nature genetics

ARTICLES

https://doi.org/10.1038/s41588-019-0364-4

Corrected: Publisher Correction

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

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 T. 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] and Eli A. Stahl[1,2,3,4]

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 genome-wide association study (GWAS) loci and may disentangle the role of different tissues in disease development. We used the largest eQTL reference panel for the dorso-lateral prefrontal 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 postnatal expression.

GWASs 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 colocalization of eQTLs and GWAS loci have been developed[10–13]. However, these methods require simplifying assumptions about genetic architecture (that is, 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 (RNA-seq), 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

[1]Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [4]Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Vanderbilt University Medical Center, Nashville, TN, USA. [6]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, 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, UK. [14]A list of members and affiliations appears at the end of the paper. [15]University of North Carolina at Chapel Hill, 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, IL, USA. *e-mail: laura.huckins@mssm.edu

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 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 SCZ GWAS data, obtained through collaboration with the Psychiatric Genomics Consortium (PGC) SCZ working group, the 'CLOZUK2' cohort, and the iPSYCH-GEMS SCZ working group. We identified 413 genome-wide significant genic associations with SCZ in our PGC+CLOZUK2 sample, constituting 67 independent associations outside the MHC region. We demonstrated the relevance of these associations to SCZ etiopathology by using gene set enrichment analysis, and by examining the effects of manipulation of these genes in mouse models. Finally, we investigated the spatiotemporal expression of these genes by using a developmental transcriptome dataset, and identified distinct spatiotemporal 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 Figs. 1 and 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 that for the uncorrected data[14,18] (KS test; $P<2.2\times10^{-16}$; Supplementary Fig. 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-seq 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 12 GTEx-derived brain tissue models[15,21] (Fig. 1 and Supplementary Fig. 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 12 GTEx models (KS test, $P<2.2\times10^{-16}$ across all analyses; Fig. 1). Median R_R^2 values were significantly correlated with sample size of the original tissue set ($\rho=0.92$, $P=7.2\times10^{-6}$), the number of genes in the prediction model ($\rho=0.9$, $P=2.6\times10^{-5}$), and the number of significant 'eGenes' in each tissue type ($\rho=0.95$, $P=5.5\times10^{-7}$; Fig. 1c). Notably, these correlations persist after removing obvious outliers (Fig. 1c).

To estimate transancestral prediction accuracy, GREX was calculated for 162 African American individuals and 280 European individuals from the NIMH Human Brain Collection Core (HBCC) dataset (Supplementary Fig. 2c). R_R^2 values were higher on average in Europeans than in African Americans (average $R_{R\_EUR}^2=0.048$, $R_{R\_AA}^2=0.040$), but were significantly correlated between African Americans and Europeans ($\rho=0.78$, $P<2.2\times10^{-16}$, Pearson test; Supplementary Fig. 3).

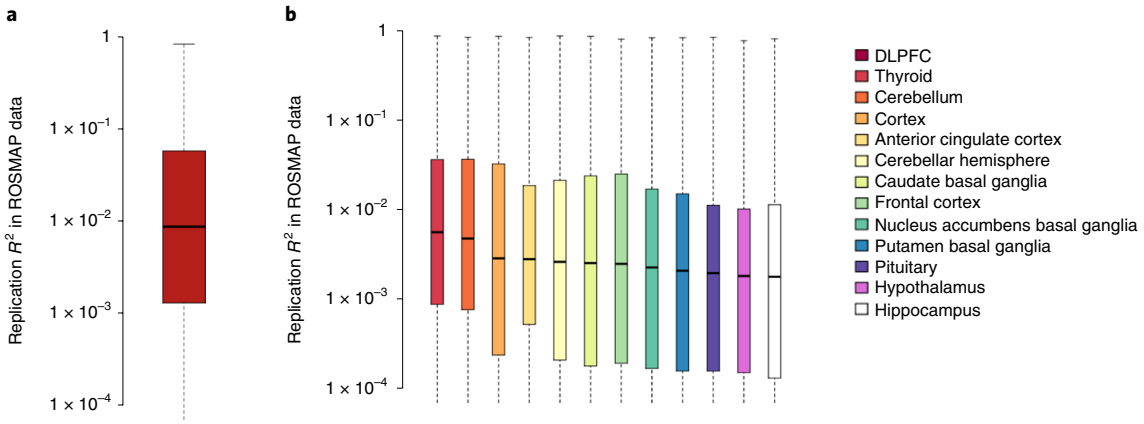
Application of transcriptomic imputation to schizophrenia. We used CMC DLPFC and 12 GTEx-derived brain tissue prediction models to impute GREX of 19,661 unique genes in cases and controls from the PGC-SCZ GWAS study[22]. Predicted expression levels were tested for association with SCZ. Additionally, 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 by using an inverse-variance-based approach in METAL. Our final analysis included 40,299 cases and 65,264 controls (Supplementary Fig. 4a).

We identified 413 genome-wide significant associations, representing 256 genes in 13 tissues (Fig. 2a). The largest number of associations was detected in the CMC-DLPFC GREX data (Fig. 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 Fig. 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 of 12 replicating loci are previous GWAS loci, compared with 38 of 88 nonreplicating loci. There was significant concordance between our discovery (PGC+CLOZUK2) and replication (iPSYCH-GEMS) samples; 72 of 100 genes have consistent direction of effect, including all 12 replicating genes (binomial $P=1.258\times10^{-5}$), and we found significant correlation of effect sizes ($P=1.784\times10^{-4}$; $\rho=0.036$) and $-\log_{10}P$ values ($P=1.073\times10^{-5}$; $\rho=0.043$).

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 and Fig. 2a,b). The largest signal was identified in the CMC-DLPFC GREX data (24 genes; Fig. 2c), followed by the putamen (seven genes). 19 out of 67 genes did not lie within 1 Mb of a previously genome-wide significant GWAS locus[22] (shown in bold in Table 1); of these, 5 of 19 genes were within 1 Mb of a locus that approached genome-wide significance ($P<5\times10^{-07}$). The remaining 14 genes all fall within nominally significant PGC-SCZ GWAS loci ($P<8\times10^{-04}$), but did not reach genome-wide significance.

We compared our CMC-DLPFC prediXcan associations statistics to COLOC results from our recent study[10,23]. Briefly, COLOC tests for colocalization 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[24]. We found a significant correlation between prediXcan P values and PP4 values; $\rho=0.35$, $P=2.3\times10^{-311}$. Thirty-one genes had 'strong' evidence of colocalization 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$=2.11\times10^{-104}$), and all had $P<1\times10^{-4}$. We identified 40 GWAS loci with no significant prediXcan associations; all of these loci also had strong evidence for no colocalization 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 gene sets comprising (1) hypothesis-driven pathways and



c

	Brain tissue	Number of samples	Number of genes	*N* significant eGenes
CMC	Dorso-lateral prefrontal cortex	646	10,929	12,813
GTex	Thyroid	278	11,180	10,610
	Cerebellum	103	10,007	4,528
	Cortex	96	9,166	2,768
	Anterior cingulate cortex	72	8,738	1,289
	Cerebellar hemisphere	89	9,458	3,403
	Caudate basal ganglia	100	9,152	2,612
	Frontal cortex	92	9,040	2,152
	Nucleus accumbens basal ganglia	93	8,921	2,202
	Putamen basal ganglia	82	8,765	1,653
	Pituitary	87	9,155	2,260
	Hypothalamus	81	8,555	1,253
	Hippocampus	81	8,540	1,164
Correlation with predictor performance		$\rho = 0.92$ $P = 7.2 \times 10^{-6}$	$\rho = 0.90$ $P = 2.6 \times 10^{-5}$	$\rho = 0.95$ $P = 5.5 \times 10^{-7}$
Correlation with predictor performance, excluding CMC DLPFC and GTEx-thyroid		$\rho = 0.57$ $P = 0.067$	$\rho = 0.84$ $P = 0.0012$	$\rho = 0.82$ $P = 0.0021$

Fig. 1 | Replication of DLPFC prediction models in independent data. Measured gene expression (ROSMAP RNA-seq) was compared with 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 $R_R^2 \geq 0.01$. Box plot midlines show means, edges show quartiles, and whiskers show full range of data. **b**, Distribution of R_R^2 values of 12 GTEx predictors in ROSMAP data. **c**, Table of sample sizes and *P*-value 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).

