK-2800, Denmark.
46. Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children’s Hospital, Boston, Massachusetts 02115,USA.
47. Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, SE-17176 Stockholm, Sweden.
48. Department of Psychiatry, Diakonhjemmet Hospital, 0319 Oslo, Norway.
49. NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, 0424 Oslo, Norway.
50. State Mental Hospital, 85540 Haar, Germany.
51. Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA.
52. Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia 30033, USA.
53. Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia 30322, USA.
54. Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia 23298, USA.
55. Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Gottingen 37075, Germany.
56. Department of Medical Genetics, University of Pécs, Pécs H-7624, Hungary.
57. Szentagothai Research Center, University of Pécs, Pécs H-7624, Hungary.

58. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-17177, Sweden.
59. Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa 52242, USA.
60. University Medical Center Groningen, Department of Psychiatry, University of Groningen NL-9700 RB, The Netherlands.
61. School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana 70112, USA.
62. Athinoula A. Martinos Center, Massachusetts General Hospital, Boston, Massachusetts 02129, USA.
63. Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA.
64. Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
65. Department of Psychiatry, University of California at San Francisco, San Francisco, California 94143, USA.
66. University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, 3584 Utrecht, The Netherlands.
67. Centre Hospitalier du Rouvray and INSERM U1079 Faculty of Medicine, 76301 Rouen, France.
68. Schizophrenia Research Institute, Sydney NSW2010, Australia.
69. School of Psychiatry, University of New South Wales, Sydney NSW2031, Australia.
70. Royal Brisbane and Women’s Hospital, University of Queensland, Brisbane, St Lucia QLD 4072, Australia.
71. Institute of Psychology, Chinese Academy of Science, Beijing 100101, China.
72. Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.
73. State Key Laboratory for Brain and Cognitive Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.
74. Castle Peak Hospital, Hong Kong, China.
75. Institute of Mental Health, Singapore 539747, Singapore.
76. Department of Psychiatry, Washington University, St.Louis, Missouri 63110, USA.
77. Department of Child and Adolescent Psychiatry, Assistance Publique Hopitaux de Paris, Pierre and Marie Curie Faculty of Medicine and Institute for Intelligent Systems and Robotics, Paris 75013, France.
78. Blue Note Biosciences, Princeton, New Jersey 08540, USA
79. Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA.

80. Department of Psychological Medicine, Queen Mary University of London, London E1 1BB, UK.
81. Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, LondonWC1E6JJ, UK.
82. Sheba Medical Center, Tel Hashomer 52621, Israel.
83. Department of Genomics, Life and Brain Center, D-53127 Bonn, Germany.
84. Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany.
85. AppliedMolecular Genomics Unit, VIB Department of Molecular Genetics, University of Antwerp, B-2610 Antwerp, Belgium.
86. First Department of Psychiatry, University of Athens Medical School, Athens 11528, Greece.
87. Department of Psychiatry, University College Cork, Co. Cork, Ireland.
88. Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway.
89. Cognitive Genetics and Therapy Group, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Co. Galway, Ireland.
90. Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois 60637, USA.
91. Department of Psychiatry and Behavioral Sciences, North Shore University Health System, Evanston, Illinois 60201, USA.
92. Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.
93. Department of Child and Adolescent Psychiatry, University Clinic of Psychiatry, Skopje 1000, Republic of Macedonia.
94. Department of Psychiatry, University of Regensburg, 93053 Regensburg, Germany.
95. Department of General Practice, Helsinki University Central Hospital, University of Helsinki P.O. Box 20, Tukholmankatu 8 B, FI-00014, Helsinki, Finland
96. Folkhälsan Research Center, Helsinki, Finland, Biomedicum Helsinki 1,Haartmaninkatu 8, FI-00290, Helsinki, Finland.
97. National Institute for Health and Welfare, P.O. Box 30, FI-00271 Helsinki, Finland.
98. Translational Technologies and Bioinformatics, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland.
99. Department of Psychiatry, Georgetown University School of Medicine, Washington DC 20057, USA.
100. Department of Psychiatry, Keck School of Medicine of the University of Southern California, Los Angeles, California 90033, USA.
101. Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, Virginia 23298, USA.
102. Mental Health Service Line, Washington VA Medical Center, Washington DC 20422, USA.

103. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg , D-68159 Mannheim, Germany.
104. Department of Genetics, University of Groningen, University Medical Centre Groningen, 9700 RB Groningen, The Netherlands.
105. Department of Psychiatry, University of Colorado Denver, Aurora, Colorado 80045, USA.
106. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California 90095, USA.
107. Department of Psychiatry, University of Halle, 06112 Halle, Germany.
108. Department of Psychiatry, University of Munich, 80336, Munich, Germany.
109. Departments of Psychiatry and Human and Molecular Genetics, INSERM, Institut de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris 75013, France.
110. Mental Health Research Centre, Russian Academy of Medical Sciences, 115522 Moscow, Russia.
111. Neuroscience Therapeutic Area, Janssen Research and Development, Raritan, New Jersey 08869, USA.
112. Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, QLD 4072, Australia.
113. Academic Medical Centre University of Amsterdam, Department of Psychiatry, 1105 AZ Amsterdam, The Netherlands.
114. Illumina, La Jolla, California, California 92122, USA.
115. Priority Research Centre for Health Behaviour, University of Newcastle, Newcastle NSW 2308, Australia.
116. School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle NSW2308, Australia.
117. Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel CH-4058, Switzerland.
118. Department of Genetics, Harvard Medical School, Boston, Massachusetts, Massachusetts 02115, USA.
119. Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Aichi,470-1192, Japan.
120. Regional Centre for Clinical Researchin Psychosis, Department of Psychiatry, Stavanger University Hospital, 4011 Stavanger, Norway.
121. Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona 08035, Spain.
122. Centre for Medical Research, The University of Western Australia, Perth WA6009, Australia.
123. The Perkins Institute for Medical Research, The University of Western Australia, Perth WA6009, Australia.

