bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Gene expression imputation across multiple brain regions reveals schizophrenia risk**
2 **throughout development.**
3
4 **Author List:**
5 Laura M. Huckins[1,2], Amanda Dobbyn[1,2], Douglas M. Ruderfer[3], Gabriel Hoffman[1,4], Weiqing
6 Wang[1,2], Antonio Pardinas[5] Veera M Rajagopal[6,15,16], Thomas D. Als[6,15,16], Hoang Nguyen[1,2],
7 Kiran Girdhar[1,2], James Boocock[7], Panos Roussos[1,2], Menachem Fromer[1,2], Robin Kramer[8],
8 Enrico Domencini[9], Eric Gamazon[3,14], Shaun Purcell[1,2,4], CommonMind Consortium, the
9 Schizophrenia Working Group of the Psychiatric Genomics Consortium, iPSYCH-GEMS
10 Schizophrenia Working Group, Ditte Demontis[6,15,16], Anders D. Børglum[6,15,16], James Walters[5],
11 Michael O'Donovan[5], Patrick Sullivan[10,11], Michael Owen[5], Bernie Devlin[12], Solveig K.
12 Sieberts[13], Nancy Cox[3,14], Hae Kyung Im[14], Pamela Sklar[1,2,4], Eli A. Stahl[1,2]
13
14

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

15 **Author Affiliations:**

16 1. Division of Psychiatric Genomic, Icahn School of Medicine at Mount Sinai, NYC, NY;
17 2. Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, NYC, NY;
18 3. Vanderbilt University Medical Center, Nashville, TN;
19 4. Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount
20 Sinai, NYC, NY, USA;
21 5. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK;
22 6. Department of Biomedicine, Aarhus University, Aarhus, Denmark
23 7. Department of Human Genetics, David Geffen School of Medicine, University of California
24 Los Angeles, Los Angeles, CA, USA;
25 8. Human Brain Collection Core, National Institute of Mental Health, Bethesda, MD, USA;
26 9. Laboratory of Neurogenomic Biomarkers, Centre for Integrative Biology (CIBIO),
27 University of Trento, Trento, Italy;
28 10. University of North Carolina at Chapel Hill, NC, USA
29 11. Karolinska Institutet, Stockholm, Sweden
30 12. Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA;
31 13. Systems Biology, Sage Bionetworks, Seattle, WA, USA;
32 14. Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago,
33 Illinois, USA
34 15. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH,
35 Denmark.
36 16. Center for Integrative Sequencing, Aarhus University, Aarhus, Denmark.
37
38

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39 **Abstract**

40 Transcriptomic imputation approaches offer an opportunity to test associations between disease
41 and gene expression in otherwise inaccessible tissues, such as brain, by combining eQTL
42 reference panels with large-scale genotype data. These genic associations could elucidate signals
43 in complex GWAS loci and may disentangle the role of different tissues in disease development.
44 Here, we use the largest eQTL reference panel for the dorso-lateral pre-frontal cortex (DLPFC),
45 collected by the CommonMind Consortium, to create a set of gene expression predictors and
46 demonstrate their utility. We applied these predictors to 40,299 schizophrenia cases and 65,264
47 matched controls, constituting the largest transcriptomic imputation study of schizophrenia to
48 date. We also computed predicted gene expression levels for 12 additional brain regions, using
49 publicly available predictor models from GTEx. We identified 413 genic associations across 13
50 brain regions. Stepwise conditioning across the genes and tissues identified 71 associated genes
51 (67 outside the MHC), with the majority of associations found in the DLPFC, and of which
52 14/67 genes did not fall within previously genome-wide significant loci. We identified 36
53 significantly enriched pathways, including hexosaminidase-A deficiency, and multiple pathways
54 associated with porphyric disorders. We investigated developmental expression patterns for all
55 67 non-MHC associated genes using BRAINSPAN, and identified groups of genes expressed
56 specifically pre-natally or post-natally.

57

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58 **Introduction**

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

72

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

87

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

88 The access to the large collection of dorso-lateral pre-frontal cortex (DLPFC) gene expression
89 data collected by the CommonMind Consortium[14] affords us a unique opportunity to study and
90 codify relationships between genotype and gene expression. Here, we present a novel set of gene
91 expression predictor models, built using CommonMind Consortium DLPFC data[14]. We compare
92 different regression approaches to building these models (including elastic net[15], Bayesian sparse
93 linear mixed models and ridge regression[16], and using max eQTLs), and benchmark performance
94 of these predictors against existing GTEx prediction models. We applied our CMC DLPFC
95 predictors and 12 GTEx-derived neurological prediction models to predict gene expression in
96 schizophrenia GWAS data, obtained through collaboration with the Psychiatric Genomics
97 Consortium (PGC) schizophrenia working group, the "CLOZUK2" cohort, and the iPSYCH-
98 GEMS schizophrenia working group. We identified 413 genome-wide significant genic
99 associations with schizophrenia in our PGC+CLOZUK2 sample, constituting 67 independent
100 associations outside the MHC region. We demonstrate the relevance of these associations to
101 schizophrenia aetiopathology using gene set enrichment analysis, and by examining the effects
102 of manipulation of these genes in mouse models. Finally, we investigated spatio-temporal
103 expression of these genes using a developmental transcriptome dataset, and identified distinct
104 spatio-temporal patterns of expression across our associated genes.
105
106
107
108
109
110
111
112

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

113 **Results**

114 **Prediction Models based on CommonMind Consortium DLPFC expression**

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

126

127 To test the predictive accuracy of the CMC-derived DLPFC models, and to benchmark this
128 against existing GTEx-derived prediction models, genetically-regulated gene expression (GREX)
129 was calculated in an independent DLPFC RNA-sequencing dataset (the Religious Orders Study
130 Memory and Ageing Project, ROSMAP[19]). We compared predicted GREX to measured
131 ROSMAP gene expression for each gene (Replication $R^{2,}$ or R_R^2) for the CMC-derived DLPFC
132 models and twelve GTEx-derived brain tissue models[15,20,21] (Figure 1, Supplementary Figure
133 2B). CMC-derived DLPFC models had higher average R_R^2 values (Mean $R_R^2 = 0.056$), more
134 genes with $R_R^2 > 0.01$, and significantly higher overall distributions of R_R^2 values than any of the
135 twelve GTEx models (ks-test, $p<2.2\times10^{-16}$ across all analyses; Figure 1). Median R_R^2 values were
136 significantly correlated with sample size of the original tissue set (rho=0.92, $p=7.2\times10^{-6}$), the
137 number of genes in the prediction model (rho=0.9, $p=2.6\times10^{-5}$), and the number of significant
138 'eGenes' in each tissue type (rho=0.95, $p=5.^{5x10-7}$; Figure 1C). Notably, these correlations persist
139 after removing obvious outliers (Figure 1C).

140

141 To estimate trans-ancestral prediction accuracy, genetically regulated gene expression was
142 calculated for 162 African-American individuals and 280 European individuals from the NIMH
143 Human Brain Collection Core (HBCC) dataset (supplementary figure 2B). R_R^2 values were

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

144 higher on average in Europeans than African-Americans (average $R_{R_EUR}^2$ = 0.048, $R_{R_AA}^2$ =
145 0.040), but were significantly correlated between African-Americans and Europeans (rho=0.78,
146 p<2.2 x10^{-16}, Pearson test; supplementary figure 3).
147
148 **Application of Transcriptomic Imputation to Schizophrenia**
149 We used CMC DLPFC and the 12 GTEx–derived brain tissue prediction models to impute
150 genetically regulated expression levels (GREX) of 19,661 unique genes in cases and controls
151 from the PGC-SCZ GWAS study[22]. Predicted expression levels were tested for association with
152 schizophrenia. Additionally, we applied CMC and GTEx-derived prediction models to summary
153 statistics from 11 PGC cohorts (for which raw genotypes were unavailable) and the CLOZUK2
154 cohort. Meta-analysis was carried out across all PGC-SCZ and CLOZUK2 cohorts using an
155 odds-ratio based approach in METAL. Our final analysis included 40,299 cases and 65,264
156 controls (Figure 2A).
157
158 We identified 413 genome-wide significant associations, representing 256 genes in 13 tissues
159 (Figure 3A). The largest number of associations were detected in the CMC DLPFC GREX data
160 (Figure 3C; 49 genes outside the MHC, 69 genes overall). We sought replication of our CMC
161 DLPFC SCZ-associations in an independent dataset of 4,133 cases and 24,788 controls in
162 collaboration with the iPSYCH-GEMS SCZ working group (Figure 2B). We found significant
163 correlation of effect sizes (p=1.784 x10^{-04}; rho=0.036) and –log10 p-values (p=1.073 x10^{-05};
164 rho=0.043) between our discovery (PGC+CLOZUK2) and replication (iPSYCH-GEMS)
165 samples. Non-MHC Genes reaching genome-wide significance in our discovery sample (49
166 genes) were significantly more likely to reach nominal significance in the replication sample, and
167 had significantly more consistent directions of effect than might be expected by chance
168 (binomial test, p=2.42 x10^{-05}, p=0.044). (Suppl. info).
169
170 To identify the top independent associations within genomic regions, which include multiple
171 associations for a single gene across tissues, or multiple nearby genes, we partitioned genic
172 associations into 58 groups defined based on genomic proximity and applied stepwise forward
173 conditional analysis within each group (Supplementary Table 1). In total, 67 genes remained
174 genome-wide significant after conditioning (Table 1; Figure 3A-B). The largest signal was

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

175 identified in the CMC DLPFC predicted expression data (24 genes; Figure 3C), followed by the
176 Putamen (7 genes). 19/67 genes did not lie within 1Mb of a previously genome-wide significant
177 GWAS locus[22] (shown in bold, Table 1); of these, 5/19 genes were within 1Mb of a locus which
178 approached genome-wide significance ($p<5x10^{-07}$). The remaining 14 genes all fall within
179 nominally significant PGC-SCZ GWAS loci ($p<8x10^{-04}$), but did not reach genome-wide
180 significance.
181

182 **Implicated genes highlight SCZ-associated molecular pathways and gene set analyses**

183 We tested for overlap between our non-MHC SCZ-associated genes and 8,657 genesets
184 comprised of 1) hypothesis-driven pathways and 2) general molecular database pathways. We
185 corrected for multiple testing using the Benjamin-Hochberg false discovery rate (FDR)
186 correction[23].
187

188 We identified three significantly associated pathways in our hypothesis-driven analysis (Table
189 2). Targets of the fragile-X mental retardation protein formed the most enriched pathway
190 (FMRP; $p=1.96x10^{-8}$). Loss of FMRP inhibits synaptic function, is comorbid with autism
191 spectrum disorder, and causes intellectual disability, as well as psychiatric symptoms including
192 anxiety, hyperactivity and social deficits[24]. Enrichment of this large group of genes has been
193 observed frequently, in the original CommonMind analysis[14], by colleagues investigating the
194 same PGC and CLOZUK2 samples[26] as well as by investigators studying autism[24,27]. There was
195 a significant enrichment among our SCZ associated genes and genes that have been shown to be
196 intolerant to loss-of-function mutations[28] ($p=5.86x10^{-5}$) as well as with CNVs associated with
197 bipolar disorder[29] ($p=7.92x10^{-8}$), in line with a recent variant-based study of the same
198 individuals[26].
199

200 Next, we performed an agnostic search for overlap between our schizophrenia-associated genes
201 and ~ 8,500 molecular pathways collated from large, publicly available databases. 33 pathways
202 were significantly enriched after FDR correction (Table 2, Suppl. Table 2), including a number
203 of pathways with some prior literature in psychiatric disease. We identified an enrichment with
204 porphyrin metabolism ($p=1.03x10^{-4}$). Deficiencies in porphyrin metabolism lead to “Porphyria”,
205 an adult-onset metabolic disorder with a host of associated psychiatric symptoms, in particular