(2) general molecular database pathways. We corrected for multiple testing by using the Benjamini–Hochberg false discovery rate (FDR) correction[25].

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[26]. Enrichment of this large group of genes has been observed frequently in studies of SCZ[27,28] and autism[26,29]. There was a significant enrichment among our SCZ-associated genes and genes that have been shown to be intolerant to loss-of-function mutations[30] ($P = 5.86 \times 10^{-5}$) and with copy number variants (CNVs) associated with bipolar disorder[31] ($P = 7.92 \times 10^{-8}$), in line with a recent GWAS study of the same individuals[28].

Next, we performed an agnostic search for overlap between our SCZ-associated genes and ~8,500 molecular pathways collated from large, publicly available databases. Thirty-three pathways were significantly enriched after FDR correction (Table 2 and Supplementary 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[32–37]. Five pathways potentially related to porphyrin metabolism, regarding abnormal iron level in the spleen, liver, and kidney, are also significantly enriched, including two or five 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

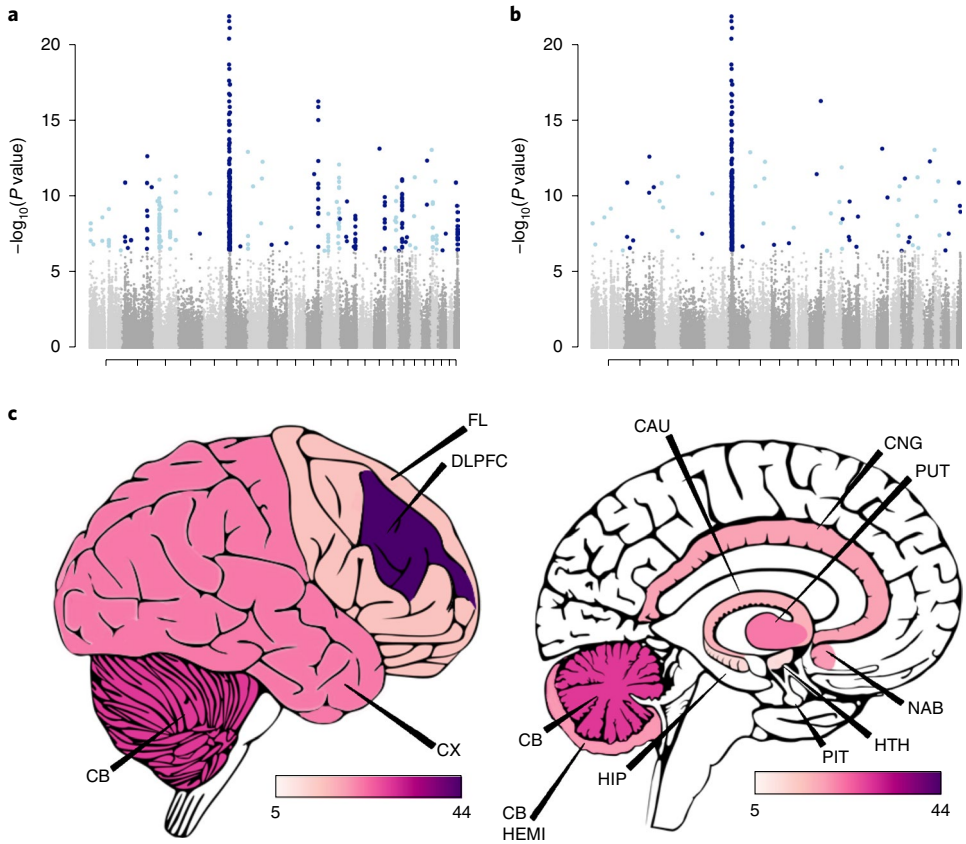


Fig. 2 | SCZ association results. a, 413 genes are associated with SCZ across 12 brain tissues. Each point represents one gene–tissue pair. **b**, 67 genes remain significant outside the MHC after stepwise conditional analysis. **c**, Number of genome-wide significant loci, outside the MHC region, identified in each brain region. These trends are partly driven by differences in power between brain regions. CB, cerebellum; CX cortex; FL, frontal cortex; DLPFC, dorso-lateral prefrontal 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.

of porphyric disorders, are also highly enriched ($P=2.2\times10^{-4}$, 2.6×10^{-4}, 4.1×10^{-4}), but do not pass FDR correction.

Hexosaminidase activity was enriched ($P=3.47\times10^{-5}$) in our results. This enrichment is not driven by a single highly associated gene, but 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[38]. Adult-onset HEX-A deficiency presents with neurological and psychiatric symptoms, notably including onset of psychosis and SCZ[39]. Five pathways corresponding to Ras and Rab signaling, protein regulation, and GTPase activity were enriched ($P<6\times10^{-5}$). These pathways have a crucial role in neuron cell differentiation[40] and migration[41], and have been implicated in the development of SCZ and autism[42–45]. We also find significant enrichment with protein phosphatase type 2A regulator activity ($P=5.24\times10^{-5}$), which was associated with major depressive disorder (MDD) and across MDD, bipolar disorder (BPD) and SCZ in the same large integrative analysis[46], and has been implicated in antidepressant response and serotonergic neurotransmission[47].

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[27] to DLPFC prediXcan results. Out of 460, 76 were nominally significant in the DLPFC prediXcan analysis, significantly more than would be expected by chance (binomial test, $P=8.75\times10^{-20}$). In particular, the Fromer et al. analysis highlighted six loci where expression levels of a single gene putatively affected SCZ 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 SCZ. 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[48–51]. We identified mutant mouse lines lacking expression of 37 out of 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. Out of 37 genes, 25 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