124. Department of Medical Genetics, Medical University, Sofia 1431, Bulgaria.
125. Department of Psychology, University of Colorado Boulder, Boulder, Colorado 80309, USA.
126. Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario M5T 1R8, Canada.
127. Department of Psychiatry, University of Toronto, Toronto, Ontario M5T 1R8, Canada.
128. Institute of Medical Science,University of Toronto, Toronto,OntarioM5S1A8, Canada.
129. Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia.
130. Latvian Biomedical Research and Study Centre, Riga, LV-1067, Latvia.
131. Department of Psychiatry and Zilkha Neurogenetics Institute, Keck School of Medicine at University of Southern California, Los Angeles, California 90089, USA.
132. Faculty of Medicine, Vilnius University, LT-01513 Vilnius, Lithuania.
133. Department of Biology and Medical Genetics, 2nd Faculty of Medicine and University Hospital Motol, 150 06 Prague, Czech Republic.
134. Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris 75013, France.
135. Duke-NUS Graduate Medical School, Singapore 169857.
136. Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel.
137. Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China.
138. Mental Health Centre and Psychiatric Laboratory, West China Hospital, Sichuan University, Chengdu, 610041 Sichuan, China.
139. Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA.
140. Department of Psychiatry, Columbia University, New York, New York 10032, USA.
141. Priority Centre for Translational Neuroscience and Mental Health, University of Newcastle, Newcastle NSW 2300, Australia.
142. Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University in Szczecin, 70-453 Szczecin, Poland.
143. Department of Mental Health and Substance Abuse Services; National Institute for Health and Welfare, P.O.BOX30, FI-00271Helsinki, Finland.
144. Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA.
145. Department of Psychiatry, University of Bonn, D-53127 Bonn, Germany.
146. Centre National de la Recherche Scientifique, Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodénégératifs, Hôpital de la Pitiè-Salpêtrière 75013 Paris, France.
147. Department of Genomics Mathematics, University of Bonn, D-53127 Bonn, Germany.

148. Research Unit, Sørlandet Hospital, 4604 Kristiansand, Norway.
149. Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA.
150. VA Boston Health Care System, Brockton, Massachusetts 02301, USA.
151. Department of Psychiatry, National University of Ireland Galway, Co. Galway, Ireland.
152. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH16 4SB, UK.
153. Division of Psychiatry, University of Edinburgh, Edinburgh EH16 4SB, UK.
154. Division of Mental Health and Addiction, Oslo University Hospital, 0424 Oslo, Norway.
155. Massachusetts Mental Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Boston, Massachusetts 02114, USA.
156. Estonian Genome Center, University of Tartu, Tartu 50090, Estonia.
157. School of Psychology, University of Newcastle, Newcastle NSW2308, Australia.
158. First Psychiatric Clinic, Medical University, Sofia 1431, Bulgaria.
159. Department P, Aarhus University Hospital, DK-8240 Risskov, Denmark.
160. Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin 2, Ireland.
161. King's College London, London SE58AF, UK.
162. Maastricht University Medical Centre, South Limburg Mental Health Research and TeachingNetwork,EURON,6229HX Maastricht, The Netherlands.
163. Institute of Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK.
164. Max Planck Institute of Psychiatry, 80336 Munich, Germany.
165. Munich Cluster for Systems Neurology (SyNergy),80336 Munich, Germany.
166. Department of Psychiatry and Psychotherapy, Jena University Hospital, 07743 Jena, Germany.
167. Department of Psychiatry, Queensland Brain Institute and Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, St Lucia QLD 4072, Australia.
168. Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.
169. Department of Psychiatry, Trinity College Dublin, Dublin 2, Ireland.
170. Eli Lilly and Company, Lilly Corporate Center, Indianapolis, 46285 Indiana, USA.
171. Department of Clinical Sciences, Psychiatry, Umeå University, SE-901 87Umeå , Sweden.
172. DETECT Early Intervention Service for Psychosis, Blackrock, Co.Dublin, Ireland.
173. Centre for Public Health, Institute of Clinical Sciences, Queen's University Belfast, Belfast BT12 6AB, UK.
174. Lawrence Berkeley National Laboratory, University of California at Berkeley, Berkeley, California 94720,USA.
175. Institute of Psychiatry, King's College London, London SE5 8AF,UK.

176. Melbourne Neuropsychiatry Centre, University of Melbourne & Melbourne Health, Melbourne, Vic 3053, Australia.
177. Department of Psychiatry, University of Helsinki, P.O. Box 590, FI-00029 HUS, Helsinki, Finland.
178. Public Health Genomics Unit, National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland
179. Medical Faculty, University of Belgrade, 11000 Belgrade, Serbia.
180. Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-7160, USA.
181. Institute for Molecular Medicine Finland, FIMM, University of Helsinki, P.O. Box 20FI-00014, Helsinki, Finland
182. Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA.
183. Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK.
184. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA.
185. Pharma Therapeutics Clinical Research, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA.
186. Department of Psychiatry and Psychotherapy, University of Gottingen, 37073 Göttingen, Germany.
187. Psychiatry and Psychotherapy Clinic, University of Erlangen, 91054 Erlangen, Germany.
188. Hunter New England Health Service, Newcastle NSW2308, Australia.
189. School of Biomedical Sciences, University of Newcastle, Newcastle NSW2308, Australia.
190. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA.
191. University of Iceland, Landspitali, National University Hospital, 101 Reykjavik, Iceland.
192. Department of Psychiatry and Drug Addiction, Tbilisi State Medical University (TSMU), N33,0177 Tbilisi, Georgia.
193. Research and Development, Bronx Veterans Affairs Medical Center, New York, New York 10468, USA.
194. WellcomeTrust Centre for Human Genetics, Oxford OX3 7BN, UK.
195. deCODE Genetics, 101 Reykjavik, Iceland.
196. Department of Clinical Neurology, Medical University of Vienna, 1090 Wien, Austria.
197. Lieber Institute for Brain Development, Baltimore, Maryland 21205, USA.
198. Department of Medical Genetics, University Medical Centre Utrecht, Universiteitsweg 100,3584CG, Utrecht, The Netherlands.
199. Berkshire Healthcare NHS Foundation Trust, Bracknell RG12 1BQ, UK.
200. Section of Psychiatry, University of Verona, 37134 Verona, Italy.

201. Department of Psychiatry, University of Oulu, P.O. Box 5000, 90014, Finland.
202. University Hospital of Oulu, P.O. Box 20, 90029 OYS, Finland.
203. Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland.
204. Health Research Board, Dublin 2, Ireland.
205. School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth WA6009, Australia.
206. Computational Sciences CoE, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA.
207. Human Genetics, Genome Institute of Singapore, A*STAR, Singapore 138672.
208. University College London, London WC1E 6BT, UK.
209. Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, 52428 Juelich, Germany.
210. Department of Genetics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel.
211. NeuroscienceDiscovery and Translational Area, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland.
212. Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Medical Research Foundation Building, Perth WA6000, Australia.
213. Virginia Institute for Psychiatric and Behavioral Genetics, Departments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA.
214. The Feinstein Institute for Medical Research, Manhasset, New York 11030, USA.
215. The Hofstra NS-LIJ School of Medicine, Hempstead, New York 11549, USA.
216. The Zucker Hillside Hospital, Glen Oaks, New York 11004, USA.
217. Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore.
218. Queensland Centre for Mental Health Research, University of Queensland, Brisbane 4076, Queensland, Australia.
219. Center for HumanGenetic Research and Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
220. Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam3000, The Netherlands.
221. Department of Complex Trait Genetics, NeuroscienceNeuroscience Campus Amsterdam, VU University Medical Center Amsterdam, Amsterdam1081, The Netherlands.
222. Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam1081, The Netherlands.
223. University of Aberdeen, Institute of Medical Sciences, Aberdeen AB25 2ZD, UK.