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

206 episodes of violence and psychosis[30–35]. Five pathways potentially related to porphyrin
207 metabolism, regarding abnormal iron level in the spleen, liver and kidney are also significantly
208 enriched, including 2/5 of the most highly enriched pathways ($p<2.0 \times 10^{-04}$). The PANTHER
209 and REACTOME pathways for Heme biosynthesis and the GO pathway for protoporphyrinogen
210 IX metabolic process, which are implicated in the development of porphyric disorders, are also
211 highly enriched ($p=2.2 \times 10^{-04}$, 2.6×10^{-04}, 4.1×10^{-04}), although do not pass FDR-correction.
212
213 Hexosaminidase activity was enriched ($p=3.47 \times 10^{-05}$) in our results; this enrichment is not
214 driven by a single highly-associated gene; rather, every single gene in the HEX-A pathway is
215 nominally significant in the SCZ association analysis (Supplementary Table 2). Deficiency of
216 hexosaminidase A (HEX-A) results in serious neurological and mental problems, most
217 commonly presenting in infants as "Tay-Sachs" disease[36]. Adult-onset HEX-A deficiency
218 presents with neurological and psychiatric symptoms, notably including onset of psychosis and
219 schizophrenia[37]. Five pathways corresponding to Ras- and Rab- signaling, protein regulation and
220 GTPase activity were enriched ($p<6 \times 10^{-05}$). These pathways have a crucial role in neuron cell
221 differentiation[38] and migration[39], and have been implicated in the development of schizophrenia
222 and autism[40–43]. We also find significant enrichment with protein phosphatase type 2A regulator
223 activity ($p=5.24 \times 10^{-05}$), which was associated with MDD and across MDD, BPD and SCZ in the
224 same large integrative analysis[44], and has been implicated in antidepressant response and
225 serotonergic neurotransmission[45].
226

227 **Predicted gene expression changes are consistent with functional validation studies**

228 To test the functional impact of our SCZ-associated predicted gene expression changes (GREX),
229 we performed two in-silico analyses. First, we compared directions of effect in our meta-analysis
230 to those in the CMC analysis of differentially expressed genes between SCZ cases and controls.
231 This analysis highlighted six loci where expression levels of a single gene putatively affected
232 schizophrenia risk. All six of these genes are nominally significant in our DLPFC analysis, and
233 two (*CLCN3* and *FURIN*) reach genome-wide significance. In the conditional analysis across all
234 brain regions, one additional gene (*SNX19*) reaches genome-wide significance. The direction of
235 effect for all six genes matches the direction of gene expression changes observed in the original
236 CMC paper, indicating that gene expression estimated in the imputed transcriptome reflects

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

237 measured expression levels in brains of individuals with Schizophrenia. Further, this observation
238 is consistent with a model where the differential expression signature observed in CMC is caused
239 by genetics rather than environment.
240
241 The original CMC analysis identified 21 eSNP genes using SHERLOCK[14,46], of which 17 were
242 present in our CMC DLPFC analysis. 14/17 genes reached nominal significance (significantly
243 more than expected by chance, $p=3.6\times10^{-16}$), and 11 reached genome-wide significance
244 (binomial p-value 6.04×10^{-55}). Additionally, 31 regions contained genes ranked highly by
245 Sherlock in the original CMC analysis (supplementary data file 2 in Fromer, M. *et al.* Gene
246 expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* 19,
247 1442–1453 (2016)[14]). Of these, 14 regions lay near one of our CMC DLPFC associated genes,
248 and 13/14 regions had common genes between SHERLOCK and PrediXcan analyses. Five loci
249 included multiple SHERLOCK genes; in every instance we are able to specifically identify one
250 or two associated genes from the longer SHERLOCK list.
251
252 To understand the impact of altered expression of our 67 SCZ-associated genes, we performed
253 an in-silico analysis of mouse mutants, by collating large, publicly available mouse databases[47–
254 51]. We identified mutant mouse lines lacking expression of 37/67 of our SCZ-associated genes,
255 and obtained 5,333 phenotypic data points relating to these lines, including 1,170 related to
256 behavioral, neurological or craniofacial phenotypes. 25/37 genes were associated with at least
257 one behavioral, neurological or related phenotype (Supplementary table 3). We repeated this
258 analysis for genes identified in 366 GWAS, including any GWAS for which at least ten mutant
259 mouse lines exist (105 GWAS). SCZ-associated genes were more likely to be associated with
260 behavior, brain development and nervous system phenotypes than genes in these GWAS sets
261 (p=0.057).
262
263 **Spatiotemporal expression of SCZ-associated genes indicated distinct patterns of risk**
264 **throughout development**
265 We assessed expression of our SCZ-associated genes throughout development using
266 BRAINSPAN[52]. Data were partitioned into eight developmental stages (four pre-natal, four post-
267 natal), and four brain regions[29,52] (Figure 4A). We noted that SCZ-associated genes were

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

268 significantly co-expressed, in both pre-natal and post-natal development and in all four brain
269 regions, based on local connectedness[53] (Figure 4B), global connectedness[53] (i.e., average path
270 length between genes, supplementary Figure 6), and network density (i.e., number of edges,
271 supplementary Figure 7). Examining pairwise gene expression correlation (suppl. Fig 8) and
272 gene co-expression networks (suppl. Fig 9) for each spatiotemporal point indicated that the same
273 genes do not drive this co-expression pattern throughout development; rather, it appears that
274 separate groups of genes drive early pre-natal, late pre-natal and post-natal clustering.
275
276 To visualize this, we calculated Z scores of gene expression for each SCZ-associated gene,
277 across all 32 time-points (Figure 5). Genes clustered into four groups (supplementary fig 10),
278 with distinct spatio-temporal expression signatures. The largest cluster (Cluster A, Figure 5A; 29
279 genes) spanned early to late-mid pre-natal development (4-24 weeks post conception), either
280 across the whole brain (22 genes) or in regions 1-3 only (7 genes). 12 genes were expressed in
281 late pre-natal development (Figure 5D; 25-38 pcw); 10 genes were expressed in regions 1-3,
282 post-natally and in the late pre-natal period (Figure 5C), and 15 genes were expressed throughout
283 development (Figure 5B), either specifically in region four (nine genes) or throughout the brain
284 (six genes). We used a stratified qq-plot approach[54] to examine whether SNPs in cis-regions of
285 genes in these four clusters are differentially enriched in psychiatric disorders. SNPs in cis-
286 regions of genes in the two pre-natal clusters are more highly enriched than SNPs in cis-regions
287 of genes in post-natal clusters, and compared to all SNPs, in childhood-onset disorders (ASD and
288 ADHD, supplementary figure 13), but not adult-onset disorders (BPD and MDD, data not
289 shown).
290
291 We noticed a relationship between patterns of gene expression and the likelihood of behavioral,
292 neurological or related phenotypes in our mutant mouse model database. Mutant mice lacking
293 genes expressed exclusively pre-natally in humans, or genes expressed pre- and post-natally,
294 were more likely to have any behavioral or neurological phenotypes than mutant mice lacking
295 expression of genes expressed primarily in the third trimester or post-natally ($p=1.7\times10^{-04}$)
296 (supplementary figure 11).
297
298

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

299 **Discussion**

300 In this study, we present gene expression prediction models for the dorso-lateral pre-frontal
301 cortex (DLPFC), constructed using CommonMind Consortium genotype and gene expression
302 data. These prediction models may be applied to either raw data or summary statistics, in order to
303 yield gene expression information in large data sets, and across a range of tissues. This has the
304 significant advantage of allowing researchers to access transcriptome data for non-peripheral
305 tissues, at scales currently prohibited by the high cost of RNA sequencing, and circumventing
306 distortions in measures of gene expression stemming from errors of measurement or
307 environmental influences. Since disease status may alter gene expression but not the germline
308 profile, analyzing genetically regulated expression ensures that we identify only the causal
309 direction of effect between gene expression and disease[15]. Large, imputed transcriptomic datasets
310 represent the first opportunity to study the role of subtle gene expression changes (and therefore
311 modest effect sizes) in disease development.

312

313 There are some inherent limitations to this approach. The accuracy of transcriptomic imputation
314 (TI) is reliant on access to large eQTL reference panels, and it is therefore vital that efforts to
315 collect and analyze these samples continue. TI has exciting advantages for gene discovery as
316 well as downstream applications[15,55,56]; however, the relative merits of existing methodologies
317 are as yet under-explored. Our analysis suggests that, overall, sparser elastic net models better
318 capture gene expression regulation than BSLMM; at the same time, the improved performance of
319 elastic net over max-eQTL models suggests that a single eQTL model is over-simplified[2,15].
320 Fundamentally, transcriptomic imputation methods model only the genetically regulated portion
321 of gene expression, and so cannot capture or interpret variance of expression induced by
322 environment or lifestyle factors, which may be of particular importance in psychiatric disorders.
323 Given the right study design, analyzing genetic components of expression together with observed
324 expression could open doors to better study the role of gene expression in disease.

325

326 Sample size and tissue matching contribute to accuracy of TI results. Our CMC-derived DLPFC
327 prediction models had higher average validation R^2 values in external DLPFC data than GTEx-
328 derived brain tissue models. Notably, the model with the second highest percent of genes passing
329 the R^2 threshold is the Thyroid, which has the largest sample size among the GTEx brain

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

330 prediction models. When looking at mean R^2 values, the second highest value comes from the
331 GTEx Frontal Cortex, despite the associated small sample size, implying at least some degree of
332 tissue specificity of eQTLs architecture.
333
334 We were able to compare TI accuracy in European and African-American individuals, and found
335 that our models were applicable to either ethnicity with only a small decrease in accuracy.
336 Common SNPs shared across ethnicities have important effects on gene expression, and as such
337 we expect GREX to have consistency across populations. There is a well-documented dearth of
338 exploration of genetic associations in non-European cohorts[57,58] We believe that these analyses
339 should be carried out in non-European cohorts.
340
341 We applied the CMC DLPFC prediction models, along with 12 GTEx-derived brain expression
342 prediction models, to schizophrenia cases and controls from the PGC2 and CLOZUK2
343 collections, constituting the largest transcriptomic analysis of schizophrenia to date. Predicted
344 gene expression levels were calculated for 19,661 unique genes across brain regions (Figure 1C)
345 and tested for association with SCZ case-control status. We identified 413 significant
346 associations, constituting 67 independent associations. We found significant replication of our
347 CMC DLPFC associations in a large independent replication cohort, in collaboration with the
348 iPSYCH-GEMS consortium. A recent TWAS study of 30 GWAS summary statistic traits[55]
349 identified 38 non-MHC genes associated at tissue-level significance with SCZ in CMC- and
350 GTEx-derived brain tissues (ie, matching those used in our study). Of these, 26 also reach
351 genome-wide significance in our study, although in many instances these genes are not identified
352 as the lead independent associated gene following our conditional analysis. Among our 67 SCZ-
353 associated genes, 19 were novel, i.e. did not fall within 1Mb of a previous GWAS locus
354 (including 5/7 of the novel brain genes identified in the recent TWAS analysis).
355
356 We used conditional analyses to identify independent associations within loci. These analyses
357 clarify the most strongly associated genes and tissues (Table 1), while we note that nearly
358 collinear gene-tissue pairs could also represent causal associations. The tissues highlighted
359 allowed us to tabulate apparently independent contributions to SCZ risk from different brain

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

360 regions, even though their transcriptomes are highly correlated generally. We find DLPFC and
361 Cerebellum effects, as well as from Putamen, Caudate and Nucleus Accumbens Basal Ganglia.
362
363 We used these genic associations to search for enrichments with molecular pathways and gene
364 sets, and identified 36 significant enriched pathways. Among novel pathways, we identified a
365 significant association with HEX-A deficiency. Despite the well-studied and documented
366 symptomatic overlap between adult-onset HEX-A deficiency and schizophrenia, we believe that
367 this is the first demonstration of shared genetics between the disorders. Notably, this overlap is
368 not driven by a single highly-associated gene which is shared by both disorders; rather, every
369 single gene in the HEX-A pathway is nominally significant in the SCZ association analysis, and
370 five genes have $p < 1x10^{-03}$, indicating that there may be substantial shared genetic aetiology
371 between the two disorders that warrants further investigation. Additionally, we identified a
372 significant overlap between our SCZ-associated genes and a number of pathways associated with
373 porphyrin metabolism. Porphyric disorders have been well characterized and are among early
374 descriptions of "schizophrenic" and psychotic presentations of schizophrenia, as described in the
375 likely eponymous mid-19th century poem "Porphyria's Lover", by Robert Browning[59], and have
376 been cited as a likely diagnosis for the various psychiatric and metabolic ailments of Vincent van
377 Gogh[60–65] and King George III[66].
378
379 Finally, we assessed patterns of expression for the 67 SCZ-associated genes throughout
380 development using spatio-temporal transcriptomic data obtained from BRAINSPAN. We
381 identified four clusters of genes, with expression in four distinct spatiotemporal regions, ranging
382 from early pre-natal to strictly post-natal expression. There are plausible hypotheses and genetic
383 evidence for SCZ disease development in adolescence, given the correlation with age of onset, as
384 well as prenatally, supported by genetic overlap with neurodevelopmental disorders[67–69] as well
385 as the earlier onset of cognitive impairments[70–73]. Understanding the temporal expression
386 patterns of SCZ-associated genes can help to elucidate gene development and trajectory, and
387 inform research and analysis design. Identification of SCZ-associated genes primarily expressed
388 prenatally is striking given our adult eQTL reference panels, and may reflect common eQTL
389 architecture across development, which is known to be partial[74–76]; therefore, our results should
390 spur interest in extending TI data and/or methods to early development[74]. Identification of SCZ-

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

391 associated genes primarily expressed in adolescence and adult-hood is of particular interest for
392 direct analysis of the brain transcriptome in adult psychiatric cases.
393
394 eQTL data have been recognized for nearly a decade as potentially important for understanding
395 complex genetic variation. Nicolae et al[1] showed that common variant-common disease
396 associations are strongly enriched for genetic regulation of gene expression. Therefore,
397 integrative approaches combining transcriptomic and genetic association data have great
398 potential. Current TI association analyses increase power for genetic discovery, even while many
399 open areas of TI remain to be developed, such as leveraging additional data types such as
400 chromatin modifications[77] (e.g. methylation, histone modification), imputing different tissues or
401 different exposures (e.g. age, smoking, trauma) and modeling trans/coexpression effects. It
402 remains critical to leverage TI associations to provide insights into specific disease mechanisms.
403 Here, the accelerated identification of disease associated genes allows the detection of novel
404 pathways and distinct spatiotemporal patterns of expression in schizophrenia risk.
405
406

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

408

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

410 **eQTL Data**

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

418

419 A 1% MAF cut-off was applied. Variants were filtered to remove any SNPs in high LD (r^2>0.9),
420 indels, and all variants with ambiguous ref/alt alleles. All protein coding genes on chromosomes
421 1-22 with at least one cis-SNP after these QC steps were included in this analysis. SNPs in trans
422 have been shown not to provide a substantial improvement in prediction accuracy[15] and were not
423 included here.