Table 1 | SCZ-associated genes following conditional analysis

Gene name	Tissue	BETA	*P* value	GVAR	Adjusted BETA	Adjusted OR
GNL3	Cerebellum	0.037	1.39×10^{-11}	0.115	0.012	1.012
THOC7	Cerebellum	−0.113	5.77×10^{-10}	0.010	−0.011	0.989
NAGA	Cerebellum	0.122	1.12×10^{-09}	0.009	0.011	1.011
TAC3	Cerebellum	−0.868	8.03×10^{-08}	0.000	−0.015	0.985
CHRNA2	Cerebellum	−0.016	1.63×10^{-07}	0.395	−0.010	0.990
ACTR5	Cerebellum	0.208	3.88×10^{-07}	0.019	0.029	1.029
INO80E	Frontal cortex	0.130	7.25×10^{-12}	0.009	0.012	1.013
PLPPR5	Frontal cortex	−0.672	2.58×10^{-09}	0.006	−0.053	0.948
FAM205A	Frontal cortex	0.043	1.21×10^{-08}	0.061	0.011	1.011
AC110781.3	Thyroid	0.342	1.31×10^{-13}	0.002	0.014	1.014
IMMP2L	Thyroid	−0.073	7.09×10^{-12}	0.046	−0.016	0.984
IGSF9B	Thyroid	−0.024	3.05×10^{-07}	0.156	−0.010	0.991
NMRAL1	Thyroid	0.038	4.03×10^{-07}	0.060	0.009	1.009
HIF1A	DLPFC	11.130	7.52×10^{-14}	0.000	0.148	1.159
TIMM29	DLPFC	11.207	9.27×10^{-14}	0.000	0.168	1.183
ST7-OT4	DLPFC	10.170	5.79×10^{-13}	0.001	0.318	1.374
H2AFY2	DLPFC	10.962	3.60×10^{-12}	0.000	0.191	1.211
STARD3	DLPFC	10.740	5.90×10^{-12}	0.001	0.304	1.355
CTC-471F3.5	DLPFC	8.535	1.11×10^{-11}	0.000	0.104	1.110
SF3A1	DLPFC	8.651	1.32×10^{-11}	0.000	0.083	1.086
ZNF512	DLPFC	10.312	1.32×10^{-11}	0.001	0.261	1.298
FURIN	DLPFC	−0.084	2.22×10^{-11}	0.022	−0.012	0.988
INHBA-AS1	DLPFC	8.399	2.24×10^{-11}	0.000	0.127	1.135
SF3B1	DLPFC	0.099	6.14×10^{-11}	0.014	0.012	1.012
EFTUD1P1	DLPFC	−0.092	1.81×10^{-10}	0.017	−0.012	0.988
MLH1	DLPFC	2.840	2.10×10^{-10}	0.001	0.069	1.071
GATAD2A	DLPFC	−0.044	2.18×10^{-10}	0.071	−0.012	0.988
METTL1	DLPFC	9.357	2.23×10^{-10}	0.000	0.166	1.181
DMC1	DLPFC	7.229	4.48×10^{-10}	0.000	0.130	1.139
RAD51D	DLPFC	7.612	2.11×10^{-09}	0.000	0.111	1.117
RERE	DLPFC	2.847	6.32×10^{-09}	0.000	0.036	1.037
PCCB	DLPFC	−0.044	2.05×10^{-08}	0.054	−0.010	0.990
CLCN3	DLPFC	0.141	2.96×10^{-08}	0.005	0.010	1.010
ATG101	DLPFC	8.086	4.90×10^{-08}	0.007	0.695	2.005
JRK	DLPFC	0.032	1.25×10^{-07}	0.091	0.010	1.010
PTPRU	DLPFC	−0.077	1.60×10^{-07}	0.016	−0.010	0.990
MARCKS	DLPFC	0.398	2.05×10^{-07}	0.001	0.015	1.015
TCF4	Anterior cingulate cortex	−0.059	5.22×10^{-13}	0.051	−0.013	0.987
DGKD	Anterior cingulate cortex	−0.937	2.63×10^{-11}	0.001	−0.022	0.979
C1QTNF4	Anterior cingulate cortex	−0.173	$1.37\times10-^{09}$	0.010	−0.017	0.983
PITPNA	Anterior cingulate cortex	−0.243	$1.77\times10-^{07}$	0.002	−0.010	0.990
FXR1	Caudate basal ganglia	0.439	$5.40\times10-^{12}$	0.001	0.017	1.017
ZDHHC1	Caudate basal ganglia	0.354	$5.36\times10-^{08}$	0.001	0.011	1.012
PDE4D	Cerebellar hemisphere	0.365	$6.81\times10-^{11}$	0.001	0.013	1.013

Continued

Table 1 | SCZ-associated genes following conditional analysis (Continued)

Gene name	Tissue	BETA	*P* value	GVAR	Adjusted BETA	Adjusted OR
DRD2	Cerebellar hemisphere	−0.182	$2.47\times10-^{10}$	0.004	−0.012	0.988
PITPNM2	Cerebellar hemisphere	−0.065	2.21×10^{-09}	0.028	−0.011	0.989
RINT1	Cerebellar hemisphere	0.086	6.32×10^{-09}	0.016	0.011	1.011
SRMS	Cerebellar hemisphere	−0.440	3.08×10^{-08}	0.001	−0.011	0.989
SETD6	Cerebellar hemisphere	−0.043	1.05×10^{-07}	0.054	−0.010	0.990
APOPT1	Cortex	−0.074	1.24×10^{-10}	0.026	−0.012	0.988
VSIG2	Cortex	−0.092	6.01×10^{-09}	0.013	−0.011	0.989
SDCCAG8	Cortex	−0.069	3.88×10^{-07}	0.002	−0.003	0.997
PIK3C2A	Cortex	−0.040	4.04×10^{-07}	0.365	−0.024	0.976
AS3MT	Frontal cortex	0.594	5.65×10^{-17}	0.001	0.017	1.017
FOXN2	Hippocampus	−0.250	2.65×10^{-07}	0.021	−0.036	0.964
RASIP1	Nucleus accumbens basal ganglia	0.055	3.80×10^{-08}	0.034	0.010	1.010
TCF23	Nucleus accumbens basal ganglia	−0.076	4.83×10^{-08}	0.019	−0.010	0.990
TTC14	Nucleus accumbens basal ganglia	−0.089	4.84×10^{-08}	0.013	−0.010	0.990
TYW5	Putamen basal ganglia	−0.080	2.63×10^{-13}	0.035	−0.015	0.985
SNX19	Putamen basal ganglia	0.031	1.31×10^{-12}	0.179	0.013	1.013
CIART	Putamen basal ganglia	0.090	6.78×10^{-10}	0.017	0.012	1.012
SH2D7	Putamen basal ganglia	0.096	7.89×10^{-09}	0.013	0.011	1.011
DGUOK	Putamen basal ganglia	0.255	8.26×10^{-08}	0.002	0.011	1.011
C12orf76	Putamen basal ganglia	0.031	2.27×10^{-07}	0.095	0.010	1.010
LRRC37A	Putamen basal ganglia	−0.035	2.69×10^{-07}	0.076	−0.010	0.991
AC005841.1	Pituitary	0.162	3.28×10^{-09}	0.005	0.011	1.011
RPS17	Pituitary	0.035	4.03×10^{-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}			

Sixty-seven non-MHC genes are significantly associated with SCZ following conditional analysis. Effect sizes (BETA) refer to predicted GREX in cases compared with controls. Effect sizes and odds ratios are also shown adjusted to 'unit' variance in gene expression. OR, odds ratio; DLPFC, dorso-lateral prefrontal cortex; GVAR, genetic variance.

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 with these 'baseline' GWAS comparisons. SCZ mouse lines also had higher rates of embryonic phenotypes, usually

Table 2 | Significantly enriched pathways and gene sets

Analysis	Gene set	Comp *P* value	FDR *P* value
Hypothesis driven	FMRP targets	1.96×10^{-08}	3.097×10^{-06}
	BP de novo CNV	7.92×10^{-08}	6.257×10^{-06}
	HIGH LOF intolerant	5.86×10^{-05}	0.00309
Agnostic	Increased spleen iron level	2.72×10^{-08}	0.000245
	Decreased IgM level	6.80×10^{-07}	0.00307
	Condensed chromosome	1.99×10^{-06}	0.00598
	Chromosome	2.80×10^{-06}	0.00632
	Abnormal spleen iron level	6.79×10^{-06}	0.00765
	Mitotic anaphase	6.39×10^{-06}	0.00765
	Mitotic metaphase and anaphase	5.13×10^{-06}	0.00765
	Resolution of sister chromatid cohesion	5.82×10^{-06}	0.00765
	Increased liver iron level	1.03×10^{-05}	0.0103
	Separation of sister chromatids	1.28×10^{-05}	0.0115
	Regulation of Rab GTPase activity	1.78×10^{-05}	0.0123
	Regulation of Rab protein signal transduction	1.78×10^{-05}	0.0123
	Protein phosphorylated amino acid binding	1.75×10^{-05}	0.0123
	Chromosome	2.57×10^{-05}	0.0165
	Hexosaminidase activity	3.47×10^{-05}	0.0174
	Abnormal learning memory conditioning	3.11×10^{-05}	0.0174
	Abnormal liver iron level	3.47×10^{-05}	0.0174
	Mitotic prometaphase	2.99×10^{-05}	0.0174
	M phase	3.70×10^{-05}	0.0176
	Positive regulation of Rab GTPase activity	5.93×10^{-05}	0.0232
	Rab GTPase activator activity	5.93×10^{-05}	0.0232
	Protein phosphatase type 2A regulator activity	5.24×10^{-05}	0.0232
	Replicative senescence	5.44×10^{-05}	0.0232
	Condensed nuclear chromosome	7.11×10^{-05}	0.0267
	Ubiquitin-specific protease activity	0.000104	0.0335
	Ras GTPase activator activity	9.61×10^{-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

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.