224. Departments of Psychiatry, Neurology, Neuroscience and Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA.
225. Department of Clinical Medicine, University of Copenhagen, Copenhagen 2200, Denmark.

References for main text

1. Nicolae, D. L. *et al.* Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genet.* **6,** e1000888 (2010).
2. Dobbyn, A. *et al.* Co-localization of Conditional eQTL and GWAS Signatures in Schizophrenia. *Rev.* (2017).
3. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24,** 408–415 (2008).
4. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10,** 184–194 (2009).
5. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16,** 197–212 (2015).
6. Moffatt, M. F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448,** 470–473 (2007).
7. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42,** 937–948 (2010).
8. Dubois, P. C. A. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42,** 295–302 (2010).
9. Libioulle, C. *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* **3,** e58 (2007).
10. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10,** e1004383 (2014).
11. Boocock, J., Giambartolomei, C. & Stahl, E. A. COLOC2. (2016).
12. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48,** 481–487 (2016).
13. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48,** 709–717 (2016).
14. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19,** 1442–1453 (2016).
15. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47,** 1091–8 (2015).

16. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48,** 245–52 (2016).
17. Geschwind, D. H. & Flint, J. Genetics and genomics of psychiatric disease. *Science (80-.).* **349,** (2015).
18. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28,** 882–3 (2012).
19. A. Bennett, D., A. Schneider, J., Arvanitakis, Z. & S. Wilson, R. Overview and Findings from the Religious Orders Study. *Curr. Alzheimer Res.* **9,** 628–645 (2012).
20. Bennett, D. A., Schneider, J. A., Buchman, A. S., Barnes, L. L. & Wilson, P. A. B. and R. S. Overview and Findings from the Rush Memory and Aging Project. *Current Alzheimer Research* **9,** 646–663 (2012).
21. Mele, M. *et al.* The human transcriptome across tissues and individuals. *Science (80-.).* **348,** 660–665 (2015).
22. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
23. Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am. J. Hum. Genet.* (2018). doi:10.1016/j.ajhg.2018.04.011
24. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10,** e1004383 (2014).
25. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9,** (2018).
26. Benjamin, Y. & Hochberg, Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 289–300 (1995).
27. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146,** 247–61 (2011).
28. Fromer, M. *et al.* *Gene Expression Elucidates Functional Impact of Polygenic Risk for Schizophrenia*. *bioRxiv* (Cold Spring Harbor Labs Journals, 2016). doi:10.1101/052209
29. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant

genes and in regions under strong background selection. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0059-2

30. Sanders, S. J. First glimpses of the neurobiology of autism spectrum disorder. *Curr. Opin. Genet. Dev.* **33,** 80–92 (2015).
31. Exome Aggregation Consortium *et al. Analysis of protein-coding genetic variation in 60,706 humans*. *bioRxiv* (Cold Spring Harbor Labs Journals, 2015). doi:10.1101/030338
32. Malhotra, D. *et al.* High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron* **72,** 951–963 (2011).
33. Bautista, O., Vázquez-Caubet, J. C., Zhivago, E. A. & Dolores Sáiz, M. From metabolism to psychiatric symptoms: psychosis as a manifestation of acute intermittent porphyria. *J. Neuropsychiatry Clin. Neurosci.* **26,** E30 (2014).
34. Zimmermann, M., Bonaccurso, C., Valerius, C. & Hamann, G. F. [Acute intermittent porphyria. A clinical chameleon: case study of a 40-year-old female patient]. *Nervenarzt* **77,** 1501–5 (2006).
35. Ventura, P. *et al.* A challenging diagnosis for potential fatal diseases: recommendations for diagnosing acute porphyrias. *Eur. J. Intern. Med.* **25,** 497–505 (2014).
36. Pischik, E. & Kauppinen, R. An update of clinical management of acute intermittent porphyria. *Appl. Clin. Genet.* **8,** 201–14 (2015).
37. Kumar, B. Acute intermittent porphyria presenting solely with psychosis: a case report and discussion. *Psychosomatics* **53,** 494–8 (2012).
38. Bonnot, O. *et al.* Diagnostic and treatment implications of psychosis secondary to treatable metabolic disorders in adults: a systematic review. *Orphanet J. Rare Dis.* **9,** 65 (2014).
39. Kaback, M. M. & Desnick, R. J. *Hexosaminidase A Deficiency*. *GeneReviews(®)* (University of Washington, Seattle, 1993).
40. Saleh, O. Late Onset Tay-Sachs Disease Pre senting as a Brief Psychotic Disorder with Catatonia: A Case Report and Review of Literature.
41. Skaper, S. D. in *Brain Protection in Schizophrenia, Mood and Cognitive Disorders* (ed. Ritsner, M. S.) 135–165 (Springer Science & Business Media, 2010, 2010).
42. Castellano, E. *et al.* RAS signalling through PI3-Kinase controls cell migration via modulation of Reelin expression. *Nat. Commun.* **7,** 11245 (2016).

43. Gururajan, A. & Buuse, M. van den. Is the mTOR-signalling cascade disrupted in Schizophrenia? *J. Neurochem.* **129,** 377–387 (2014).

44. Ritsner, M. S. *Brain Protection in Schizophrenia, Mood and Cognitive Disorders*. (Springer Science & Business Media, 2010, 2010). doi:10.1007/978-90-481-8553-5

45. Enriquez-Barreto, L. & Morales, M. The PI3K signaling pathway as a pharmacological target in Autism related disorders and Schizophrenia. *Mol. Cell. Ther.* **4,** 2 (2016).

46. Glessner, J. T. *et al.* Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10584–9 (2010).

47. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18,** 199–209 (2015).

48. Bauman, A. L. *et al.* Cocaine and antidepressant-sensitive biogenic amine transporters exist in regulated complexes with protein phosphatase 2A. *J. Neurosci.* **20,** 7571–8 (2000).

49. Ayadi, A. *et al.* Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm. Genome* **23,** 600–610 (2012).

50. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477,** 289–294 (2011).

51. Howe, D. G. *et al.* ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic Acids Res.* **41,** D854–D860 (2013).

52. Smith, C. L., Blake, J. A., Kadin, J. A., Richardson, J. E. & Bult, C. J. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res.* **46,** D836–D842 (2018).

53. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508,** 199–206 (2014).

54. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393,** 440–442 (1998).