424

425 **Building gene expression prediction databases**

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

431

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

436

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

437 Prediction models were compared across four different regression methods; elastic net
438 (prediXcan), ridge regression (using the TWAS method[16]), Bayesian sparse linear mixed
439 modelling (BSLMM; TWAS), and linear regression using the best eQTL for each gene
440 (Supplementary Figure 1A). Mean R_{cv}^2 values were significantly higher for elastic net regression
441 (mean R_{cv}^2=0.056) than for eQTL-based prediction (mean R_{cv}^2=0.025), BSLMM (mean
442 R_{cv}^2=0.021) or Ridge Regression (mean R_{cv}^2=0.020). The distribution of R_{cv}^2 values was also
443 significantly higher for elastic net regression than for any other method (ks-test, $p<2.2\times10^{-16}$).
444

445 **Replication of gene expression prediction models in independent data**

446 Predictive accuracy of CMC DLPFC models were tested in two independent datasets.
447 First, we used data from the Religious Orders Study and Memory and Aging Project
448 (ROSMAP[19]). This study included genotype data and DLPFC RNA-seq data[78] for 451
449 individuals of European descent (Supplementary Figure 2B).
450

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

455 *Equation 1: R_R^2 calculation.*

456 $$R_{R1}^2 = (M \sim GREX + PC_1 + PC_2 + \cdots + PC_{10})$$
457 $$R_{R2}^2 = (M \sim PC_1 + PC_2 + \cdots + PC_{10})$$
458 $$R_R^2 = R_{R1}^2 - R_{R2}^2$$
459

460 Where:

M	Measured expression (RNA-seq)
$GREX$	GREX imputed expression
PC_n	n^{th} Principal Component

461

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

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

465
466 Prediction accuracy was also assessed for 11 publicly available GTEx neurological predictor
467 databases, and R_R^2 values used to compare to CMC DLPFC performance. CMC DLPFC models
468 had higher average R_R^2 values, more genes with $R_R^2 > 0.01$, and significantly higher overall
469 distributions of R_R^2values than any of the twelve GTEx brain tissue models (ks-test, p<2.2e-16;
470 Figure 1A,B).
471
472 To estimate trans-ancestral prediction accuracy, genetically regulated gene expression was
473 calculated for 162 African-American individuals and 280 European individuals from the NIMH
474 Human Brain Collection Core (HBCC) dataset[79] (Supplementary Figure 2C). Predicted gene
475 expression levels were compared to DLPFC expression levels measured using microarray. There
476 was a significant correlation between the European and African-American samples for R_{CV}^2
477 values and R_R^2 values (rho=0.66, 0.56; Supplementary figure 3B-C). R_R^2 values were higher on
478 average in Europeans, but were significantly correlated between African-Americans and
479 Europeans (rho=0.78, p<2.2e-16, Pearson test; supplementary figure 3D).
480

481 **Extension to Summary Statistics**

482 Transcriptomic Imputation may be applied to summary statistics instead of raw dosages, in
483 instances where raw data is unavailable. However, this method suffers from slightly reduced
484 accuracy, requires covariance matrices calculated in an ancestrally-matched reference
485 population[80] (usually only possible for European cohorts), and precludes testing of
486 endophenotypes within the data, and so should not be applied when raw data is available.
487
488 We assessed concordance between CMC DLPFC transcriptomic imputation results using
489 summary-statistics (MetaXcan[80]) and raw genotypes (PrediXcan[15]) using nine European and
490 three Asian PGC-SCZ cohorts[22] for which both data types were available. Cohorts were chosen
491 to encompass a range of case : control ratios, to test previous suggestions that accuracy is
492 reduced in unbalanced cohorts[80]. Covariances for all variants included in the DLPFC predictor
493 models were computed using MetaXcan[80]. For all European cohorts, Pearson correlation of log-
494 10 p-values and effect sizes was above 0.95. The mean correlation was 0.963 (Supplementary
495 Figure 4). There was no correlation between total sample size, case-control ratio, p-value or

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

496 effect-size. Seven genes were removed due to discordant p-values. For the three Asian cohorts
497 tested, the mean correlation was 0.91 (Supplementary Figure 5).
498
499 Concordance was also tested for the same nine European PGC-SCZ cohorts, across 12
500 neurological GTEx prediction databases. All correlations were significant (rho>0.95, p<2.2e-16).
501 There was a significant correlation between p-value concordance and case-control ratio
502 (rho=0.37, p=7.606 x10^{-15}). 114 genes had discordant p-values between the two methods and
503 were excluded from future analyses.
504

505 **Application to Schizophrenia**

506 **Dataset Collection**

507 We obtained 53 discovery cohorts for this study, including 40,299 SCZ cases and 65,264
508 controls (Figure 2). 52/53 cohorts (35,079 cases, 46,441 controls) were obtained through
509 collaboration with the Psychiatric Genomics Consortium, and are described in the 2014 PGC
510 Schizophrenia GWAS[22]. The remaining cohort, referred to as CLOZUK2, constitutes the largest
511 single cohort of individuals with Schizophrenia (5,220 cases and 18,823 controls), collected as
512 part of an effort to investigate treatment-resistant Schizophrenia[26].
513
514 50/53 datasets included individuals of European ancestry, while three datasets include
515 individuals of Asian ancestry (1,836 cases, 3,383 controls). All individuals were ancestrally
516 matched to controls. Information on genotyping, quality control and other data management
517 issues may be found in the original papers describing these collections[22,26]. All sample
518 collections complied with ethical regulations. Details regarding ethical compliance and consent
519 procedures may be found in the original manuscripts describing these collections[22,26].
520
521 Access to dosage data was available for 44/52 PGC-SCZ cohorts. The remaining PGC cohorts,
522 and the CLOZUK2 cohort provided summary statistics. Three European PGC cohorts were trio-
523 based, rather than case-control.
524
525 Additionally, we tested for replication of our CMC DLPFC associations in an independent
526 dataset of 4,133 cases and 24,788 controls obtained through collaboration with the iPSYCH-

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

527 GEMS schizophrenia working group (effective sample size 14,169.5; Figure 2B, supplementary
528 information).
529

530 **Transcriptomic Imputation and association testing**

531 Transcriptomic Imputation was carried out individually for each case-control PGC-SCZ cohort
532 with available dosage data (44/52 cohorts). Predicted gene expression levels were computed
533 using the DLPFC predictors described in this manuscript, as well as for 11 other brain tissues
534 prediction databases created using GTEx tissues[15,20,21,81] (Figure 1C). Associations between
535 predicted gene expression values and case-control status were calculated using a linear
536 regression test in R. Ten ancestry principal components were included as covariates. Association
537 tests were carried out independently for each cohort, across 12 brain tissues.
538

539 For the 8 PGC cohorts with no available dosage data, the three PGC trio-based analyses, and the
540 CLOZUK2 cohort, a summary-statistic based transcriptomic imputation approach was used
541 ("MetaXcan"), as described previously.
542

543 **Meta-analysis**

544 Meta-analysis was carried out across all 53 cohorts using METAL[82]. Cochran's Q test for
545 heterogeneity was implemented in METAL[82,83], and a heterogeneity p-value threshold of $p >$
546 $1x10^{-3}$ applied to results. A conservative significance threshold was applied to these data,
547 correcting for the total number of genes tested across all tissues (121,611 gene-region tests in
548 total). This resulted in a genome-wide significance threshold of $4.1x10^{-7}$.
549

550 Effect sizes and direction of effect quoted in this manuscript refer to changes in predicted
551 expression in cases compared to controls i.e., genes with negative effect sizes have decreased
552 predicted expression in cases compared to controls.
553

554 **Identifying independent associations**

555 We identified a number of genomic regions which contained multiple gene associations and/or
556 genes associated across multiple tissues. We identified 58 of these regions, excluding the MHC,
557 based on distance between associated genes, and verified them using visual inspection. In order

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

558 to identify independent genic associations within these regions, we carried out a stepwise
559 forward conditional analysis following "GCTA-COJO" theory[84] using "CoCo"
560 (https://github.com/theboocock/coco/), an R implementation of GCTA-COJO. CoCo allows the
561 specification of custom correlation matrices by the user (for example, ancestrally specific LD
562 matrices). For each region, we generated a predicted gene expression correlation matrix for all
563 significant genes ($p \leq 1x10^{-6}$), as the root-effective sample size (N_{eff}, eqn 2) weighted average
564 correlation across all cohorts where we had access to dosage data.

565 *Equation 2: Effective Sample Size, N_{eff}*

566
$$N_{eff} = \frac{4}{(\frac{1}{N_{cases}} + \frac{1}{N_{controls}})}$$

567

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

573

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

576 *Equation 3: GREX Beta adjustment*

577
$$\beta = \beta_{CoCo}\, x \sqrt{GVAR}$$

578

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

580

581 **Gene set Analyses**

582 Pathway analyses were carried out using an extension to MAGMA[85]. P-values were assigned to
583 genes using the most significant p-value achieved by each gene in the meta-analysis. We then
584 carried out a competitive gene-set analysis test using these p-values, using two gene sets:

585

586 1. 159 gene sets with prior hypotheses for involvement in SCZ development, including loss-
587 of-function intolerant genes, CNV-intolerant genes, targets of the fragile-X mental

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

588 retardation protein, CNS related gene sets, and 104 behavioural and neurological
589 pathways from the Mouse Genome Informatics database[14,26,67,86].
590 2. An agnostic analysis, including ~8,500 gene sets collated from publicly available
591 databases including GO[87,88], KEGG[89], REACTOME[90], PANTHER[91,92], BIOCARTA[93]
592 and MGI[48]. Sets were filtered to include only gene sets with at least ten genes.
593
594 Significance levels were adjusted across all pathways included in either test using the
595 Benjamini-Hochberg "FDR" correction in R[23].
596
597 **Coexpression of SCZ genes throughout development**
598 We investigate spatiotemporal expression of our associated genes using publicly available
599 developmental transcriptome data, obtained from the BRAINSPAN consortium[94]. We partitioned
600 these data into biologically relevant spatio-temporal data sets[95], corresponding to four general
601 brain regions; the frontal cortex, temporal and parietal regions, sensory-motor regions, and
602 subcortical regions (Figure 4A[96]), and eight developmental time-points (four pre-natal, four post-
603 natal)[95].
604
605 First, we tested for correlation of gene expression for all SCZ-associated genes at each
606 spatiotemporal time-point. Genes with pearson correlation coefficients >= 0.8 or <=-0.8 were
607 considered co-expressed. 100,000 iterations of this analysis were carried out using random gene
608 sets with equivalent expression level distributions to the SCZ-associated genes. For each gene
609 set, a gene co-expression network was created, with edges connecting all co-expressed genes.
610 Networks were assessed using three criteria; first, the number of edges within the network, as a
611 crude measured of connectedness; second, the Watts-Strogatz average path length between
612 nodes, as a global measure of connectedness across all genes in the network[53]; third, the Watts-
613 Strogatz clustering coefficient, to measure tightness of the clusters within the network[53]. For
614 each spatio-temporal time point, we plotted gene-pair expression correlation (suppl. Fig 8) and
615 co-expression networks (suppl. Fig 9).
616
617 For each of the 67 SCZ-associated genes, we calculated average expression at each
618 spatiotemporal point. We then calculated Z-Score of expression specificity using these values,