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[52]. Data were partitioned into eight developmental stages (four prenatal, four postnatal), and four brain regions[31,52] (Fig. 3a). SCZ-associated genes were significantly coexpressed in both prenatal and postnatal development and in all four brain regions, based on local connectedness[53] (Fig. 3b), global connectedness[53] (that is, average path length between genes; Supplementary Fig. 5), and network density (that is, number of edges; Supplementary Fig. 6). Examining pairwise gene expression correlation (Supplementary Fig. 7) and gene coexpression networks (Supplementary Fig. 8) for each spatiotemporal point indicated that the same genes do not drive this coexpression pattern throughout development, but rather, it appears that separate groups of genes drive early prenatal, late prenatal, and postnatal clustering.

To visualize this, we calculated *z* scores measuring the spatiotemporal specificity of gene expression for each SCZ-associated gene, across all 32 time points (Fig. 4). Genes clustered into four groups (Supplementary Fig. 9) with distinct spatiotemporal expression signatures. The largest cluster (cluster A, Fig. 4a, 29 genes) spanned early to late mid-prenatal development (4–24 weeks post conception (p.c.w.)), either across the whole brain (22 genes) or in regions 1–3 only (seven genes). Twelve genes were expressed in late prenatal development (Fig. 4d; 25–38 p.c.w.), ten genes were expressed in regions 1–3, postnatally and in the late prenatal period

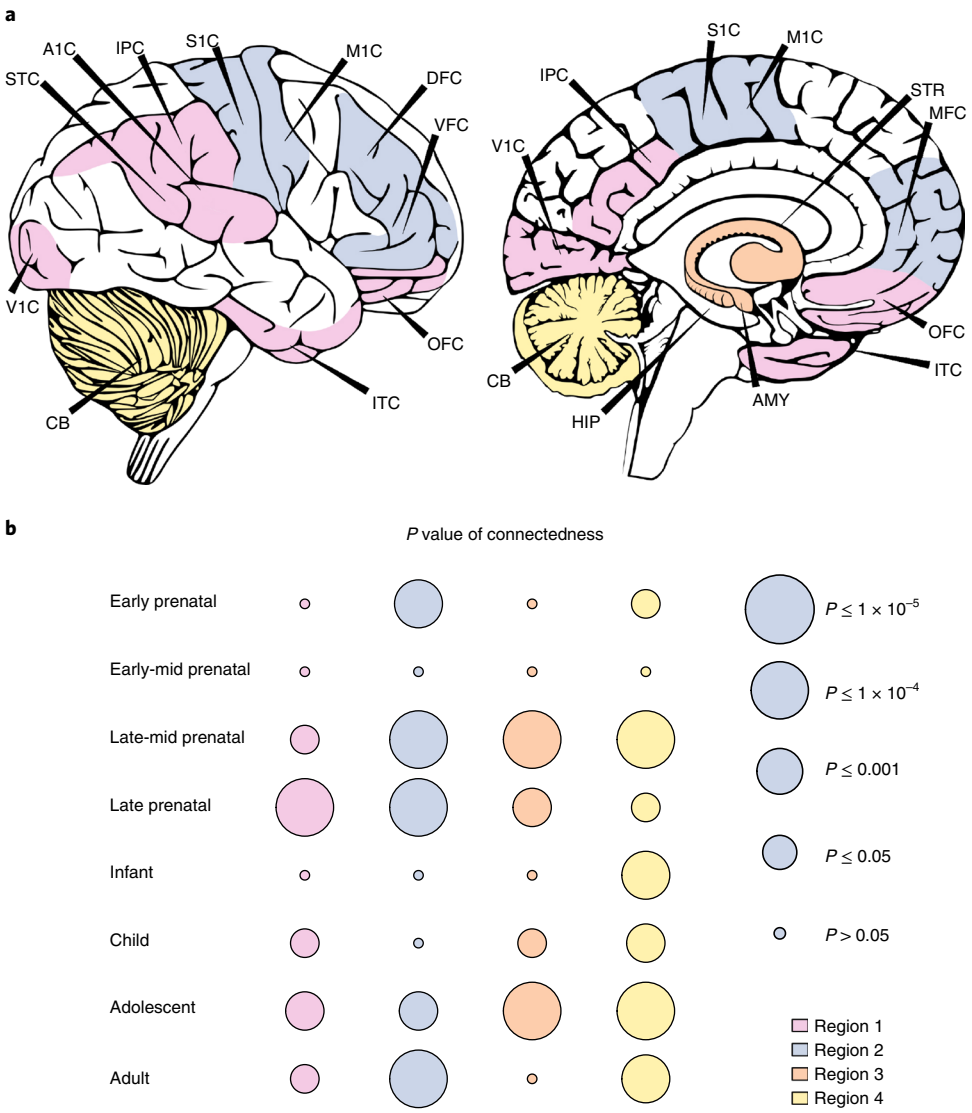


Fig. 3 | SCZ-associated genes are coexpressed throughout development and across brain regions. a, Brain tissues selected for each of four BrainSpan regions. BrainSpan includes 525 samples from 43 unique individuals. IPC, inferior parietal cortex; V1C, primary visual cortex; ITC, inferior temporal cortex; OFC, orbital frontal cortex; STC, posterior (caudal) superior temporal cortex; A1C, primary auditory cortex; S1C, primary somatosensory cortex; M1C, primary motor cortex; DFC, dorsolateral prefrontal cortex; VFC, ventrolateral prefrontal cortex; MFC, medial prefrontal cortex; HIP, hippocampus; AMY, amygdala; STR, striatum; CB, cerebellum. Region 1: IPC, V1C, ITC, OFC, STC, A1C; region 2: S1C, M1C, DFC, VFC, MFC; region 3: HIP, AMY, STR; region 4: CB. **b**, Average clustering coefficients were calculated for all pairs of SCZ-associated genes, and compared with average clustering coefficients for 100,000 permuted gene networks to obtain empirical significance levels.