55. Nguyen, H. T. *et al.* Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *doi.org* 135293 (2017). doi:10.1101/135293

56. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100,** 473–487 (2017).
57. Gottlieb, A., Daneshjou, R., DeGorter, M., Montgomery, S. & Altman, R. Population-specific imputation of gene expression improves prediction of pharmacogenomic traits for African Americans. *bioRxiv* 115451 (2017). doi:10.1101/115451
58. Need, A. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends Genet* **25,** 489–94 (2009).
59. Popejoy, A. & Fullerton, S. Genomics is failing on diversity. *Nature* **538,** 161–164 (2016).
60. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. *bioRxiv* 68593 (2016). doi:10.1101/068593
61. Browning, R. in *The Poems of Robert Browning* (eds. Porter, C. & Clarke, H. A.) 257–271 (Thomas Y. Cromwell and Company, 1896).
62. Loftus, L. S. & Arnold, W. N. Vincent van Gogh's illness: acute intermittent porphyria? *BMJ* **303,** 1589–91
63. Strik, W. K. [The psychiatric illness of Vincent van Gogh]. *Nervenarzt* **68,** 401–9 (1997).
64. Arnold, W. N. The illness of Vincent van Gogh. *J. Hist. Neurosci.* **13,** 22–43 (2004).
65. Hughes, J. R. A reappraisal of the possible seizures of Vincent van Gogh. *Epilepsy Behav.* **6,** 504–10 (2005).
66. Bhattacharyya, K. B. & Rai, S. The neuropsychiatric ailment of Vincent Van Gogh. *Ann. Indian Acad. Neurol.* **18,** 6–9 (2014).
67. Correa, R. Vincent van Gogh: A pathographic analysis. *Med. Hypotheses* **82,** 141–144 (2014).
68. Peters, T. J. & Beveridge, A. The madness of King George III: a psychiatric re-assessment. *Hist. Psychiatry* **21,** 20–37 (2010).
69. Szatkiewicz, J. P. *et al.* Copy number variation in schizophrenia in Sweden. *Mol. Psychiatry* **19,** 762–773 (2014).
70. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506,** 179–184 (2014).
71. Consortium, C.-D. G. of the P. G. Identification of risk loci with shared effects on five

major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).

72. Keefe, R. S. E. & Fenton, W. S. How Should DSM-V Criteria for Schizophrenia Include Cognitive Impairment? *Schizophr. Bull.* **33,** 912–920 (2007).

73. Reichenberg, A. *et al.* Static and Dynamic Cognitive Deficits in Childhood Preceding Adult Schizophrenia: A 30-Year Study. *Am. J. Psychiatry* **167,** 160–169 (2010).

74. Gold, J. M. Cognitive deficits as treatment targets in schizophrenia. *Schizophr. Res.* **72,** 21–28 (2004).

75. Cannon, M. *et al.* Evidence for Early-Childhood, Pan-Developmental Impairment Specific to Schizophreniform Disorder. *Arch. Gen. Psychiatry* **59,** 449 (2002).

76. Parikshak, N. N., Gandal, M. J., Geschwind, D. H. & Angeles, L. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16,** 441–458 (2015).

77. Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* **14,** R75 (2013).

78. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478,** 519–523 (2012).

79. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *bioRxiv* 67355 (2016). doi:10.1101/067355

Figure Legends for main text

Figure 1: Replication of DLPFC prediction models in independent data.
Measured gene expression (ROSMAP RNA-seq) was compared to predicted genetically-regulated gene expression for CMC DLPFC and 12 GTeX predictor databases. Replication R^2 values are significantly higher for the DLPFC than for the 12 GTEX brain expression models.

A. Distribution of R_R^2 values of CMC DLPFC predictors in ROSMAP data. Mean R_R^2 = 0.056. 47.7% of genes have R_R^2 >= 0.01. Boxplots show mean, quartiles,; whiskers show full range of data.
B. Distribution of R_R^2 values of 12 GTeX predictors in ROSMAP data.

Table of sample sizes and p-val thresholds for CMC DLPFC and GTeX data. Number of samples, number of genes in the prediXcan model and number of eGenes are all significantly correlated with predictor performance in ROSMAP data (spearman correlation test) .

Figure 2: SCZ associations results

A) 413 genes are associated with SCZ across 12 brain tissues. Each point represents one gene-tissue pair.
B) 67 genes remain significant outside the MHC after stepwise conditional analysis
C) Number of genome-wide significant loci, outside the MHC region, identified in each brain region. These trends are partly driven by differences in power between brain regions.

Abbreviations are as follows; CB- Cerebellum; CX- Cortex; FL- Frontal Cortex; DLPFC- Dorso-lateral pre-frontal cortex; CB HEMI- Cerebellar Hemisphere; HIP- Hippocampus; PIT- Pituitary Gland; HTH- Hypothalamus; NAB- Nucleus Accumbens (Basal Ganglia); PUT- Putamen (Basal Ganglia); CAU- Caudate (Basal Ganglia); CNG- Anterior Cingulate Cortex

Figure 3: SCZ-associated genes are co-expressed throughout development and across brain regions

A) Brain tissues selected for each of four brainspan regions. Brainspan includes 525 samples from 43 unique individuals. Region 1: IPC, V1C, ITC, OFC, STC, A1C; Region 2:S1C, M1C, DFC, VFC, MFC; Region 3:HIP, AMY, STR; Region 4: CB

Average clustering coefficients were calculated for all pairs of SCZ-associated genes, and compared to average clustering coefficients for 100,000 permuted gene networks to obtain empirical significance levels.

Figure 4: Gene expression patterns for SCZ-associated genes cluster into four groups, relating to distinct spatiotemporal expression.
Brain regions are shown in figure 3a.

A. 29 genes are expressed in the early-mid pre-natal period (4-24 post-conception weeks)
B. 15 genes are expressed throughout development; subclusters correspond to either specific expression in region 4, or expression across the brain
C. Ten genes are expressed in the late-prenatal (25-38pcw) and post-natal period
D. 12 genes are expressed in the late pre-natal period (25-39pcw)

Table 1: SCZ-associated genes following conditional analysis. *67 non-MHC genes are significantly associated with schizophrenia following conditional analysis. Effect sizes (BETA) refer to predicted genetically regulated gene expression (GREX) in cases compared to controls. Effect sizes and odds ratios are also shown adjusted to 'unit' variance in gene expression. OR, odd's ratio; DLPFC, dorso-lateral prefrontal cortex*