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

619 and plotted Z-Scores to visually examine patterns of gene expression throughout development
620 and across brain regions. Clusters were formally identified using a dendrogram cut at height 10
621 (Suppl. Fig 10).
622

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

624 We downloaded genotype, knock-out allele information and phenotyping data for ~10,000
625 mouse mutant models from five large mouse phenotyping and genotyping projects; Mouse
626 Genome Informatics (MGI[48]), EuroPhenome[47,97], Mouse Genome Project (MGP[47,49]),
627 International Mouse Phenotyping Consortium (IMPC[50]), and Infection and Immunity
628 Immunophenotyping (3I[98]). Where possible, we also downloaded raw phenotyping data
629 regarding specific assays. In total, we obtained 175,012 phenotypic measurements, across 10,288
630 mutant mouse models. We searched for any mouse lines with phenotypes related to behavior
631 (natural, observed, stereotypic or assay-induced); cognition or working memory; brain, head or
632 craniofacial dysmorphology; retinal or eye morphology, and/or vision or visual dysfunction or
633 impairment; ear morphology or hearing dysfunction or impairment; neural tube defects; brain
634 and/or nervous system development; abnormal nociception.
635

636 We compared the prevalence of psychiatric phenotypes in mutant mice for our SCZ-associated
637 genes to the prevalence among other disease-associated gene sets. We selected 366 GWAS gene
638 sets, and removed any for which fewer than ten mutant mouse models were included in our
639 databases, leaving 105 gene sets. We compared the prevalence of 13 different categories of
640 psychiatric phenotypes, relating to adrenal gland, behavior, brain development, craniofacial
641 dysmorphology, ear/auditory phenotypes, eye dysmorphology, head dysmorphology, nervous
642 system development, abnormal nociception, seizures, thyroid gland, vision phenotypes. For each
643 GWAS gene set, we counted the number of categories with at least one phenotype, and
644 compared to the number in our SCZ-associated gene set to obtain an empirical p-value.
645

646 **Data Availability**

647 Our CMC-derived DLPFC prediction models will be made publicly available.
648

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

649 **Acknowledgements**

650 Data were generated as part of the CommonMind Consortium supported by funding from Takeda
651 Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd and NIH grants R01MH085542,
652 R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916,
653 P50M096891, P50MH084053S1, R37MH057881 and R37MH057881S1,
654 HHSN271201300031C, AG02219, AG05138 and MH06692.
655
656 Brain tissue for the study was obtained from the following brain bank collections: the Mount
657 Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease
658 Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories and
659 the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum
660 (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of
661 Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai,
662 Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent
663 Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks),
664 Thomas Lehner, Barbara Lipska (NIMH).
665
666 ROSMAP study data were provided by the Rush Alzheimer's Disease Center, Rush University
667 Medical Center, Chicago. Data collection was supported through funding by NIA grants
668 P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984,
669 U01AG46152, the Illinois Department of Public Health, and the Translational Genomics
670 Research Institute.
671
672 The iPSYCH-GEMS team would like to acknowledge funding from the Lundbeck Foundation
673 (grant no R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, an
674 Advanced Grant from the European Research Council (project no: 294838), the Danish Strategic
675 Research Council the Novo Nordisk Foundation for supporting the Danish National Biobank
676 resource, and grants from Aarhus and Copenhagen Universities and University Hospitals,
677 including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center.
678
679

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

680 The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the
681 Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA,
682 NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained
683 from the GTEx Portal on 09/05/16. BrainSpan: Atlas of the Developing Human Brain [Internet].
684 Funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01.
685
686
687

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

688 **CommonMind Consortium Working Group**
689
690 Jessica S Johnson 1, Hardik R Shah 2,3, Lambertus L Klein 4, Kristen K Dang 5, Benjamin A
691 Logsdon 5, Milind C Mahajan 2,3, Lara M Mangravite 5, Hiroyoshi Toyoshiba 6, Raquel E Gur
692 7, Chang-Gyu Hahn 8, Eric Schadt 2,3, David A Lewis 4, Vahram Haroutunian 1,5,9,10, Mette
693 A Peters 5, Barbara K Lipska 11, Joseph D Buxbaum 1, 12, 13, Keisuke Hirai 14, Thanneer M
694 Perumal 5, Laurent Essioux 15,
695
696 1. Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at
697 Mount Sinai, New York, New York, USA
698 2. Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai,
699 New York, New York, USA
700 3. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai,
701 New York, New York, USA
702 4. Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh,
703 Pennsylvania, USA
704 5. Sage Bionetworks, Seattle, Washington, USA
705 6. Integrated Technology Research Laboratories, Pharmaceutical Research Division, Takeda
706 Pharmaceutical Company Limited, Fujisawa, Kanagawa, Japan
707 7. Neuropsychiatry Section, Department of Psychiatry, Perelman School of Medicine,
708 University of Pennsylvania, Philadelphia, Pennsylvania, USA
709 8. Neuropsychiatric Signaling Program, Department of Psychiatry, Perelman School of
710 Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
711 9. Psychiatry, JJ Peters Virginia Medical Center, Bronx, New York, USA.
712 10. Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New
713 York, USA
714 11. Human Brain Collection Core, National Institutes of Health, NIMH, Bethesda, Maryland,
715 USA
716 12. Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York,
717 USA.
718 13. Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount
719 Sinai, New York, New York, USA
720 14. CNS Drug Discovery Unit, Pharmaceutical Research Division, Takeda Pharmaceutical
721 Company Limited, Fujisawa, Kanagawa, Japan
722 15. F. Hoffman-La Roche Ltd
723
724
725
726
727
728
729
730
731
732

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

733 **iPSYCH-GEMS SCZ working group**

734

735 Anders D. Børglum 1,2,3, Ditte Demontis 1,2,3, Veera Manikandan Rajagopal 1,2,3, Thomas D.
736 Als 1,2,3, Manuel Mattheisen 1,2,3, Jakob Grove 1,2,3,4, Thomas Werge 1,7,8, Preben Bo
737 Mortensen 1,2,9,10, Carsten Bøcker Pedersen 1,9,10, Esben Agerbo 1,9,10, Marianne Giørtz
738 Pedersen 1, 9, 10, Ole Mors 1,6, Merete Nordentoft 1, 11, David M. Hougaard 1,5, Jonas
739 Bybjerg-Grauholm 1,5, Marie Bækvad-Hansen 1,5, Christine Søholm Hansen 1,5

740

741 1. iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research,
742 Denmark
743 2. iSEQ, Center for Integrative Sequencing, Aarhus University, Aarhus, Denmark
744 3. Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark
745 4. Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark
746 5. Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut,
747 Copenhagen, Denmark
748 6. Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark
749 7. Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen,
750 Roskilde, Denmark
751 8. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
752 9. National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark
753 10. Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark
754 11. Mental Health Services in the Capital Region of Denmark, Mental Health Center
755 Copenhagen, University of Copenhagen, Copenhagen, Denmark

756

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

758
759
760

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

761 **Schizophrenia Working Group of the Psychiatric Genomics Consortium**
762 Stephan Ripke 1,2, Benjamin M. Neale1,2,3,4, Aiden Corvin 5, James T. R. Walters 6, Kai-How
763 Farh1, Peter A. Holmans 6,7, Phil Lee1,2,4, Brendan Bulik-Sullivan1,2, David A. Collier 8,9,
764 Hailiang Huang 1,3, Tune H. Pers 3,10,11, Ingrid Agartz 12,13,14, Esben Agerbo 15,16,17,
765 Margot Albus 18, Madeline Alexander 19, Farooq Amin 20,21, Silviu A. Bacanu 22, Martin
766 Begemann 23, Richard A. Belliveau Jr 2, Judit Bene 24,25, Sarah E. Bergen 2,26, Elizabeth
767 Bevilacqua 2, Tim B. Bigdeli 22, Donald W. Black 27, Richard Bruggeman28, Nancy G.
768 Buccola29, Randy L. Buckner30,31,32, William Byerley33, Wiepke Cahn34, Guiqing Cai35,36,
769 DominiqueCampion37, Rita M. Cantor38, Vaughan J. Carr39,40, Noa Carrera6, Stanley V.
770 Catts39,41, Kimberly D. Chambert2, Raymond C. K. Chan42, Ronald Y. L. Chen43, Eric Y. H.
771 Chen43,44, Wei Cheng45, Eric F. C. Cheung46, Siow Ann Chong47, C. Robert Cloninger48,
772 David Cohen49, Nadine Cohen50, Paul Cormican5, Nick Craddock6,7, James J. Crowley51,
773 David Curtis52,53, Michael Davidson54, Kenneth L. Davis36, Franziska Degenhardt55,56,
774 Jurgen Del Favero57, Ditte Demontis17,58,59, Dimitris Dikeos60, Timothy Dinan61, Srdjan
775 Djurovic14,62, Gary Donohoe5,63, Elodie Drapeau36, Jubao Duan64,65, Frank Dudbridge66,
776 Naser Durmishi67, Peter Eichhammer68, Johan Eriksson69,70,71, Valentina Escott-Price6,
777 Laurent Essioux72, Ayman H. Fanous73,74,75,76, Martilias S. Farrell51, Josef Frank77, Lude
778 Franke78, Robert Freedman79, Nelson B. Freimer80, Marion Friedl81, Joseph I. Friedman36,
779 Menachem Fromer1,2,4,82, Giulio Genovese2, Lyudmila Georgieva6, Ina Giegling81,83, Paola
780 Giusti-Rodrı´guez51, Stephanie Godard84, Jacqueline I. Goldstein1,3, Vera Golimbet85, Srihari
781 Gopal86, Jacob Gratten87, Lieuwe de Haan88, Christian Hammer23, Marian L. Hamshere6,
782 Mark Hansen89, Thomas Hansen17,90, Vahram Haroutunian36,91,92, Annette M. Hartmann81,
783 Frans A. Henskens39,93,94, Stefan Herms55,56,95, Joel N. Hirschhorn3,11,96, Per
784 Hoffmann55,56,95, Andrea Hofman55,56, Mads V. Hollegaard97, David M. Hougaard97,
785 Masashi Ikeda98, Inge Joa99, Antonio Julia 100, Rene S. Kahn34, Luba Kalaydjieva101,102,
786 Sena Karachanak-Yankova103, Juha Karjalainen78, David Kavanagh6, Matthew C. Keller104,
787 James L. Kennedy105,106,107, Andrey Khrunin108, Yunjung Kim51, Janis Klovins109, James
788 A. Knowles110, Bettina Konte81, Vaidutis Kucinskas111, Zita Ausrele Kucinskiene111, Hana
789 Kuzelova-Ptackova112, Anna K. Kahler26, Claudine Laurent19,113, Jimmy Lee Chee
790 Keong47,114, S. Hong Lee87, Sophie E. Legge6, Bernard Lerer115, Miaoxin Li43,44,116, Tao
791 Li117, Kung-Yee Liang118, Jeffrey Lieberman119, Svetlana Limborska108, Carmel M.
792 Loughland39,120, Jan Lubinski121, Jouko Lonnqvist122, Milan Macek Jr112, Patrik K. E.
793 Magnusson26, Brion S. Maher123, Wolfgang Maier124, Jacques Mallet125, Sara Marsal100,
794 Manuel Mattheisen17,58,59,126, Morten Mattingsdal14,127, Robert W. McCarley128,129,
795 ColmMcDonald130, Andrew M. McIntosh131,132, Sandra Meier77, Carin J. Meijer88, Bela
796 Melegh24,25, Ingrid Melle14,133, Raquelle I. Mesholam-Gately128,134, Andres Metspalu135,
797 Patricia T. Michie39,136, Lili Milani135, Vihra Milanova137, Younes Mokrab8, Derek W.
798 Morris5,63, Ole Mors17,58,138, Kieran C. Murphy139, Robin M. Murray140, Inez Myin-
799 Germeys141, Bertram Muller-Myhsok142,143,144, Mari Nelis135, Igor Nenadic145, Deborah
800 A. Nertney146, Gerald Nestadt147, Kristin K. Nicodemus148, Liene Nikitina-Zake109, Laura