(Fig. 4c), and 15 genes were expressed throughout development (Fig. 4b), either specifically in region 4 (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 SCZ[54]. Genes in clusters A and B (clusters with prenatal 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 (genes with late prenatal and postnatal 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 prenatally in humans, or genes expressed pre- and postnatally, were more likely to have any behavioral or neurological phenotypes than mutant mice lacking expression of genes expressed primarily in the third trimester or postnatally ($P = 1.7 \times 10^{-4}$) (Supplementary Fig. 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-seq and circumvents distortions in measures of gene expression stemming from errors of measurement or

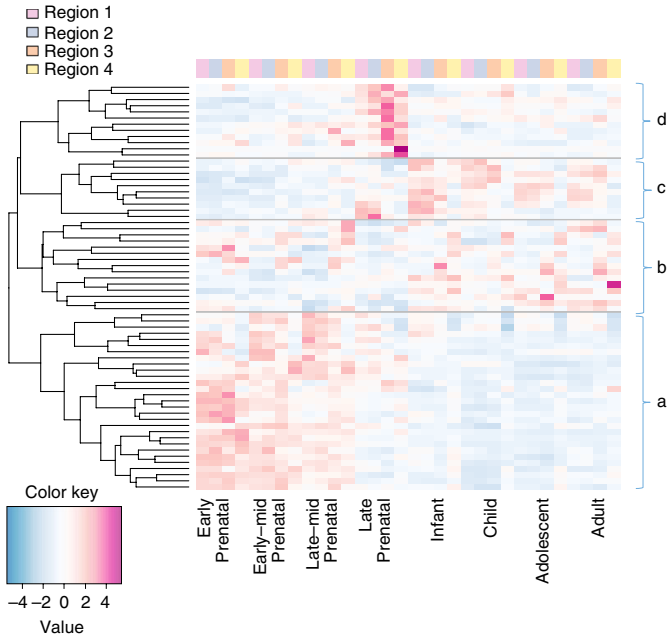


Fig. 4 | Gene expression patterns for SCZ-associated genes cluster into four groups, relating to distinct spatiotemporal expression. a, Brain regions are shown in Fig. 3a. Twenty-nine genes are expressed in the early to mid-prenatal period (4–24 p.c.w.). **b**, Fifteen genes are expressed throughout development. Subclusters correspond to either specific expression in region 4 or expression across the brain. **c**, Ten genes are expressed in the late-prenatal (25–38 p.c.w.) and postnatal period. **d**, Twelve genes are expressed in the late prenatal period (25–39 p.c.w.).

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 imputation has exciting advantages for gene discovery as well as downstream applications[15,55,56]; however, the relative merits of existing methodologies are as yet underexplored. 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 oversimplified[2,15]. Fundamentally, transcriptomic imputation methods model only the genetically regulated portion of gene expression and thus 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 eQTL architecture.

We compared transcriptomic imputation accuracy in European and African American individuals and found that our models were applicable to either ancestry with only a small decrease in accuracy. Common SNPs shared across ancestries 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[57,58]. We believe that these analyses should be carried out in non-European cohorts.

We applied the CMC DLPFC and GTEx-derived prediction models to SCZ 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 (Fig. 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 colocalization estimates ('PP4') from COLOC. Importantly, GWAS loci with no significant prediXcan associations also had no evidence for colocalization 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. One caveat is that four of our associations (SNX19, NAGA, TYW5, and GNL3) have no evidence for colocalization in COLOC results, or after 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 SCZ GWAS[59], 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[55] identified 38 non-MHC genes associated at tissue-level significance with SCZ in CMC- and GTEx-derived brain tissues (that is, 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, that is, did not fall within 1 Mb of a previous GWAS locus (including five of seven 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), though we note that nearly colinear 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 significantly 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 SCZ, 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 that is shared by both disorders, but 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 SCZ, as described in the likely eponymous mid-19th century poem 'Porphyria's Lover', by Robert Browning[60], and have been cited as a likely diagnosis for the various psychiatric and metabolic ailments of Vincent van Gogh[61–66] and King George III (ref. [67]).

Finally, we assessed patterns of expression for the 67 SCZ-associated genes throughout development using spatiotemporal transcriptomic data obtained from BrainSpan. We identified four clusters of genes, with expression in four distinct spatiotemporal regions, ranging from early prenatal to strictly postnatal 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[68–70] and the earlier onset of cognitive impairments[71–74]. 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[75–77]; therefore, our results should spur interest in extending transcriptomic imputation data and/or methods to early development[75]. 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[78] (for example, methylation or histone modification), imputing different tissues or different exposures (for example, age, smoking, or trauma) and modeling trans/coexpression effects. It remains critical to leverage transcriptomic imputation 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 SCZ risk.

URLs. 'CoCo', an R implementation of GCTA-COJO, https://github.com/theboocock/coco/; Gene2pheno, gene2pheno.org; publicly available whole-blood-derived S-PrediXcan results (as of March 2018), https://github.com/laurahuckins/CMC_DLPFC_prediXcan.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0364-4.

Received: 28 June 2017; Accepted: 30 January 2019;
Published online: 25 March 2019

References

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. Preprint at https://www.biorxiv.org/content/10.1101/129429v2 (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**, 1091–1098 (2015).
16. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
17. Geschwind, D. H. & Flint, J. Genetics and genomics of psychiatric disease. *Science* **349**, 1489–94 (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**, 882–883 (2012).
19. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. 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, R. S. Overview and findings from the rush memory and aging project. *Curr. Alzheimer Res.* **9**, 646–663 (2012).
21. Mele, M. et al. The human transcriptome across tissues and individuals. *Science* **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.* https://doi.org/10.1016/j.ajhg.2018.04.011 (2018).
24. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
25. Benjamin, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
26. Darnell, J. C. et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
27. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
28. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
29. Sanders, S. J. First glimpses of the neurobiology of autism spectrum disorder. *Curr. Opin. Genet. Dev.* **33**, 80–92 (2015).
30. Monkol, Lek. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
31. Malhotra, D. et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron* **72**, 951–963 (2011).
32. 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).
33. 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**, 1501–1505 (2006).

34. Ventura, P. et al. A challenging diagnosis for potential fatal diseases: recommendations for diagnosing acute porphyrias. *Eur. J. Intern. Med.* **25**, 497–505 (2014).
35. Pischik, E. & Kauppinen, R. An update of clinical management of acute intermittent porphyria. *Appl. Clin. Genet.* **8**, 201–214 (2015).
36. Kumar, B. Acute intermittent porphyria presenting solely with psychosis: a case report and discussion. *Psychosomatics* **53**, 494–498 (2012).
37. 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).
38. Kaback, M. M. & Desnick, R. J. *Hexosaminidase A Deficiency*: *GeneReviews* (University of Washington, Seattle, 1993).
39. Osama, S. Late onset Tay-Sachs disease presenting as a brief psychotic disorder with catatonia: a case report and review of literature. *Jefferson J. Psych.* **15**, 4 (2000).
40. Skaper, S. D. in *Brain Protection in Schizophrenia, Mood and Cognitive Disorders* (ed. Ritsner, M. S.) 135–165 (Springer Science & Business Media, 2010).
41. Castellano, E. et al. RAS signalling through PI3-Kinase controls cell migration via modulation of Reelin expression. *Nat. Commun.* **7**, 11245 (2016).
42. Gururajan, A. & Buuse, M. van den. Is the mTOR-signalling cascade disrupted in Schizophrenia? *J. Neurochem.* **129**, 377–387 (2014).
43. Ritsner, M. S. *Brain Protection in Schizophrenia, Mood and Cognitive Disorder*s (Springer Science & Business Media, 2010).
44. 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).
45. Glessner, J. T. et al. Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc. Natl Acad. Sci. USA* **107**, 10584–10589 (2010).
46. 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).
47. Bauman, A. L. et al. Cocaine and antidepressant-sensitive biogenic amine transporters exist in regulated complexes with protein phosphatase 2A. *J. Neurosci.* **20**, 7571–7578 (2000).
48. 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).
49. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
50. Howe, D. G. et al. ZFIN, the zebrafish model organism database: increased support for mutants and transgenics. *Nucleic Acids Res.* **41**, D854–D860 (2013).
51. 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).
52. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
53. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
54. Nguyen, H. T. et al. Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome Med.* **9**, 114 (2017).
55. 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).
56. Gottlieb, A., Daneshjou, R., DeGorter, M., Montgomery, S. & Altman, R. Population-specific imputation of gene expression improves prediction of pharmacogenomic traits for African Americans. Preprint at https://www.biorxiv.org/content/10.1101/115451v1 (2017).
57. Need, A. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends Genet* **25**, 489–494 (2009).
58. Popejoy, A. & Fullerton, S. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
59. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. Preprint at https://www.biorxiv.org/content/10.1101/068593v1 (2016).
60. Browning, R. in *The Poems of Robert Browning* (eds Porter, C. & Clarke, H. A.) 257–271 (Thomas Y. Cromwell and Company, 1896).
61. Loftus, L. S. & Arnold, W. N. Vincent van Gogh's illness: acute intermittent porphyria? *BMJ* **303**, 1589–1591 (1991).
62. Strik, W. K. The psychiatric illness of Vincent van Gogh. *Nervenarzt* **68**, 401–409 (1997).
63. Arnold, W. N. The illness of Vincent van Gogh. *J. Hist. Neurosci.* **13**, 22–43 (2004).
64. Hughes, J. R. A reappraisal of the possible seizures of Vincent van Gogh. *Epilepsy Behav.* **6**, 504–510 (2005).
65. Bhattacharyya, K. B. & Rai, S. The neuropsychiatric ailment of Vincent van Gogh. *Ann. Indian Acad. Neurol.* **18**, 6–9 (2014).
66. Correa, R. Vincent van Gogh: A pathographic analysis. *Med. Hypotheses* **82**, 141–144 (2014).
67. Peters, T. J. & Beveridge, A. The madness of King George III: a psychiatric re-assessment. *Hist. Psychiatry* **21**, 20–37 (2010).
68. Szatkiewicz, J. P. et al. Copy number variation in schizophrenia in Sweden. *Mol. Psychiatry* **19**, 762–773 (2014).
69. Fromer, M. et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
70. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
71. Keefe, R. S. E. & Fenton, W. S. How should DSM-V criteria for schizophrenia include cognitive impairment? *Schizophr. Bull.* **33**, 912–920 (2007).
72. 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).
73. Gold, J. M. Cognitive deficits as treatment targets in schizophrenia. *Schizophr. Res.* **72**, 21–28 (2004).
74. Cannon, M. et al. Evidence for early-childhood, pan-developmental impairment specific to schizophreniform disorder. *Arch. Gen. Psychiatry* **59**, 449 (2002).
75. 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).
76. Glass, D. et al. Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* **14**, R75 (2013).
77. Colantuoni, C. et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2012).
78. Gusev, A. et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Preprint at https://www.biorxiv.org/content/10.1101/067355v1 (2016).