Gene name	Tissue	BETA	P	GVAR	Adjusted BETA	Adjusted OR
GNL3	Cerebellum	0.037	$1.39x10^{-11}$	0.115	0.012	1.012
THOC7	Cerebellum	-0.113	$5.77x10^{-10}$	0.010	-0.011	0.989
NAGA	Cerebellum	0.122	$1.12x10^{-09}$	0.009	0.011	1.011
TAC3	Cerebellum	-0.868	$8.03x10^{-08}$	0.000	-0.015	0.985
CHRNA2	Cerebellum	-0.016	$1.63x10^{-07}$	0.395	-0.010	0.990
ACTR5	Cerebellum	0.208	$3.88x10^{-07}$	0.019	0.029	1.029
INO80E	Frontal Cortex	0.130	$7.25x10^{-12}$	0.009	0.012	1.013
PLPPR5	Frontal Cortex	-0.672	$2.58x10^{-09}$	0.006	-0.053	0.948
FAM205A	Frontal Cortex	0.043	$1.21x10^{-08}$	0.061	0.011	1.011
AC110781.3	Thyroid	0.342	$1.31x10^{-13}$	0.002	0.014	1.014
IMMP2L	Thyroid	-0.073	$7.09x10^{-12}$	0.046	-0.016	0.984
IGSF9B	Thyroid	-0.024	$3.05x10^{-07}$	0.156	-0.010	0.991
NMRAL1	Thyroid	0.038	$4.03x10^{-07}$	0.060	0.009	1.009
HIF1A	DLPFC	11.130	$7.52x10^{-14}$	0.000	0.148	1.159
TIMM29	DLPFC	11.207	$9.27x10^{-14}$	0.000	0.168	1.183
ST7-OT4	DLPFC	10.170	$5.79x10^{-13}$	0.001	0.318	1.374
H2AFY2	DLPFC	10.962	$3.60x10^{-12}$	0.000	0.191	1.211
STARD3	DLPFC	10.740	$5.90x10^{-12}$	0.001	0.304	1.355
CTC-471F3.5	DLPFC	8.535	$1.11x10^{-11}$	0.000	0.104	1.110
SF3A1	DLPFC	8.651	$1.32x10^{-11}$	0.000	0.083	1.086
ZNF512	DLPFC	10.312	$1.32x10^{-11}$	0.001	0.261	1.298
FURIN	DLPFC	-0.084	$2.22x10^{-11}$	0.022	-0.012	0.988
INHBA-AS1	DLPFC	8.399	$2.24x10^{-11}$	0.000	0.127	1.135
SF3B1	DLPFC	0.099	$6.14x10^{-11}$	0.014	0.012	1.012
EFTUD1P1	DLPFC	-0.092	$1.81x10^{-10}$	0.017	-0.012	0.988
MLH1	DLPFC	2.840	$2.10x10^{-10}$	0.001	0.069	1.071
GATAD2A	DLPFC	-0.044	$2.18x10^{-10}$	0.071	-0.012	0.988
METTL1	DLPFC	9.357	$2.23x10^{-10}$	0.000	0.166	1.181
DMC1	DLPFC	7.229	$4.48x10^{-10}$	0.000	0.130	1.139
RAD51D	DLPFC	7.612	$2.11x10^{-09}$	0.000	0.111	1.117

RERE	DLPFC	2.847	6.32x10^{-09}	0.000	0.036	1.037
PCCB	DLPFC	-0.044	2.05x10^{-08}	0.054	-0.010	0.990
CLCN3	DLPFC	0.141	2.96x10^{-08}	0.005	0.010	1.010
ATG101	DLPFC	8.086	4.90x10^{-08}	0.007	0.695	2.005
JRK	DLPFC	0.032	1.25x10^{-07}	0.091	0.010	1.010
PTPRU	DLPFC	-0.077	1.60x10^{-07}	0.016	-0.010	0.990
MARCKS	DLPFC	0.398	2.05x10^{-07}	0.001	0.015	1.015
TCF4	Anterior Cingulate Cortex	-0.059	5.22x10^{-13}	0.051	-0.013	0.987
DGKD	Anterior Cingulate Cortex	-0.937	2.63x10^{-11}	0.001	-0.022	0.979
C1QTNF4	Anterior Cingulate Cortex	-0.173	1.37x10^{-09}	0.010	-0.017	0.983
PITPNA	Anterior Cingulate Cortex	-0.243	1.77x10^{-07}	0.002	-0.010	0.990
FXR1	Caudate Basal Ganglia	0.439	5.40x10^{-12}	0.001	0.017	1.017
ZDHHC1	Caudate Basal Ganglia	0.354	5.36x10^{-08}	0.001	0.011	1.012
PDE4D	Cerebellar Hemisphere	0.365	6.81x10^{-11}	0.001	0.013	1.013
DRD2	Cerebellar Hemisphere	-0.182	2.47x10^{-10}	0.004	-0.012	0.988
PITPNM2	Cerebellar Hemisphere	-0.065	2.21x10^{-09}	0.028	-0.011	0.989
RINT1	Cerebellar Hemisphere	0.086	6.32x10^{-09}	0.016	0.011	1.011
SRMS	Cerebellar Hemisphere	-0.440	3.08x10^{-08}	0.001	-0.011	0.989
SETD6	Cerebellar Hemisphere	-0.043	1.05x10^{-07}	0.054	-0.010	0.990
APOPT1	Cortex	-0.074	1.24x10^{-10}	0.026	-0.012	0.988
VSIG2	Cortex	-0.092	6.01x10^{-09}	0.013	-0.011	0.989
SDCCAG8	Cortex	-0.069	3.88x10^{-07}	0.002	-0.003	0.997
PIK3C2A	Cortex	-0.040	4.04x10^{-07}	0.365	-0.024	0.976
AS3MT	Frontal Cortex	0.594	5.65x10^{-17}	0.001	0.017	1.017
FOXN2	Hippocampus	-0.250	2.65x10^{-07}	0.021	-0.036	0.964
RASIP1	Nucleus Accumbens Basal Ganglia	0.055	3.80x10^{-08}	0.034	0.010	1.010
TCF23	Nucleus Accumbens Basal Ganglia	-0.076	4.83x10^{-08}	0.019	-0.010	0.990
TTC14	Nucleus Accumbens Basal Ganglia	-0.089	4.84x10^{-08}	0.013	-0.010	0.990
TYW5	Putamen Basal Ganglia	-0.080	2.63x10^{-13}	0.035	-0.015	0.985
SNX19	Putamen Basal Ganglia	0.031	1.31x10^{-12}	0.179	0.013	1.013
CIART	Putamen Basal Ganglia	0.090	6.78x10^{-10}	0.017	0.012	1.012
SH2D7	Putamen Basal Ganglia	0.096	7.89x10^{-09}	0.013	0.011	1.011
DGUOK	Putamen Basal Ganglia	0.255	8.26x10^{-08}	0.002	0.011	1.011
C12orf76	Putamen Basal Ganglia	0.031	2.27x10^{-07}	0.095	0.010	1.010
LRRC37A	Putamen Basal Ganglia	-0.035	2.69x10^{-07}	0.076	-0.010	0.991
AC005841.1	Pituitary	0.162	3.28x10^{-09}	0.005	0.011	1.011
RPS17	Pituitary	0.035	4.03x10^{-08}	0.082	0.010	1.010

Associations in the MHC region

BTN1A1	Caudate Basal Ganglia	-0.261	1.67×10^{-22}
VARS2	Anterior Cingulate Cortex	0.075	7.48×10^{-15}
HIST1H3H	Putamen Basal Ganglia	-1.106	3.22×10^{-10}
NUDT3	Nucleus Accumbens Basal Ganglia	0.104	6.55×10^{-9}

Table 2: Significantly enriched pathways and gene sets. We tested for enrichment of 8,657 pathways among our prediXcan results using a competitive p-value in MAGMA and calculated an FDR-corrected p-value to determine significance. FMRP, fragile-X mental retardation protein; BP, bipolar; CNV, copy number variant; LOF, loss of function.