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

801 Nisenbaum149, Annelie Nordin150, Eadbhard O'Callaghan151, ColmO'Dushlaine2,F. Anthony
802 O'Neill152, Sang-YunOh153, Ann Olincy79, Line Olsen17,90, Jim Van Os141,154, Psychosis
803 Endophenotypes International Consortium155, Christos Pantelis39,156, George N.
804 Papadimitriou 60, Sergi Papiol 23, Elena Parkhomenko36, Michele T. Pato110, Tiina
805 Paunio157,158, Milica Pejovic-Milovancevic159, Diana O. Perkins160, Olli Pietiläinen158,161,
806 Jonathan Pimm53, Andrew J. Pocklington6, John Powell140, Alkes Price3,162, Ann E. Pulver
807 147, Shaun M. Purcell 82, Digby Quested 163, Henrik B. Rasmussen 17,90, Abraham
808 Reichenberg36, Mark A. Reimers164, Alexander L. Richards6, Joshua L. Roffman30,32, Panos
809 Roussos82,165, Douglas M. Ruderfer6,82, Veikko Salomaa71, Alan R. Sanders64,65, Ulrich
810 Schall39,120, Christian R. Schubert166, Thomas G. Schulze77,167, Sibylle G. Schwab168,
811 Edward M. Scolnick2, Rodney J. Scott39,169,170, Larry J. Seidman128,134, Jianxin Shi171,
812 Engilbert Sigurdsson172, Teimuraz Silagadze173, Jeremy M. Silverman36,174, Kang Sim47,
813 Petr Slominsky108, Jordan W. Smoller2,4, Hon-Cheong So43, Chris C. A. Spencer175, Eli A.
814 Stahl3,82, Hreinn Stefansson176, Stacy Steinberg176, Elisabeth Stogmann177, Richard E.
815 Straub178, Eric Strengman179,34, Jana Strohmaier77, T. Scott Stroup119, Mythily
816 Subramaniam47, Jaana Suvisaari122, Dragan M. Svrakic48, Jin P. Szatkiewicz51, Erik
817 Soderman12, Srinivas Thirumalai180, Draga Toncheva103, Sarah Tosato181,
818 JuhaVeijola182,183, John Waddington184, Dermot Walsh185, Dai Wang86, Qiang Wang117,
819 Bradley T. Webb22, Mark Weiser54, Dieter B. Wildenauer186, Nigel M. Williams6, Stephanie
820 Williams51, Stephanie H. Witt77, Aaron R. Wolen164, Emily H. M. Wong43, Brandon K.
821 Wormley22, Hualin Simon Xi187, Clement C. Zai105,106, Xuebin Zheng188, Fritz
822 Zimprich177, Naomi R. Wray87, Kari Stefansson176, Peter M. Visscher87, Wellcome Trust
823 Case-Control Consortium 2189, Rolf Adolfsson150, Ole A. Andreassen14,133, Douglas H. R.
824 Blackwood132, Elvira Bramon190, Joseph D. Buxbaum35,36,91,191, Anders D.
825 Børglum17,58,59,138, Sven Cichon55,56,95,192, Ariel Darvasi193, Enrico Domenici194,
826 Hannelore Ehrenreich23,Tonu Esko3,11,96,135, Pablo V. Gejman64,65, Michael Gill5,
827 Hugh Gurling53, Christina M. Hultman26, Nakao Iwata98, Assen V. Jablensky39,102,186,195,
828 Erik G. Jonsson12,14, Kenneth S. Kendler196, George Kirov6, Jo Knight105,106,107,
829 ToddLencz197,198,199, Douglas F. Levinson19, QingqinS. Li86, Jianjun Liu188,200, Anil K.
830 Malhotra197,198,199, Steven A. McCarroll2,96, Andrew McQuillin53, Jennifer L. Moran2,
831 Preben B. Mortensen15,16,17, Bryan J. Mowry87,201, Markus M. Nothen55,56, Roel A.
832 Ophoff38,80,34, Michael J. Owen6,7, Aarno Palotie2,4,161, Carlos N. Pato110, Tracey L.
833 Petryshen2,128,202, Danielle Posthuma203,204,205, Marcella Rietschel77, Brien P. Riley196,
834 Dan Rujescu81,83, Pak C. Sham43,44,116, Pamela Sklar 82,91,165, David St Clair206, Daniel
835 R. Weinberger1 78,207, Jens R. Wendland166, Thomas Werge 17,90,208, Mark J. Daly1,2,3,
836 Patrick F. Sullivan 26,51,160 & Michael C. O'Donovan 6,7
837
838 1. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston,
839 Massachusetts 02114, USA. 2. Stanley Center for Psychiatric Research, Broad Institute of MIT
840 and Harvard, Cambridge, Massachusetts 02142, USA. 3. Medical and Population Genetics

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

841 Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. 4.
842 Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston,
843 Massachusetts 02114, USA. 5. Neuropsychiatric Genetics Research Group, Department of
844 Psychiatry, Trinity College Dublin, Dublin 8, Ireland. 6. MRC Centre for Neuropsychiatric
845 Genetics and Genomics, Institute of Psychological Medicine and ClinicalNeurosciences, School
846 of Medicine, Cardiff University, Cardiff CF24 4HQ,UK. 7. NationalCentre for Mental Health,
847 Cardiff University, Cardiff CF244HQ,UK. 8. Eli Lilly and Company Limited, Erl Wood Manor,
848 Sunninghill Road, Windlesham, Surrey GU20 6PH, UK. 9. Social, Genetic and Developmental
849 Psychiatry Centre, Institute of Psychiatry, King's College London, London SE58AF,UK. 10.
850 Center for BiologicalSequence Analysis, Department of Systems Biology, Technical University
851 of Denmark, DK-2800, Denmark. 11. Division of Endocrinology and Center for Basic and
852 Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts 02115,USA.
853 12. Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, SE-17176
854 Stockholm, Sweden. 13. Department of Psychiatry, Diakonhjemmet Hospital, 0319 Oslo,
855 Norway. 14. NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical
856 Medicine, University of Oslo, 0424 Oslo, Norway. 15. Centre for Integrative Register-based
857 Research, CIRRAU, Aarhus University, DK-8210 Aarhus, Denmark. 16. National Centre for
858 Register-based Research, Aarhus University, DK-8210Aarhus,Denmark. 17. The Lundbeck
859 Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark. 18. State Mental
860 Hospital, 85540 Haar, Germany. 19. Department of Psychiatry and Behavioral Sciences,
861 Stanford University, Stanford, California 94305, USA. 20. Department of Psychiatry and
862 Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia 30033, USA.
863 21. Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia
864 30322, USA. 22. Virginia Institute for Psychiatric and Behavioral Genetics, Department of
865 Psychiatry, Virginia Commonwealth University, Richmond, Virginia 23298, USA. 23. Clinical
866 Neuroscience, Max Planck Institute of Experimental Medicine, Gottingen 37075, Germany. 24.
867 Department of Medical Genetics, University of Pécs, Pécs H-7624, Hungary. 25. Szentagothai
868 Research Center, University of Pécs, Pécs H-7624, Hungary. 26. Department of Medical
869 Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-17177, Sweden. 27.
870 Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa
871 52242, USA. 28. University Medical Center Groningen, Department of Psychiatry, University of
872 Groningen NL-9700 RB, The Netherlands. 29. School of Nursing, Louisiana State University
873 Health Sciences Center, New Orleans, Louisiana 70112, USA. 30. Athinoula A. Martinos
874 Center, Massachusetts General Hospital, Boston, Massachusetts 02129, USA. 31. Center for
875 Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. 32. Department of
876 Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. 33.
877 Department of Psychiatry, University of California at San Francisco, San Francisco, California
878 94143, USA. 34. University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus
879 Institute of Neuroscience, 3584 Utrecht, The Netherlands. 35. Department of Human

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

880 Genetics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. 36.
881 Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York
882 10029, USA. 37. Centre Hospitalier du Rouvray and INSERM U1079 Faculty of Medicine,
883 76301 Rouen, France. 38. Department of Human Genetics, David Geffen School of Medicine,
884 University of California, Los Angeles, California 90095, USA. 39. Schizophrenia Research
885 Institute, Sydney NSW2010, Australia. 40. School of Psychiatry, University of New
886 South Wales, Sydney NSW2031, Australia. 41. Royal Brisbane and Women's Hospital,
887 University of Queensland, Brisbane, St Lucia QLD 4072, Australia. 42. Institute of Psychology,
888 Chinese Academy of Science, Beijing 100101, China. 43. Department of Psychiatry, Li Ka Shing
889 Faculty of Medicine, The University of Hong Kong, Hong Kong, China. 44. State Key
890 Laboratory for Brain and Cognitive Sciences, Li Ka Shing Faculty of Medicine, The University
891 of Hong Kong, Hong Kong, China. 45. Department of Computer Science, University of North
892 Carolina, Chapel Hill, North Carolina 27514, USA. 46Castle Peak Hospital, Hong Kong, China.
893 47Institute of Mental Health, Singapore 539747, Singapore. 48Department of Psychiatry,
894 Washington University, St.Louis, Missouri 63110, USA. 49. Department of Child and
895 Adolescent Psychiatry, Assistance Publique Hopitaux de Paris, Pierre and Marie Curie Faculty of
896 Medicine and Institute for Intelligent Systems and Robotics, Paris 75013, France. 50. Blue Note
897 Biosciences, Princeton, New Jersey 08540, USA 51. Department of Genetics, University of
898 North Carolina, Chapel Hill, North Carolina 27599-7264, USA. 52. Department of Psychological
899 Medicine, Queen Mary University of London, London E1 1BB, UK. 53. Molecular Psychiatry
900 Laboratory, Division of Psychiatry, University College London, LondonWC1E6JJ, UK. 54.
901 Sheba Medical Center, Tel Hashomer 52621, Israel. 55. Department of Genomics, Life and Brain
902 Center, D-53127 Bonn, Germany. 56. Institute of Human Genetics, University of Bonn, D-53127
903 Bonn, Germany. 57. AppliedMolecular Genomics Unit, VIB Department of Molecular Genetics,
904 University of Antwerp, B-2610 Antwerp, Belgium. 58. Centre for Integrative Sequencing, iSEQ,
905 Aarhus University, DK-8000 Aarhus C, Denmark. 59. Department of Biomedicine, Aarhus
906 University, DK-8000 Aarhus C, Denmark. 60. First Department of Psychiatry, University of
907 Athens Medical School, Athens 11528, Greece. 61. Department of Psychiatry, University
908 College Cork, Co. Cork, Ireland. 62. Department of Medical Genetics, Oslo University Hospital,
909 0424 Oslo, Norway. 63. Cognitive Genetics and Therapy Group, School of Psychology and
910 Discipline of Biochemistry, National University of Ireland Galway, Co. Galway, Ireland. 64.
911 Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois
912 60637, USA. 65. Department of Psychiatry and Behavioral Sciences, North Shore University
913 Health System, Evanston, Illinois 60201, USA. 66. Department of Non-Communicable Disease
914 Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. 67.
915 Department of Child and Adolescent Psychiatry, University Clinic of Psychiatry, Skopje 1000,
916 Republic of Macedonia. 68Department of Psychiatry, University of Regensburg, 93053
917 Regensburg, Germany. 69Department of General Practice, Helsinki University Central Hospital,
918 University of Helsinki P.O. Box 20, Tukholmankatu 8 B, FI-00014, Helsinki, Finland 70.
919 Folkhälsan Research Center, Helsinki, Finland, Biomedicum Helsinki 1,Haartmaninkatu 8, FI-