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: P. Sklar, J. Buxbaum (Icahn School of Medicine at Mount Sinai), B. Devlin, D. Lewis (University of Pittsburgh), R. Gur, C.-G. Hahn (University of Pennsylvania), K. Hirai, H. Toyoshiba (Takeda Pharmaceuticals Company Limited), E. Domenici, L. Essioux (F. Hoffman-La Roche Ltd), L. Mangravite, M. Peters (Sage Bionetworks), T. Lehner, B. 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 acknowledges 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 September 5, 2016. BrainSpan: Atlas of the Developing Human Brain (Internet). Funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01.

H.K.I. was supported by R01 MH107666-01.

Author contributions

L.M.H. designed the study and specific subanalyses, ran analyses, and wrote the manuscript. A.D. designed and ran analyses and contributed to the writing group.

D.M.R. contributed to study and analytical design, and writing. G.H. contributed to analytical design and writing. W.W., H.T.N., and J.B. designed and ran specific analyses. A.F.P., V.M.R., T.D.A., K.G., M.F. all ran specific analyses. S.K.S. designed the study and analyses and contributed to the writing group. P.R. and R.K. designed the study and contributed data. E.D. designed the study, contributed data, and contributed to the writing group. E.R.G. designed specific analyses, and contributed to the writing group. S.P. designed the study. All three consortia (CMC, PGC-SCZ, iPSYCH-GEMS) contributed data. D.D., A.D.B., J.T.R.W., M.C.O'D., M.J.O. contributed data, advised on analyses, and contributed to the writing group. P. Sullivan advised on analyses and contributed to the writing group. B.D. designed the study, contributed data, advised on analyses, and contributed to the writing group. N.J.C. and H.K.I. designed the study, advised on analyses, and contributed to the writing group. P. Sklar and E.A.S. designed the study and specific analyses, ran analyses, and contributed to the writing group.

Competing interests

E.D. has received research support from Roche during 2016–2018. T.W. has acted as advisor and lecturer to H. Lundbeck A/S. All other authors declare no conflicts of interest.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0364-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.M.H.

CommonMind Consortium

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] and Laurent Essioux[28]

iPSYCH-GEMS Schizophrenia 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] and Christine Søholm Hansen[8,36]

The 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 Chong[75], 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 Eichhammer[94], 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-Rodríguez[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], Colm McDonald[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. Murphy[160], 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'Callaghan[172], Colm O'Dushlaine[38], F. Anthony O'Neill[173], Sang-Yun Oh[174], Ann Olincy[126], Line Olsen[8,64], Jim Van Os[162,175], Christos Pantelis[68,176], George N. Papadimitriou[86], Sergi Papiol[55], Elena Parkhomenko[3], Michele T. Pato[131], Tiina Paunio[177,178], Milica Pejovic-Milovancevic[179], 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. Richards[6], 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], Juha Veijola[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], 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], Todd Lencz[214,215,216], Douglas F. Levinson[51], Qingqin S. 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] and Michael C. O'Donovan[6,42]

[20]Integrated Technology Research Laboratories, Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, Fujisawa, Japan. [21]Neuropsychiatry Section, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [22]Neuropsychiatric Signaling Program, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [23]Psychiatry, JJ Peters Virginia Medical Center, Bronx, NY, USA. [24]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, NY, USA. [25]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [26]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [27]CNS Drug Discovery Unit, Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, Fujisawa, Japan. [28]F. Hoffman-La Roche Ltd, Basel, Switzerland. [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. [37]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [38]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [39]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [40]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [41]Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [42]NationalCentre for Mental Health, Cardiff University, Cardiff, UK. [43]Eli Lilly and Company Limited, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey, UK. [44]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK. [45]Center for BiologicalSequence Analysis, Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark. [46]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. [47]Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, Stockholm, Sweden. [48]Department of Psychiatry, Diakonhjemmet Hospital, Oslo, Norway. [49]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [50]State Mental Hospital, Haar, Germany. [51]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [52]Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, GA, USA. [53]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, USA. [54]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [55]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Gottingen, Germany. [56]Department of Medical Genetics, University of Pécs, Pécs, Hungary. [57]Szentagothai Research Center, University of Pécs, Pécs, Hungary. [58]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [59]Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA, USA. [60]University Medical Center Groningen, Department of Psychiatry, University of Groningen, Groningen, the Netherlands. [61]School of Nursing, Louisiana State University Health Sciences Center, New Orleans, LA, USA. [62]Athinoula A. Martinos Center, Massachusetts General Hospital, Boston, MA, USA. [63]Center for Brain Science, Harvard University, Cambridge, MA, USA. [64]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [65]Department of Psychiatry, University of California at San Francisco, San Francisco, CA, USA. [66]University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, Utrecht, the Netherlands. [67]Centre Hospitalier du Rouvray and INSERM U1079 Faculty of Medicine, Rouen, France. [68]Schizophrenia Research Institute, Sydney, New South Wales, Australia. [69]School of Psychiatry, University of New South Wales, Sydney, New South Wales, Australia. [70]Royal Brisbane and Women's Hospital, University of Queensland, Brisbane, Queensland, Australia. [71]Institute of Psychology, Chinese Academy of Science, Beijing, 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, Singapore. [76]Department of Psychiatry, Washington University, St. Louis, MO, 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, France. [78]Blue Note Biosciences, Princeton, NJ, USA. [79]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [80]Department of Psychological Medicine, Queen Mary University of London, London, UK. [81]Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London, UK. [82]Sheba Medical Center, Tel Hashomer, Israel. [83]Department of Genomics, Life and Brain Center, Bonn, Germany. [84]Institute of Human Genetics, University of Bonn, Bonn, Germany. [85]AppliedMolecular Genomics Unit, VIB Department of Molecular Genetics, University of Antwerp, Antwerp, Belgium. [86]First Department of Psychiatry, University of Athens Medical School, Athens, Greece. [87]Department of Psychiatry, University College Cork, Co, Cork, Ireland. [88]Department of Medical Genetics, Oslo University Hospital, 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, IL, USA. [91]Department of Psychiatry and Behavioral Sciences, North Shore University Health System, Evanston, IL, USA. [92]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. [93]Department of Child and Adolescent Psychiatry, University Clinic of Psychiatry, Skopje, Republic of Macedonia. [94]Department of Psychiatry, University of Regensburg, Regensburg, Germany. [95]Department of General Practice, Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland. [96]Folkhälsan Research Center, Helsinki, Finland, Biomedicum Helsinki, Helsinki, Finland. [97]National Institute for Health and Welfare, Helsinki, Finland. [98]Translational Technologies and Bioinformatics, Pharma Research and Early Development, F. Hoffman-La Roche, Basel, Switzerland. [99]Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA. [100]Department of Psychiatry, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA. [101]Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, VA, USA. [102]Mental Health Service Line, Washington VA Medical Center, Washington, DC, USA. [103]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Mannheim, Germany. [104]Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands. [105]Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA. [106]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, USA. [107]Department of Psychiatry, University of Halle, Halle, Germany. [108]Department of Psychiatry, University of Munich, Munich, Germany. [109]Departments of Psychiatry and Human and Molecular Genetics, INSERM, Institut