Analysis	Gene Set	Comp P	FDR P
Hypothesis driven	FMRP-targets	$1.96x10^{-08}$	$3.097x10^{-06}$
	BP denovo CNV	$7.92x10^{-08}$	$6.257x10^{-06}$
	HIGH LOF intolerant	$5.86x10^{-05}$	0.00309
Agnostic	Increased spleen iron level	$2.72x10^{-08}$	0.000245
	Decreased IgM level	$6.80x10^{-07}$	0.00307
	Condensed chromosome	$1.99x10^{-06}$	0.00598
	Chromosome	$2.80x10^{-06}$	0.00632
	Abnormal spleen iron level	$6.79x10^{-06}$	0.00765
	Mitotic Anaphase	$6.39x10^{-06}$	0.00765
	Mitotic Metaphase and Anaphase	$5.13x10^{-06}$	0.00765
	Resolution of Sister Chromatid Cohesion	$5.82x10^{-06}$	0.00765
	Increased liver iron level	$1.03x10^{-05}$	0.0103
	Separation of Sister Chromatids	$1.28x10^{-05}$	0.0115
	Regulation of Rab GTPase activity	$1.78x10^{-05}$	0.0123
	Regulation of Rab protein signal transduction	$1.78x10^{-05}$	0.0123
	Protein phosphorylated amino acid binding	$1.75x10^{-05}$	0.0123
	Chromosome	$2.57x10^{-05}$	0.0165
	Hexosaminidase activity	$3.47x10^{-05}$	0.0174
	Abnormal learning memory conditioning	$3.11x10^{-05}$	0.0174
	Abnormal liver iron level	$3.47x10^{-05}$	0.0174
	Mitotic Prometaphase	$2.99x10^{-05}$	0.0174
	M Phase	$3.70x10^{-05}$	0.0176
	Positive regulation of Rab GTPase activity	$5.93x10^{-05}$	0.0232
	Rab GTPase activator activity	$5.93x10^{-05}$	0.0232
	Protein phosphatase type 2A regulator activity	$5.24x10^{-05}$	0.0232
	Replicative senescence	$5.44x10^{-05}$	0.0232
	Condensed nuclear chromosome	$7.11x10^{-05}$	0.0267
	Ubiquitin-specific protease activity	0.000104	0.0335
	Ras GTPase activator activity	$9.61x10^{-05}$	0.0335
	Metabolism of porphyrins	0.000103	0.0335
	Kinetochore	0.000103	0.0335
	Decreased physiological sensitivity to xenobiotic	0.000127	0.0381
	Antigen Activates B Cell Receptor Leading to Generation of Second Messengers	0.000124	0.0381

Phosphoprotein binding	0.000146	0.0424
Abnormal dorsal-ventral axis patterning	0.000152	0.0429

Online Methods (Limit 3,000 words, at end of manuscript, currently 2,064)

Creating gene expression predictors for the dorso-lateral pre-frontal cortex

eQTL Data

Genotype and RNAseq data were obtained for 538 European individuals through the CommonMind Project[14]. The mean age of these individuals was 67.4 years. RNA-seq data were generated from post-mortem human dorsolateral prefrontal cortex (DLPFC). The gene expression matrix was normalized to log(counts per million) using voom. Adjustments were made for known covariates (including sample ascertainment, quality, experimental parameters, ancestry) and surrogate variables, using linear modelling with voom-derived regression weights. Details on genotyping, imputation and RNA-seq generation may be found in the CommonMind Consortium (CMC) flagship paper[14].

The samples used here include 254 schizophrenia and 52 bipolar cases, as well as controls. The CMC flagship paper[14] applied a permutation test and an explicit disease-genotype interaction term to demonstrate that there is no significant effect of disease on eQTLs. We have therefore included both cases and controls in this analysis, to maximize sample size.

A 1% minor allele frequency (MAF) cut-off was applied. Variants were filtered to remove any SNPs in high LD (r^2>0.9), indels, and all variants with ambiguous ref/alt alleles. All protein coding genes on chromosomes 1-22 with at least one cis-SNP after these QC steps were included in this analysis (15,362 genes in total). SNPs in trans have been shown not to provide a substantial improvement in prediction accuracy[15] and were not included here.

Building gene expression prediction databases

Gene expression prediction models were created following the “PrediXcan” method[15]. Matched genotype and gene expression data were used to identify a set of variants that influence gene expression (Supplementary Figure 2a). Weights for these variants are calculated using regression in a ten-fold cross-validation framework. All cross-validation folds were balanced for diagnoses, ethnicity, and other clinical variables.

All SNPs within the cis-region (+/- 1 Mb) of each gene were included in the regression analysis. Accuracy of prediction was estimated by comparing predicted expression to measured expression, across all 10 cross-validation folds; this correlation was termed cross-validation R^2 or R_{cv}^2. Genes with $R_{cv}^2 > 0.01$ (~$p < 0.05$) were included in our final predictor database.

Prediction models were compared across four different regression methods; elastic net (prediXcan), ridge regression (using the TWAS method[16]), Bayesian sparse linear mixed modelling (BSLMM; TWAS), and linear regression using the best eQTL for each gene (Supplementary Figure 1a). Mean R_{cv}^2 values were significantly higher for elastic net regression (mean $R_{cv}^2 = 0.056$) than for eQTL-based prediction (mean $R_{cv}^2 = 0.025$), BSLMM (mean $R_{cv}^2 = 0.021$) or Ridge Regression (mean $R_{cv}^2 = 0.020$). The distribution of R_{cv}^2 values was also significantly higher for elastic net regression than for any other method (Kolgorov-Smirnov test, $p < 2.2 \times 10^{-16}$).

Replication of gene expression prediction models in independent data

Predictive accuracy of CMC DLPFC models were tested in two independent datasets.

First, we used data from the Religious Orders Study and Memory and Aging Project (ROSMAP[19,20]). This study included genotype data and DLPFC RNA-seq data for 451 individuals of European descent (Supplementary Figure 2b).

DLPFC GREX was calculated using the CMC DLPFC predictor models. Correlation between RNA-seq expression and CMC DLPFC GREX ("Replication R^2 values" or R_R^2) was used as a measure of predictive accuracy. R_R^2 was calculated including correction for ten ancestry components, as follows:

Equation 1: R_R^2 calculation.

$$\mathrm{R}_{\mathrm{R1}}^2 = (\mathrm{M} \sim \mathrm{GREX} + \mathrm{PC}_1 + \mathrm{PC}_2 + \cdots + \mathrm{PC}_{10})$$

$$\mathrm{R}_{\mathrm{R2}}^2 = (\mathrm{M} \sim \mathrm{PC}_1 + \mathrm{PC}_2 + \cdots + \mathrm{PC}_{10})$$

$$\mathrm{R}_{\mathrm{R}}^2 = \mathrm{R}_{\mathrm{R1}}^2 \text{-} \mathrm{R}_{\mathrm{R2}}^2$$

Where:

M	Measured expression (RNA-seq)

GREX	GREX imputed expression
PC_n	n^{th} Principal Component

A small number of genes (158) had very low predictive accuracy and were removed from further analyses. Cross-validation R^2 (R_{cv}^2) values and R_R^2 values were highly correlated (rho = 0.62, $p < 2.2 \times 10^{-16}$; Supplementary Figure 3a). 55.7% of CMC DLPFC genes had R_R^2 values > 0.01.