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

920 00290, Helsinki, Finland. 71. National Institute for Health and Welfare, P.O. Box 30, FI-00271
921 Helsinki, Finland. 72. Translational Technologies and Bioinformatics, Pharma Research
922 and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland. 73. Department of
923 Psychiatry, Georgetown University School of Medicine, Washington DC 20057, USA. 74.
924 Department of Psychiatry, Keck School of Medicine of the University of Southern California,
925 Los Angeles, California 90033, USA. 75. Department of Psychiatry, Virginia Commonwealth
926 University School of Medicine, Richmond, Virginia 23298, USA. 76. Mental Health Service
927 Line, Washington VA Medical Center, Washington DC 20422, USA. 77. Department of Genetic
928 Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim,
929 University of Heidelberg, Heidelberg , D-68159 Mannheim, Germany. 78. Department of
930 Genetics, University of Groningen, University Medical Centre Groningen, 9700 RB Groningen,
931 The Netherlands. 79. Department of Psychiatry, University of Colorado Denver, Aurora,
932 Colorado 80045, USA. 80. Center for Neurobehavioral Genetics, Semel Institute for
933 Neuroscience and Human Behavior, University of California, Los Angeles, California 90095,
934 USA. 81Department of Psychiatry, University of Halle, 06112 Halle, Germany. 82. Division of
935 Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New
936 York, NewYork 10029,USA. 83. Department of Psychiatry, University of Munich, 80336,
937 Munich, Germany. 84. Departments of Psychiatry and Human and Molecular Genetics,
938 INSERM, Institut de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris 75013, France. 85. Mental
939 Health Research Centre, Russian Academy of Medical Sciences, 115522 Moscow, Russia. 86.
940 Neuroscience Therapeutic Area, Janssen Research and Development, Raritan, New
941 Jersey 08869, USA. 87. Queensland Brain Institute, The University of Queensland, Brisbane,
942 Queensland, QLD 4072, Australia. 88. Academic Medical Centre University of Amsterdam,
943 Department of Psychiatry, 1105 AZ Amsterdam, The Netherlands. 89. Illumina, La Jolla,
944 California, California 92122, USA. 90Institute of Biological Psychiatry, Mental Health Centre
945 Sct.Hans, Mental Health Services Copenhagen, DK-4000, Denmark. 91. Friedman Brain
946 Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. 92. J. J.
947 Peters VA Medical Center, Bronx, New York, New York 10468, USA. 93. Priority Research
948 Centre for Health Behaviour, University of Newcastle, Newcastle NSW 2308, Australia. 94.
949 School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle
950 NSW2308, Australia. 95. Division of Medical Genetics, Department of Biomedicine, University
951 of Basel, Basel CH-4058, Switzerland. 96. Department of Genetics, Harvard Medical School,
952 Boston, Massachusetts, Massachusetts 02115, USA. 97. Section of Neonatal Screening and
953 Hormones, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum
954 Institut, Copenhagen DK-2300, Denmark. 98. Department of Psychiatry, Fujita Health
955 University School of Medicine, Toyoake, Aichi,470-1192, Japan. 99. Regional Centre for
956 Clinical Researchin Psychosis, Department of Psychiatry, Stavanger University Hospital, 4011
957 Stavanger, Norway. 100. Rheumatology Research Group, Vall d'Hebron Research Institute,
958 Barcelona 08035, Spain. 101. Centre for Medical Research, The University of Western Australia,
959 Perth WA6009, Australia. 102. The Perkins Institute for Medical Research, The University of

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

960 Western Australia, Perth WA6009, Australia. 103. Department of Medical Genetics, Medical
961 University, Sofia 1431, Bulgaria. 104. Department of Psychology, University of Colorado
962 Boulder, Boulder, Colorado 80309, USA. 105. Campbell Family Mental Health Research
963 Institute, Centre for Addiction and Mental Health, Toronto, Ontario M5T 1R8, Canada. 106.
964 Department of Psychiatry, University of Toronto, Toronto, Ontario M5T 1R8, Canada. 107.
965 Institute of Medical Science,University of Toronto, Toronto,OntarioM5S1A8, Canada. 108.
966 Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia. 109.
967 Latvian Biomedical Research and Study Centre, Riga, LV-1067, Latvia. 110. Department of
968 Psychiatry and Zilkha Neurogenetics Institute, Keck School of Medicine at University of
969 Southern California, Los Angeles, California 90089, USA. 111. Faculty of Medicine, Vilnius
970 University, LT-01513 Vilnius, Lithuania. 112. Department of Biology and Medical Genetics,
971 2nd Faculty of Medicine and University Hospital Motol, 150 06 Prague, Czech Republic. 113.
972 Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine,
973 Paris 75013, France. 114. Duke-NUS Graduate Medical School, Singapore 169857. 115.
974 Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem 91120,
975 Israel. 116. Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China.
976 117. Mental Health Centre and Psychiatric Laboratory, West China Hospital, Sichuan
977 University, Chengdu, 610041 Sichuan, China. 118. Department of Biostatistics, Johns Hopkins
978 University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. 119.
979 Department of Psychiatry, Columbia University, New York, New York 10032, USA. 120.
980 Priority Centre for Translational Neuroscience and Mental Health, University of Newcastle,
981 Newcastle NSW 2300, Australia. 121. Department of Genetics and Pathology, International
982 Hereditary Cancer Center, Pomeranian Medical University in Szczecin, 70-453 Szczecin,
983 Poland. 122. Department of Mental Health and Substance Abuse Services; National Institute for
984 Health and Welfare, P.O.BOX30, FI-00271Helsinki, Finland. 123. Department of Mental Health,
985 Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205,
986 USA. 124. Department of Psychiatry, University of Bonn, D-53127 Bonn, Germany. 125. Centre
987 National de la Recherche Scientifique, Laboratoire de Génétique Moléculaire de la
988 Neurotransmission et des Processus Neurodénégératifs, Hôpital de la Pitiè-Salpêtrière 75013
989 Paris, France. 126. Department of Genomics Mathematics, University of Bonn, D-53127 Bonn,
990 Germany. 127. Research Unit, Sørlandet Hospital, 4604 Kristiansand, Norway. 128. Department
991 of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA. 129. VA Boston
992 Health Care System, Brockton, Massachusetts 02301, USA. 130. Department of Psychiatry,
993 National University of Ireland Galway, Co. Galway, Ireland. 131. Centre for Cognitive Ageing
994 and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH16 4SB, UK. 132. Division
995 of Psychiatry, University of Edinburgh, Edinburgh EH16 4SB, UK. 133. Division of Mental
996 Health and Addiction, Oslo University Hospital, 0424 Oslo, Norway. 134. Massachusetts Mental
997 Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Boston,
998 Massachusetts 02114, USA. 135. Estonian Genome Center, University of Tartu, Tartu 50090,
999 Estonia. 136. School of Psychology, University of Newcastle, Newcastle NSW2308, Australia.

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1000 137. First Psychiatric Clinic, Medical University, Sofia 1431, Bulgaria. 138. Department P,
1001 Aarhus University Hospital, DK-8240 Risskov, Denmark. 139. Department of Psychiatry, Royal
1002 College of Surgeons in Ireland, Dublin 2, Ireland. 140. King's College London, London
1003 SE58AF, UK. 141. Maastricht University Medical Centre, South Limburg Mental Health
1004 Research and TeachingNetwork,EURON,6229HX Maastricht, The Netherlands. 142. Institute of
1005 Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK. 143. Max Planck
1006 Institute of Psychiatry, 80336 Munich, Germany. 144. Munich Cluster for Systems Neurology
1007 (SyNergy),80336 Munich, Germany. 145. Department of Psychiatry and Psychotherapy, Jena
1008 University Hospital, 07743 Jena, Germany. 146. Department of Psychiatry, Queensland Brain
1009 Institute and Queensland Centre for Mental Health Research, University of Queensland,
1010 Brisbane, Queensland, St Lucia QLD 4072, Australia. 147. Department of Psychiatry and
1011 Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
1012 21205, USA. 148. Department of Psychiatry, Trinity College Dublin, Dublin 2, Ireland. 149. Eli
1013 Lilly and Company, Lilly Corporate Center, Indianapolis, 46285 Indiana, USA. 150. Department
1014 of Clinical Sciences, Psychiatry, Umeå University, SE-901 87Umeå , Sweden. 151. DETECT
1015 Early Intervention Service for Psychosis, Blackrock, Co.Dublin, Ireland. 152. Centre for Public
1016 Health, Institute of Clinical Sciences, Queen's University Belfast, Belfast BT12 6AB, UK. 153.
1017 Lawrence Berkeley National Laboratory, University of California at Berkeley, Berkeley,
1018 California 94720,USA. 154. Institute of Psychiatry, King's College London, London SE5
1019 8AF,UK. 155. A list of authors and affiliations appear in the Supplementary Information. 156.
1020 Melbourne Neuropsychiatry Centre, University of Melbourne & Melbourne Health, Melbourne,
1021 Vic 3053, Australia. 157. Department of Psychiatry, University of Helsinki, P.O. Box 590, FI-
1022 00029 HUS, Helsinki, Finland. 158. Public Health Genomics Unit, National Institute for Health
1023 and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland 159. Medical Faculty, University of
1024 Belgrade, 11000 Belgrade, Serbia. 160. Department of Psychiatry, University of North Carolina,
1025 Chapel Hill, North Carolina 27599-7160, USA. 161. Institute for Molecular Medicine Finland,
1026 FIMM, University of Helsinki, P.O. Box 20FI-00014, Helsinki, Finland 162. Department of
1027 Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. 163.
1028 Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK. 164. Virginia Institute
1029 for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond,
1030 Virginia 23298, USA. 165. Institute for Multiscale Biology, Icahn School of Medicine at Mount
1031 Sinai, New York, New York 10029, USA. 166. Pharma Therapeutics Clinical Research, Pfizer
1032 Worldwide Research and Development, Cambridge, Massachusetts 02139, USA. 167.
1033 Department of Psychiatry and Psychotherapy, University of Gottingen, 37073 Göttingen,
1034 Germany. 168. Psychiatry and Psychotherapy Clinic, University of Erlangen, 91054 Erlangen,
1035 Germany. 169Hunter New England Health Service, Newcastle NSW2308, Australia. 170. School
1036 of Biomedical Sciences, University of Newcastle, Newcastle NSW2308, Australia. 171. Division
1037 of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892,
1038 USA. 172. University of Iceland, Landspitali, National University Hospital, 101 Reykjavik,
1039 Iceland. 173. Department of Psychiatry and Drug Addiction, Tbilisi State Medical University

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1040 (TSMU), N33,0177 Tbilisi, Georgia. 174. Research and Development, Bronx Veterans Affairs
1041 Medical Center, New York, New York 10468, USA. 175. WellcomeTrust Centre for Human
1042 Genetics, Oxford OX3 7BN, UK. 176. deCODE Genetics, 101 Reykjavik, Iceland. 177/
1043 Department of Clinical Neurology, Medical University of Vienna, 1090 Wien, Austria. 178.
1044 Lieber Institute for Brain Development, Baltimore, Maryland 21205, USA. 179. Department of
1045 Medical Genetics, University Medical Centre Utrecht, Universiteitsweg 100,3584CG, Utrecht,
1046 The Netherlands. 180. Berkshire Healthcare NHS Foundation Trust, Bracknell RG12 1BQ, UK.
1047 181. Section of Psychiatry, University of Verona, 37134 Verona, Italy. 182. Department of
1048 Psychiatry, University of Oulu, P.O. Box 5000, 90014, Finland. 183. University Hospital of
1049 Oulu, P.O. Box 20, 90029 OYS, Finland. 184. Molecular and Cellular Therapeutics, Royal
1050 College of Surgeons in Ireland, Dublin 2, Ireland. 185. Health Research Board, Dublin 2,
1051 Ireland. 186. School of Psychiatry and Clinical Neurosciences, The University of Western
1052 Australia, Perth WA6009, Australia. 187. Computational Sciences CoE, Pfizer Worldwide
1053 Research and Development, Cambridge, Massachusetts 02139, USA. 188. Human Genetics,
1054 Genome Institute of Singapore, A*STAR, Singapore 138672. 189. A list of authors and
1055 affiliations appear in the Supplementary Information. 190. University College London, London
1056 WC1E 6BT, UK. 191. Department of Neuroscience, Icahn School of Medicine at Mount
1057 Sinai,NewYork,NewYork10029,USA. 192. Institute of Neuroscience and Medicine (INM-1),
1058 Research Center Juelich, 52428 Juelich, Germany. 193. Department of Genetics, The Hebrew
1059 University of Jerusalem, 91905 Jerusalem, Israel. 194. NeuroscienceDiscovery and Translational
1060 Area, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel,
1061 Switzerland. 195. Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and
1062 Clinical Neurosciences, The University of Western Australia, Medical Research Foundation
1063 Building, Perth WA6000, Australia. 196. Virginia Institute for Psychiatric and Behavioral
1064 Genetics, Departments of Psychiatry and Human and Molecular Genetics, Virginia
1065 Commonwealth University, Richmond, Virginia 23298, USA. 197. The Feinstein Institute for
1066 Medical Research, Manhasset, New York 11030, USA. 198. The Hofstra NS-LIJ School of
1067 Medicine, Hempstead, New York 11549, USA. 199. The Zucker Hillside Hospital, Glen Oaks,
1068 New York 11004, USA. 200Saw Swee Hock School of Public Health, National University of
1069 Singapore, Singapore 117597, Singapore. 201. Queensland Centre for Mental Health Research,
1070 University of Queensland, Brisbane 4076, Queensland, Australia. 202. Center for HumanGenetic
1071 Research and Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts
1072 02114, USA. 203. Department of Child and Adolescent Psychiatry, Erasmus University Medical
1073 Centre, Rotterdam3000, The Netherlands. 204. Department of Complex Trait Genetics,
1074 NeuroscienceNeuroscience Campus Amsterdam, VU University Medical Center Amsterdam,
1075 Amsterdam1081, The Netherlands. 205. Department of Functional Genomics, Center for
1076 Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University,
1077 Amsterdam1081, The Netherlands. 206. University of Aberdeen, Institute of Medical Sciences,
1078 Aberdeen AB25 2ZD, UK. 207. Departments of Psychiatry, Neurology, Neuroscience and
1079 Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205,

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1080 USA. 208. Department of Clinical Medicine, University of Copenhagen, Copenhagen 2200,
1081 Denmark.