de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris, France. [110]Mental Health Research Centre, Russian Academy of Medical Sciences, Moscow, Russia. [111]Neuroscience Therapeutic Area, Janssen Research and Development, Raritan, NJ, USA. [112]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia. [113]Academic Medical Centre University of Amsterdam, Department of Psychiatry, Amsterdam, the Netherlands. [114]Illumina, La Jolla, CA, USA. [115]Priority Research Centre for Health Behaviour, University of Newcastle, Newcastle, New South Wales, Australia. [116]School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle, New South Wales, Australia. [117]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland. [118]Department of Genetics, Harvard Medical School, Boston, MA, USA. [119]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan. [120]Regional Centre for Clinical Researchin Psychosis, Department of Psychiatry, Stavanger University Hospital, Stavanger, Norway. [121]Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, Spain. [122]Centre for Medical Research, The University of Western Australia, Perth, Western Australia, Australia. [123]The Perkins Institute for Medical Research, The University of Western Australia, Perth, Western Australia, Australia. [124]Department of Medical Genetics, Medical University, Sofia, Bulgaria. [125]Department of Psychology, University of Colorado Boulder, Boulder, CO, USA. [126]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. [127]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [128]Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada. [129]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russia. [130]Latvian Biomedical Research and Study Centre, Riga, Latvia. [131]Department of Psychiatry and Zilkha Neurogenetics Institute, Keck School of Medicine at University of Southern California, Los Angeles, CA, USA. [132]Faculty of Medicine, Vilnius University, Vilnius, Lithuania. [133]Department of Biology and Medical Genetics, 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic. [134]Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris, France. [135]Duke-NUS Graduate Medical School, Singapore, Singapore. [136]Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem, 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, Sichuan, China. [139]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA. [140]Department of Psychiatry, Columbia University, New York, New York, NY, USA. [141]Priority Centre for Translational Neuroscience and Mental Health, University of Newcastle, Newcastle, New South Wales, Australia. [142]Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University in Szczecin, Szczecin, Poland. [143]Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland. [144]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [145]Department of Psychiatry, University of Bonn, 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, Paris, France. [147]Department of Genomics Mathematics, University of Bonn, Bonn, Germany. [148]Research Unit, Sørlandet Hospital, Kristiansand, Norway. [149]Department of Psychiatry, Harvard Medical School, Boston, MA, USA. [150]VA Boston Health Care System, Brockton, MA, 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, UK. [153]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [154]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [155]Massachusetts Mental Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Boston, MA, USA. [156]Estonian Genome Center, University of Tartu, Tartu, Estonia. [157]School of Psychology, University of Newcastle, Newcastle, New South Wales, Australia. [158]First Psychiatric Clinic, Medical University, Sofia, Bulgaria. [159]Department P, Aarhus University Hospital, Risskov, Denmark. [160]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin, Ireland. [161]King's College London, London, UK. [162]Maastricht University Medical Centre, South Limburg Mental Health Research and TeachingNetwork, EURON, Maastricht, the Netherlands. [163]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [164]Max Planck Institute of Psychiatry, Munich, Germany. [165]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [166]Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany. [167]Department of Psychiatry, Queensland Brain Institute and Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, Australia. [168]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [169]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [170]Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN, USA. [171]Department of Clinical Sciences, Psychiatry, Umeå University, Umeå, 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, UK. [174]Lawrence Berkeley National Laboratory, University of California at Berkeley, Berkeley, CA, USA. [175]Institute of Psychiatry, King's College London, London, UK. [176]Melbourne Neuropsychiatry Centre, University of Melbourne & Melbourne Health, Melbourne, Victoria, Australia. [177]Department of Psychiatry, University of Helsinki, Helsinki, Finland. [178]Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland. [179]Medical Faculty, University of Belgrade, Belgrade, Serbia. [180]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA. [181]Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Helsinki, Finland. [182]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. [183]Department of Psychiatry, University of Oxford, Oxford, UK. [184]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [185]Pharma Therapeutics Clinical Research, Pfizer Worldwide Research and Development, Cambridge, MA, USA. [186]Department of Psychiatry and Psychotherapy, University of Gottingen, Göttingen, Germany. [187]Psychiatry and Psychotherapy Clinic, University of Erlangen, Erlangen, Germany. [188]Hunter New England Health Service, Newcastle, New South Wales, Australia. [189]School of Biomedical Sciences, University of Newcastle, Newcastle, New South Wales, Australia. [190]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [191]University of Iceland, Landspitali, National University Hospital, Reykjavik, Iceland. [192]Department of Psychiatry and Drug Addiction, Tbilisi State Medical University (TSMU), Tbilisi, Georgia. [193]Research and Development, Bronx Veterans Affairs Medical Center, New York, NY, USA. [194]WellcomeTrust Centre for Human Genetics, Oxford, UK. [195]deCODE Genetics, Reykjavik, Iceland. [196]Department of Clinical Neurology, Medical University of Vienna, Wien, Austria. [197]Lieber Institute for Brain Development, Baltimore, MD, USA. [198]Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, the Netherlands. [199]Berkshire Healthcare NHS Foundation Trust, Bracknell, UK. [200]Section of Psychiatry, University of Verona, Verona, Italy. [201]Department of Psychiatry, University of Oulu, Oulu, Finland. [202]University Hospital of Oulu, Oulu, Finland. [203]Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland. [204]Health Research Board, Dublin, Ireland. [205]School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth, Western Australia, Australia. [206]Computational Sciences CoE, Pfizer Worldwide Research and Development, Cambridge, MA, USA. [207]Human Genetics, Genome Institute of Singapore, A*STAR, Singapore, Singapore. [208]University College London, London, UK. [209]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany. [210]Department of Genetics, The Hebrew University of Jerusalem, Jerusalem, Israel. [211]NeuroscienceDiscovery and Translational Area, Pharma Research and Early Development, F. Hoffman-La Roche, 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, Western Australia, Australia. [213]Virginia Institute for Psychiatric and Behavioral Genetics, Departments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA. [214]The Feinstein Institute for Medical Research, Manhasset, NY, USA. [215]The Hofstra NS-LIJ School of Medicine, Hempstead, NY, USA. [216]The Zucker Hillside Hospital, Glen Oaks, NY, USA. [217]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. [218]Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, Australia. [219]Center for HumanGenetic Research and Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [220]Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam, the Netherlands. [221]Department of Complex Trait Genetics, Neuroscience Campus Amsterdam, VU University Medical Center Amsterdam, Amsterdam, the Netherlands. [222]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus

Amsterdam, VU University, Amsterdam, the Netherlands. [223]University of Aberdeen, Institute of Medical Sciences, Aberdeen, UK. [224]Departments of Psychiatry, Neurology, Neuroscience and Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. [225]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

Methods

Creating gene expression predictors for the dorsolateral prefrontal cortex. *eQTL data.* Genotype and RNA-seq 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 modeling with voom-derived regression weights. Details on genotyping, imputation and RNA-seq generation can be found in the CommonMind Consortium (CMC) flagship paper[14].