Prediction accuracy was also assessed for 11 publicly available GTEx neurological predictor databases, and R_R^2 values used to compare to CMC DLPFC performance. CMC DLPFC models had higher average R_R^2 values, more genes with $R_R^2 > 0.01$, and significantly higher overall distributions of R_R^2 values than any of the twelve GTEx brain tissue models (ks-test, $p < 2.2 \times 10^{-16}$; Figure 1a,b).

To estimate trans-ancestral prediction accuracy, GREX was calculated for 162 African-American individuals and 280 European individuals from the NIMH Human Brain Collection Core (HBCC) dataset (Supplementary Figure 2c). Predicted gene expression levels were compared to DLPFC expression levels measured using microarray. There was a significant correlation between the European and African-American samples for R_{CV}^2 values and R_R^2 values (rho = 0.66, 0.56; Supplementary Figure 3b,c). R_R^2 values were higher on average in Europeans, but were significantly correlated between African-Americans and Europeans (rho = 0.78, $p < 2.2 \times 10^{-16}$, Pearson test; Supplementary Figure 3d).

Extension to Summary Statistics

Transcriptomic Imputation may be applied to summary statistics instead of raw data, in instances where raw data is unavailable. However, this method suffers from slightly reduced accuracy, requires covariance matrices calculated in an ancestrally-matched reference population[25] (usually only possible for European cohorts), and precludes testing of endophenotypes within the data, and so should not be applied when raw data are available.

We assessed concordance between CMC DLPFC transcriptomic imputation results using summary-statistics (S-PrediXcan[25]) and raw genotypes (PrediXcan[15]) using nine European and three Asian PGC-SCZ cohorts[22] for which both data types were available. Cohorts were chosen to encompass a range of case : control ratios, to test previous suggestions that accuracy is reduced in unbalanced cohorts[80]. Covariances for all variants included in the DLPFC predictor models were computed using S-PrediXcan[25]. For all European cohorts, Pearson correlation of log-10 p-values and effect sizes was above 0.95. The mean correlation was 0.963 (Supplementary Figure 11). There was no correlation between total sample size, case-control ratio, p-value or effect-size. Seven genes were removed due to discordant p-values. For the three Asian cohorts tested, the mean correlation was 0.91 (Supplementary Figure 12).

Concordance was also tested for the same nine European PGC-SCZ cohorts, across 12 neurological GTEx prediction databases. All correlations were significant ($\rho > 0.95$, $p < 2.2 \times 10^{-16}$). There was a significant correlation between p-value concordance and case-control ratio ($\rho = 0.37$, $p = 7.606 \times 10^{-15}$). 114 genes had discordant p-values between the two methods and were excluded from future analyses.

Application to Schizophrenia

Dataset Collection

We obtained 53 discovery cohorts for this study, including 40,299 SCZ cases and 65,264 controls (Supplementary Figure 4). 52/53 cohorts (35,079 cases, 46,441 controls) were obtained through collaboration with the Psychiatric Genomics Consortium, and are described in the 2014 PGC Schizophrenia GWAS[22]. The remaining cohort, referred to as CLOZUK2, constitutes the largest single cohort of individuals with Schizophrenia (5,220 cases and 18,823 controls), collected as part of an effort to investigate treatment-resistant Schizophrenia[60].

50/53 datasets included individuals of European ancestry, while three datasets include individuals of Asian ancestry (1,836 cases, 3,383 controls). All individuals were ancestrally matched to controls. Information on genotyping, quality control and other data management issues may be found in the original papers describing these collections[22,60]. All sample collections complied with ethical regulations. Details regarding ethical compliance and consent procedures may be found in the original manuscripts describing these collections[22,60].

Access to dosage data was available for 44/52 PGC-SCZ cohorts. The remaining PGC cohorts, and the CLOZUK2 cohort provided summary statistics. Three European PGC cohorts were trio-based, rather than case-control.

Additionally, we tested for replication of our CMC DLPFC associations in an independent dataset of 4,133 cases and 24,788 controls obtained through collaboration with the iPSYCH-GEMS schizophrenia working group (effective sample size 14,169.5; Supplementary Figure 4b, Supplementary Note).

Transcriptomic Imputation and association testing

Transcriptomic Imputation was carried out individually for each case-control PGC-SCZ cohort with available dosage data (44/52 cohorts). Predicted gene expression levels were computed using the DLPFC predictors described in this manuscript, as well as for 11 other brain tissues prediction databases created using GTEx tissues[15,21,81,82] (Figure 1c). Associations between predicted gene expression values and case-control status were calculated using a linear regression test in R. Ten ancestry principal components were included as covariates. Association tests were carried out independently for each cohort, across 12 brain tissues.

For the eight PGC cohorts with no available dosage data, the three PGC trio-based analyses, and the CLOZUK2 cohort, a summary-statistic based transcriptomic imputation approach was used ("S-PrediXcan[25]"), as described previously.

Meta-analysis

Meta-analysis was carried out across all 53 cohorts using METAL[83]. Cochran's Q test for heterogeneity was implemented in METAL[83,84], and a heterogeneity p-value threshold of $p > 1 \times 10^{-3}$ applied to results. A conservative significance threshold was applied to these data, correcting for the total number of genes tested across all tissues (121,611 gene-region tests in total). This resulted in a genome-wide significance threshold of 4.1×10^{-7}.

Effect sizes and direction of effect quoted in this manuscript refer to changes in predicted expression in cases compared to controls i.e., genes with negative effect sizes have decreased predicted expression in cases compared to controls.

Identifying independent associations

We identified a number of genomic regions which contained multiple gene associations and/or genes associated across multiple tissues. We identified 58 of these regions, excluding the MHC, based on distance between associated genes, and verified them using visual inspection. In order to identify independent genic associations within these regions, we carried out a stepwise forward conditional analysis following "GCTA-COJO" theory[85] using "CoCo" (see URLs), an R implementation of GCTA-COJO. CoCo allows the specification of custom correlation matrices by the user (for example, ancestrally specific LD matrices). For each region, we generated a predicted gene expression correlation matrix for all significant genes ($p \leq 1x10^{-6}$), as the root-effective sample size[83] (N_{eff}, eqn 2) weighted average correlation across all cohorts where we had access to dosage data.