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1082 **References**

1083 1. Nicolae, D. L. *et al.* Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to
1084 Enhance Discovery from GWAS. *PLoS Genet.* **6,** e1000888 (2010).
1085 2. Dobbyn, A. *et al.* Co-localization of Conditional eQTL and GWAS Signatures in
1086 Schizophrenia. *Rev.* (2017).
1087 3. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation:
1088 the promise of eQTL studies. *Trends Genet.* **24,** 408–415 (2008).
1089 4. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex
1090 disease traits with global gene expression. *Nat. Rev. Genet.* **10,** 184–194 (2009).
1091 5. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and
1092 disease. *Nat. Rev. Genet.* **16,** 197–212 (2015).
1093 6. Moffatt, M. F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the
1094 risk of childhood asthma. *Nature* **448,** 470–473 (2007).
1095 7. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci
1096 associated with body mass index. *Nat. Genet.* **42,** 937–948 (2010).
1097 8. Dubois, P. C. A. *et al.* Multiple common variants for celiac disease influencing immune
1098 gene expression. *Nat. Genet.* **42,** 295–302 (2010).
1099 9. Libioulle, C. *et al.* Novel Crohn disease locus identified by genome-wide association
1100 maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* **3,**
1101 e58 (2007).
1102 10. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic
1103 Association Studies Using Summary Statistics. *PLoS Genet.* **10,** e1004383 (2014).
1104 11. Boocock, J., Giambartolomei, C. & Stahl, E. A. COLOC2. (2016).
1105 12. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts
1106 complex trait gene targets. *Nat. Genet.* **48,** 481–487 (2016).
1107 13. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human
1108 traits. *Nat. Genet.* **48,** 709–717 (2016).
1109 14. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for
1110 schizophrenia. *Nat. Neurosci.* **19,** 1442–1453 (2016).
1111 15. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
1112 transcriptome data. *Nat. Genet.* **47,** 1091–8 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1113 16. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association
1114 studies. *Nat. Genet.* **48,** 245–52 (2016).
1115 17. Geschwind, D. H. & Flint, J. Genetics and genomics of psychiatric disease. *Science (80-.
1116).* **349,** (2015).
1117 18. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for
1118 removing batch effects and other unwanted variation in high-throughput experiments.
1119 *Bioinformatics* **28,** 882–3 (2012).
1120 19. A. Bennett, D., A. Schneider, J., Arvanitakis, Z. & S. Wilson, R. Overview and Findings
1121 from the Religious Orders Study. *Curr. Alzheimer Res.* **9,** 628–645 (2012).
1122 20. Mele, M. *et al.* The human transcriptome across tissues and individuals. *Science (80-.).*
1123 **348,** 660–665 (2015).
1124 21. (GTex Consortium). GTEx Portal. (2015). at
1125 <http://gtexportal.org/home/documentationPage>
1126 22. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature*
1127 **511,** 421–427 (2014).
1128 23. Benjamin, Y. & Hochberg, Y. Controlling the False Discovery Rate: a Practical and
1129 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 289–300
1130 (1995).
1131 24. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic
1132 function and autism. *Cell* **146,** 247–61 (2011).
1133 25. Fromer, M. *et al. Gene Expression Elucidates Functional Impact of Polygenic Risk for
1134 Schizophrenia*. *bioRxiv* (Cold Spring Harbor Labs Journals, 2016). doi:10.1101/052209
1135 26. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant
1136 genes and maintained by background selection. *bioRxiv* 68593 (2016).
1137 doi:10.1101/068593
1138 27. Sanders, S. J. First glimpses of the neurobiology of autism spectrum disorder. *Curr. Opin.
1139 Genet. Dev.* **33,** 80–92 (2015).
1140 28. Exome Aggregation Consortium *et al. Analysis of protein-coding genetic variation in
1141 60,706 humans*. *bioRxiv* (Cold Spring Harbor Labs Journals, 2015). doi:10.1101/030338
1142 29. Malhotra, D. *et al.* High frequencies of de novo CNVs in bipolar disorder and
1143 schizophrenia. *Neuron* **72,** 951–963 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1144 30. Bautista, O., Vázquez-Caubet, J. C., Zhivago, E. A. & Dolores Sáiz, M. From metabolism
1145 to psychiatric symptoms: psychosis as a manifestation of acute intermittent porphyria. *J.*
1146 *Neuropsychiatry Clin. Neurosci.* **26,** E30 (2014).
1147 31. Zimmermann, M., Bonaccurso, C., Valerius, C. & Hamann, G. F. [Acute intermittent
1148 porphyria. A clinical chameleon: case study of a 40-year-old female patient]. *Nervenarzt*
1149 **77,** 1501–5 (2006).
1150 32. Ventura, P. *et al.* A challenging diagnosis for potential fatal diseases: recommendations
1151 for diagnosing acute porphyrias. *Eur. J. Intern. Med.* **25,** 497–505 (2014).
1152 33. Pischik, E. & Kauppinen, R. An update of clinical management of acute intermittent
1153 porphyria. *Appl. Clin. Genet.* **8,** 201–14 (2015).
1154 34. Kumar, B. Acute intermittent porphyria presenting solely with psychosis: a case report
1155 and discussion. *Psychosomatics* **53,** 494–8 (2012).
1156 35. Bonnot, O. *et al.* Diagnostic and treatment implications of psychosis secondary to
1157 treatable metabolic disorders in adults: a systematic review. *Orphanet J. Rare Dis.* **9,** 65
1158 (2014).
1159 36. Kaback, M. M. & Desnick, R. J. *Hexosaminidase A Deficiency*. *GeneReviews(®)*
1160 (University of Washington, Seattle, 1993).
1161 37. Saleh, O. Late Onset Tay-Sachs Disease Pre senting as a Brief Psychotic Disorder with
1162 Catatonia: A Case Report and Review of Literature.
1163 38. Skaper, S. D. in *Brain Protection in Schizophrenia, Mood and Cognitive Disorders* (ed.
1164 Ritsner, M. S.) 135–165 (Springer Science & Business Media, 2010, 2010).
1165 39. Castellano, E. *et al.* RAS signalling through PI3-Kinase controls cell migration via
1166 modulation of Reelin expression. *Nat. Commun.* **7,** 11245 (2016).
1167 40. Gururajan, A. & Buuse, M. van den. Is the mTOR-signalling cascade disrupted in
1168 Schizophrenia? *J. Neurochem.* **129,** 377–387 (2014).
1169 41. Ritsner, M. S. *Brain Protection in Schizophrenia, Mood and Cognitive Disorders*.
1170 (Springer Science & Business Media, 2010, 2010). doi:10.1007/978-90-481-8553-5
1171 42. Enriquez-Barreto, L. & Morales, M. The PI3K signaling pathway as a pharmacological
1172 target in Autism related disorders and Schizophrenia. *Mol. Cell. Ther.* **4,** 2 (2016).
1173 43. Glessner, J. T. *et al.* Strong synaptic transmission impact by copy number variations in
1174 schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10584–9 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1175 44. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium.
1176 Psychiatric genome-wide association study analyses implicate neuronal, immune and
1177 histone pathways. *Nat. Neurosci.* **18,** 199–209 (2015).
1178 45. Bauman, A. L. *et al.* Cocaine and antidepressant-sensitive biogenic amine transporters
1179 exist in regulated complexes with protein phosphatase 2A. *J. Neurosci.* **20,** 7571–8
1180 (2000).
1181 46. He, X. *et al.* Sherlock: detecting gene-disease associations by matching patterns of
1182 expression QTL and GWAS. *Am J Hum Genet* **92,** 667–680 (2013).
1183 47. Ayadi, A. *et al.* Mouse large-scale phenotyping initiatives: overview of the European
1184 Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse
1185 Genetics Project. *Mamm. Genome* **23,** 600–610 (2012).
1186 48. MGI-About the Mouse Genome Informatics database resource. at
1187 <http://www.informatics.jax.org/mgihome/projects/aboutmgi.shtml>
1188 49. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene
1189 regulation. *Nature* **477,** 289–294 (2011).
1190 50. About KOMP | IMPC.
1191 51. Howe, D. G. *et al.* ZFIN, the Zebrafish Model Organism Database: increased support for
1192 mutants and transgenics. *Nucleic Acids Res.* **41,** D854–D860 (2013).
1193 52. TRANSCRIPTOME PROFILING BY RNA SEQUENCING AND EXON
1194 MICROARRAY TISSUE ACQUISITION AND QUALIFICATION.
1195 53. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393,**
1196 440–442 (1998).
1197 54. Schork, A. J. *et al.* All SNPs Are Not Created Equal: Genome-Wide Association Studies
1198 Reveal a Consistent Pattern of Enrichment among Functionally Annotated SNPs. *PLoS*
1199 *Genet.* **9,** e1003449 (2013).
1200 55. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to
1201 Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100,** 473–487
1202 (2017).
1203 56. Gottlieb, A., Daneshjou, R., DeGorter, M., Montgomery, S. & Altman, R. Population-
1204 specific imputation of gene expression improves prediction of pharmacogenomic traits for
1205 African Americans. *bioRxiv* 115451 (2017). doi:10.1101/115451

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1206 57. Need, A. & Goldstein, D. B. Next generation disparities in human genomics: concerns and
1207 remedies. *Trends Genet* **25,** 489–94 (2009).
1208 58. Popejoy, A. & Fullerton, S. Genomics is failing on diversity. *Nature* **538,** 161–164 (2016).
1209 59. Browning, R. in *The Poems of Robert Browning* (eds. Porter, C. & Clarke, H. A.) 257–271
1210 (Thomas Y. Cromwell and Company, 1896).
1211 60. Loftus, L. S. & Arnold, W. N. Vincent van Gogh's illness: acute intermittent porphyria?
1212 *BMJ* **303,** 1589–91
1213 61. Strik, W. K. [The psychiatric illness of Vincent van Gogh]. *Nervenarzt* **68,** 401–9 (1997).
1214 62. Arnold, W. N. The illness of Vincent van Gogh. *J. Hist. Neurosci.* **13,** 22–43 (2004).
1215 63. Hughes, J. R. A reappraisal of the possible seizures of Vincent van Gogh. *Epilepsy Behav.*
1216 **6,** 504–10 (2005).
1217 64. Bhattacharyya, K. B. & Rai, S. The neuropsychiatric ailment of Vincent Van Gogh. *Ann.*
1218 *Indian Acad. Neurol.* **18,** 6–9 (2014).
1219 65. Correa, R. Vincent van Gogh: A pathographic analysis. *Med. Hypotheses* **82,** 141–144
1220 (2014).
1221 66. Peters, T. J. & Beveridge, A. The madness of King George III: a psychiatric re-
1222 assessment. *Hist. Psychiatry* **21,** 20–37 (2010).
1223 67. Szatkiewicz, J. P. *et al.* Copy number variation in schizophrenia in Sweden. *Mol.*
1224 *Psychiatry* **19,** 762–773 (2014).
1225 68. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature*
1226 **506,** 179–184 (2014).
1227 69. Consortium, C.-D. G. of the P. G. Identification of risk loci with shared effects on five
1228 major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).
1229 70. Keefe, R. S. E. & Fenton, W. S. How Should DSM-V Criteria for Schizophrenia Include
1230 Cognitive Impairment? *Schizophr. Bull.* **33,** 912–920 (2007).
1231 71. Reichenberg, A. *et al.* Static and Dynamic Cognitive Deficits in Childhood Preceding
1232 Adult Schizophrenia: A 30-Year Study. *Am. J. Psychiatry* **167,** 160–169 (2010).
1233 72. Gold, J. M. Cognitive deficits as treatment targets in schizophrenia. *Schizophr. Res.* **72,**
1234 21–28 (2004).
1235 73. Cannon, M. *et al.* Evidence for Early-Childhood, Pan-Developmental Impairment Specific
1236 to Schizophreniform Disorder. *Arch. Gen. Psychiatry* **59,** 449 (2002).