The samples used here include 254 SCZ 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) cutoff was applied. Variants were filtered to remove any SNPs in high linkage disequilibrium ($r^2 > 0.9$), 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 quality control 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 Fig. 2a). Weights for these variants are calculated using regression in a ten-fold cross-validation framework. All cross-validation folds were balanced for diagnoses, ancestry, 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 ten cross-validation folds; this correlation was termed cross-validation R^2 or R_{cv}^2. Genes with $R_{cv}^2 > 0.01$ ($\sim P < 0.05$) 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 modeling (BSLMM; TWAS), and linear regression, using the best eQTL for each gene (Supplementary Fig. 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 (Kolmogorov–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 data sets.

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 Fig. 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:

R_R^2 calculation:

$$\begin{aligned} R_{R1}^2 &= (M \sim GREX + PC_1 + PC_2 + \ldots + PC_{10}) \\ R_{R2}^2 &= (M \sim PC_1 + PC_2 + \ldots + PC_{10}) \\ R_R^2 &= R_{R1}^2 - R_{R2}^2 \end{aligned} \quad (1)$$

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 Fig. 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 were used to compare with 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 12 GTEx brain tissue models (KS test, $P < 2.2 \times 10^{-16}$; Fig. 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 Fig. 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 Fig. 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 Fig. 3d).

Extension to summary statistics. Transcriptomic imputation may be applied to summary statistics instead of raw data, in instances where raw data are unavailable. However, this method suffers from slightly reduced accuracy, requires covariance matrices calculated in an ancestrally matched reference population[24] (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[24]) 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[24]. Covariances for all variants included in the DLPFC predictor models were computed using S-PrediXcan[24]. 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 Fig. 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 Fig. 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 Fig. 4). Of 53 cohorts, 52 (35,079 cases, 46,441 controls) were obtained through collaboration with the Psychiatric Genomics Consortium and are described in the 2014 PGC SCZ GWAS[22]. The remaining cohort, referred to as CLOZUK2, constitutes the largest single cohort of individuals with SCZ (5,220 cases and 18,823 controls), collected as part of an effort to investigate treatment-resistant SCZ[59].

Of 53 datasets, 50 included individuals of European ancestry, and three datasets included 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[22,59]. All sample collections complied with ethical regulations. Details regarding ethical compliance and consent procedures may be found in the original manuscripts describing these collections[22,59].

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 SCZ working group (effective sample size 14,169.5; Supplementary Fig. 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 (http://gtexportal.org/home/documentationPage)[15,21,79] (Fig. 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[24]'), as described previously.

Meta-analysis. Meta-analysis was carried out across all 53 cohorts using METAL[80]. Cochran's Q test for heterogeneity was implemented in METAL[80,81], 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, that is, 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[82] 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 ($p \leq 1 \times 10^{-6}$), as the root-effective sample size[80] (N_{eff}, equation (2)) weighted average correlation across all cohorts where we had access to dosage data.

Effective sample size, N_{eff}:

$$N_{eff} = \frac{4}{\left(\frac{1}{N_{cases}} + \frac{1}{N_{controls}}\right)} \quad (2)$$

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, equation (3)).

GREX beta adjustment:

$$\beta = \beta_{CoCo} \times \sqrt{GVAR} \quad (3)$$

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[83]. *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[14,59,68,84].
2. An agnostic analysis, including ~8,500 gene sets collated from publicly available databases including GO[85,86], KEGG[87], REACTOME[88], PANTHER[89,90], BIOCARTA (MSigDB Collections, 2017), and MGI[51]. 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 investigated spatiotemporal expression of our associated genes using publicly available developmental transcriptome data, obtained from the BrainSpan consortium[52]. We partitioned these data into biologically relevant spatiotemporal data sets[91], corresponding to four general brain regions, the frontal cortex, temporal and parietal regions, sensory-motor regions, and subcortical regions[92] (Fig. 3a), and eight developmental time points (four prenatal, four postnatal)[91].

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 coexpressed. 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 coexpression network was created, with edges connecting all coexpressed 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[53]; third, the Watts–Strogatz clustering coefficient, to measure tightness of the clusters within the network[53]. For each spatiotemporal time point, we plotted gene-pair expression correlation (Supplementary Fig. 7) and coexpression networks (Supplementary Fig. 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 Fig. 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[51]), EuroPhenome[48,93], Mouse Genome Project (MGP[48,49]), International Mouse Phenotyping Consortium (IMPC[94]), and Infection and Immunity Immunophenotyping (3I[94]). 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.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

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

References

79. Ardlie, K. G. et al. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
80. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
81. COCHRAN, W. G. THE comparison of percentages in matched samplES. *Biometrika* **37**, 256–266 (1950).
82. 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).
83. 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).
84. Kirov, G. et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2012).
85. Ashburner, M. et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.* **25**, 25–29 (2000).
86. The Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2014).
87. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
88. Croft, D. et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
89. Thomas, P. D. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
90. 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–D386 (2013).
91. 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).
92. Bahl, E., Koomar, T. & Michaelson, J. J. cerebroViz: An R package for anatomical visualization of spatiotemporal brain data. *Bioinformatics* **33**, btw726 (2016).
93. 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).
94. Brown, S. D. M. & Moore, M. W. The international mouse phenotyping consortium: past and future perspectives on mouse phenotyping. *Mamm. Genome* **23**, 632–640 (2012).

nature research

Corresponding author(s): Laura M Huckins

☒ Initial submission ☐ Revised version ☐ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

All available schizophrenia cases and controls in the Psychiatric Genomics Consortium (PGC) and CLOZUK2 studies were used in this study.

2. Data exclusions

Describe any data exclusions.

No data were excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.

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. We tested for replication of all Bonferroni-significant genes identified in our CMC-DLPFC analysis. Of 100 genes, 12 replicated in the iPSYCH-GEMS data, significantly more than might be expected by chance (binomial test, p=0.0043).

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Cases and controls are organized into ancestrally-matched cohorts. These matchings were carried out by the PGC and CLOZUK2 groups.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable to this study. This is a secondary analysis, using existing data from schizophrenia case-control consortia, and publicly available transcriptome data. We did not collect samples at any stage.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	PrediXcan and MetaXcan Python code, publicly available through github: https://github.com/hakyimlab/PrediXcan GTEX models version 6 were used. Some custom code was used to make PrediXcan compliant with PGC data formats. Code will be made available on request.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Summary statistics are publcily available. Genotypes are available to approved researchers through collaboration with the PGC and CLOZUK2 groups

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	NA
c. Report whether the cell lines were tested for mycoplasma contamination.	NA
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	NA

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No research animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	We obtained 53 discovery cohorts for this study, including 40,299 SCZ cases and 65,264 controls (Figure 2). 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 GWAS22. 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. 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. Specifically, please see: https://www.nature.com/article-assets/npg/nature/journal/v511/n7510/extref/nature13595-s1.pdf