Equation 2: Effective Sample Size, N_{eff}

$$N_{eff} = \frac{4}{(\frac{1}{N_{cases}} + \frac{1}{N_{controls}})}$$

Forward stepwise conditional analysis of all significant genes was carried out using joint linear regression modeling. First, the top-ranked gene was added to the model, then the next most significant gene in a joint model is added if significant at a given p-value threshold, and so on until either all genes are added to the model, or no joint statistic reaches the significance threshold.

We calculated effect sizes and odds ratios for SCZ-associated genes by adjusting "CoCo" betas to have unit variance (Table 1, eqn. 3).

Equation 3: GREX Beta adjustment

$$\beta = \beta_{CoCo} \text{ x } \sqrt{GVAR}$$

Where GVAR is the variance of the GREX predictor for each gene.

Gene set Analyses

Pathway analyses were carried out using an extension to MAGMA[86]. P-values were assigned to genes using the most significant p-value achieved by each gene in the meta-analysis. We then carried out a competitive gene-set analysis test using these p-values, using two gene sets:

1. 159 gene sets with prior hypotheses for involvement in SCZ development, including loss-of-function intolerant genes, CNV-intolerant genes, targets of the fragile-X mental retardation protein, CNS related gene sets, and 104 behavioral and neurological pathways from the Mouse Genome Informatics database[14,60,69,87].
2. An agnostic analysis, including ~8,500 gene sets collated from publicly available databases including GO[88,89], KEGG[90], REACTOME[91], PANTHER[92,93], BIOCARTA[94] and MGI[52]. Sets were filtered to include only gene sets with at least ten genes.

Significance levels were adjusted across all pathways included in either test using the Benjamini-Hochberg “FDR” correction in R[26].

Coexpression of SCZ genes throughout development

We investigate spatiotemporal expression of our associated genes using publicly available developmental transcriptome data, obtained from the BRAINSPAN consortium[53]. We partitioned these data into biologically relevant spatio-temporal data sets[95], corresponding to four general brain regions; the frontal cortex, temporal and parietal regions, sensory-motor regions, and subcortical regions (Figure 3a[96]), and eight developmental time-points (four pre-natal, four post-natal)[95].

First, we tested for correlation of gene expression for all SCZ-associated genes at each spatiotemporal time-point. Genes with pearson correlation coefficients >= 0.8 or <=-0.8 were considered co-expressed. 100,000 iterations of this analysis were carried out using random gene sets with equivalent expression level distributions to the SCZ-associated genes. For each gene set, a gene co-expression network was created, with edges connecting all co-expressed genes. Networks were assessed using three criteria; first, the number of edges within the network, as a crude measured of connectedness; second, the Watts-Strogatz average path length between nodes, as a global measure of connectedness across all genes in the network[54]; third, the Watts-Strogatz clustering coefficient, to measure tightness of the clusters within the network[54]. For each spatio-temporal time point, we plotted gene-pair expression correlation (Supplementary Figure 7) and co-expression networks (Supplementary Figure 8).

For each of the 67 SCZ-associated genes, we calculated average expression at each spatiotemporal point. We then calculated Z-Score of expression specificity using these values, and plotted Z-Scores to visually examine patterns of gene expression throughout development and across brain regions. Clusters were formally identified using a dendrogram cut at height 10 (Supplementary Figure 9).

In-silico replication of SCZ-associated genes in mouse models

We downloaded genotype, knock-out allele information and phenotyping data for ~10,000 mouse mutant models from five large mouse phenotyping and genotyping projects; Mouse Genome Informatics (MGI[52]), EuroPhenome[49,97], Mouse Genome Project (MGP[49,50]), International Mouse Phenotyping Consortium (IMPC[98]), and Infection and Immunity Immunophenotyping (3I[98]). Where possible, we also downloaded raw phenotyping data regarding specific assays. In total, we obtained 175,012 phenotypic measurements, across 10,288 mutant mouse models. We searched for any mouse lines with phenotypes related to behavior (natural, observed, stereotypic or assay-induced); cognition or working memory; brain, head or craniofacial dysmorphology; retinal or eye morphology, and/or vision or visual dysfunction or impairment; ear morphology or hearing dysfunction or impairment; neural tube defects; brain and/or nervous system development; abnormal nociception.

We calculated the rate of phenotypic abnormalities in all mouse lines with reduced expression of genes identified in our prediXcan analysis ("SCZ-associated mouse lines"). We compared these to (1) the 'baseline' rate of phenotypic abnormalities across all 10,288 mouse lines; and (2) the rate of abnormalities in mouse lines associated with other disorders. To do this, we downloaded all publicly-available whole-blood-derived S-PrediXcan results (as of March 2018, see URLs). In total, we obtained data for 1,907 genes reaching $p < 5 \times 10^{-6}$, across 65 studies. We calculated rates of phenotypic abnormalities for each of these 65 studies.

Data Availability

Our CMC-derived DLPFC prediction models are publicly available at https://github.com/laurahuckins/CMC_DLPFC_prediXcan

Methods-only References

80. Barbeira, A. *et al.* MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results. *bioRxiv* (2016).

81. (GTex Consortium). GTEx Portal. (2015). at <http://gtexportal.org/home/documentationPage>

82. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.).* **348,** 648–660 (2015).

83. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

84. COCHRAN, W. G. THE COMPARISON OF PERCENTAGES IN MATCHED SAMPLES. *Biometrika* **37,** 256–266 (1950).

85. Yang, 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).

86. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11,** e1004219 (2015).

87. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17,** 142–53 (2012).

88. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).

89. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43,** D1049-1056 (2014).

90. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

91. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42,** D472-7 (2014).

92. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13,** 2129–2141 (2003).

93. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of

gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41,** D377-86 (2013).

94. BioCarta. MSigDB Collections. (2017).

95. Lin, G. N. *et al.* Spatiotemporal 16p11.2 Protein Network Implicates Cortical Late Mid-Fetal Brain Development and KCTD13- Cul3-RhoA Pathway in Psychiatric Diseases. *Neuron* **85,** 742–754 (2015).

96. Bahl, E., Koomar, T. & Michaelson, J. J. cerebroViz: An R package for anatomical visualization of spatiotemporal brain data. *Bioinformatics* **33,** btw726 (2016).

97. van der Weyden, L., White, J. K., Adams, D. J. & Logan, D. W. The mouse genetics toolkit: revealing function and mechanism. *Genome Biol.* **12,** 224 (2011).

98. Brown, S. D. M. & Moore, M. W. The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping. *Mammalian genome : official journal of the International Mammalian Genome Society* **23,** 632–640 (2012).

Editorial summary: This study uses gene expression predictors for dorso-lateral pre-frontal cortex and other brain regions to perform a transcriptomic imputation analysis of schizophrenia, identifying 413 genic associations across 13 brain regions and 36 significantly enriched pathways.