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1237 74. Parikshak, N. N., Gandal, M. J., Geschwind, D. H. & Angeles, L. Systems biology and
1238 gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.*
1239 **16,** 441–458 (2015).
1240 75. Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain.
1241 *Genome Biol.* **14,** R75 (2013).
1242 76. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human
1243 prefrontal cortex. *Nature* **478,** 519–523 (2012).
1244 77. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin
1245 activity yields mechanistic disease insights. *bioRxiv* 67355 (2016). doi:10.1101/067355
1246 78. ROSMAP Study - syn3219045. at <https://www.synapse.org/#!Synapse:syn3219045>
1247 79. NIMH » Human Brain Collection Core (HBCC). at <https://www.nimh.nih.gov/labs-at-
1248 nimh/research-areas/research-support-services/hbcc/human-brain-collection-core-
1249 hbcc.shtml>
1250 80. Barbeira, A. *et al.* MetaXcan: Summary Statistics Based Gene-Level Association Method
1251 Infers Accurate PrediXcan Results. *bioRxiv* (2016).
1252 81. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue
1253 gene regulation in humans. *Science (80-.).* **348,** 648–660 (2015).
1254 82. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
1255 genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
1256 83. COCHRAN, W. G. THE COMPARISON OF PERCENTAGES IN MATCHED
1257 SAMPLES. *Biometrika* **37,** 256–266 (1950).
1258 84. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide
1259 complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).
1260 85. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-
1261 set analysis of GWAS data. *PLoS Comput. Biol.* **11,** e1004219 (2015).
1262 86. Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic
1263 signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17,** 142–53
1264 (2012).
1265 87. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene
1266 Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
1267 88. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic*

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1268 *Acids Res.* **43,** D1049-1056 (2014).
1269 89. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic*
1270 *Acids Res.* **28,** 27–30 (2000).
1271 90. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42,** D472-7
1272 (2014).
1273 91. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by
1274 function. *Genome Res.* **13,** 2129–2141 (2003).
1275 92. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of
1276 gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic*
1277 *Acids Res.* **41,** D377-86 (2013).
1278 93. BioCarta. MSigDB Collections. (2017).
1279 94. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508,**
1280 199–206 (2014).
1281 95. Lin, G. N. *et al.* Spatiotemporal 16p11.2 Protein Network Implicates Cortical Late Mid-
1282 Fetal Brain Development and KCTD13- Cul3-RhoA Pathway in Psychiatric Diseases.
1283 *Neuron* **85,** 742–754 (2015).
1284 96. Bahl, E., Koomar, T. & Michaelson, J. J. cerebroViz: An R package for anatomical visu-
1285 alization of spatiotemporal brain data. *Bioinformatics* **33,** btw726 (2016).
1286 97. van der Weyden, L., White, J. K., Adams, D. J. & Logan, D. W. The mouse genetics
1287 toolkit: revealing function and mechanism. *Genome Biol.* **12,** 224 (2011).
1288 98. Home | 3i Consortium - Wellcome Trust Sanger Institute. at
1289 <http://www.immunophenotype.org/>
1290 99. Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research
1291 Register. *Scand. J. Public Health* **39,** 54–57 (2011).
1292 100. Børglum, A. D. *et al.* Genome-wide study of association and interaction with maternal
1293 cytomegalovirus infection suggests new schizophrenia loci. *Mol. Psychiatry* **19,** 325–33
1294 (2014).
1295 101. Hollegaard, M. V *et al.* Robustness of genome-wide scanning using archived dried blood
1296 spot samples as a DNA source. *BMC Genet.* **12,** 58 (2011).
1297 102. Illumina. illumina GenCall Data Analysis Software. Illumina Tech Note. (2005). at
1298 <https://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysi

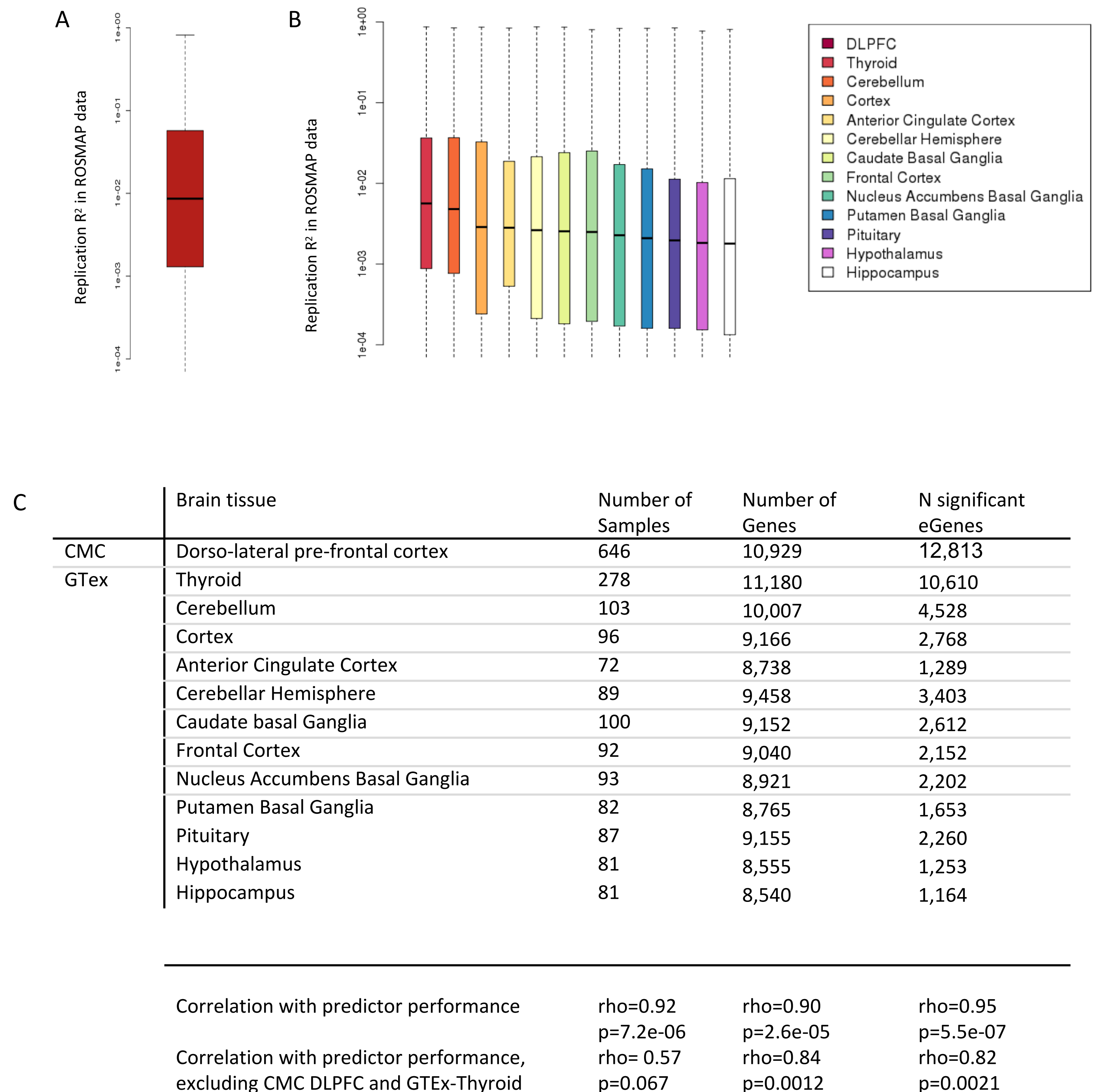
bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

s_software.pdf>

103. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40,** 1253–60 (2008).

104. Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28,** 2543–2545 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



C

	Brain tissue	Number of Samples	Number of Genes	N significant eGenes
CMC	Dorso-lateral pre-frontal 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=7.2e-06	rho=0.90 p=2.6e-05	rho=0.95 p=5.5e-07
	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

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

A. Distribution of R_R^2 values of CMC DLPFC predictors in ROSMAP data. Mean $R_R^2 = 0.056$. 47.7% of genes have $R_R^2 >= 0.01$.
B. Distribution of R_R^2 values of 12 GTeX predictors in ROSMAP data.
C. Table of sample sizes and p-val thresholds for CMC DLPFC and GTeX data. Number of samples, number of genes in the prediXcan model and number of eGenes are all significantly correlated with predictor performance in ROSMAP data.

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

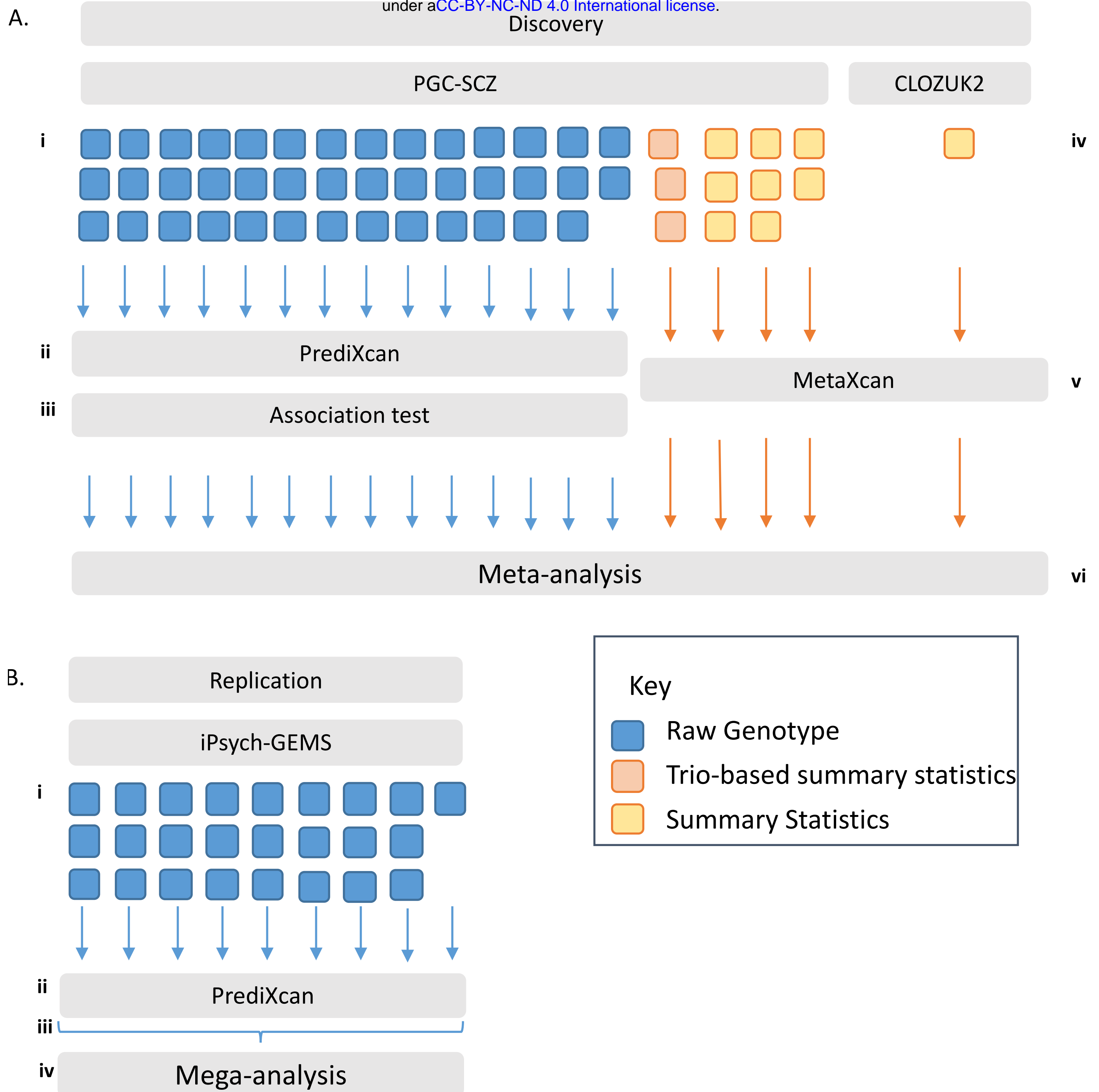


Figure 2: Analysis outline.

A) Discovery Samples. 41 PGC-SCZ cohorts had available raw genotypes (i). Predicted DLPFC gene expression was calculated in each cohort using prediXcan (ii) and tested for association with case-control status (iii). 11 PGC cohorts (3 trio, 8 case-control) and the CLOZUK2 cohort had only summary statistics available (iv). MetaXcan was used to calculate DLPFC associations for each cohort (v). Results were meta-analysed across all 53 cohorts (vi). This procedure was repeated for 12 GTEx prediction models.

B) Replication Samples. iPSYCH-GEMS samples were collected in 25 waves (i). Predicted DLPFC gene expression was calculated in each wave separately using prediXcan (ii) and merged for association testing (iii). A mega-analysis was run across all 25 waves, using wave membership as a covariate in the regression (iv)

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

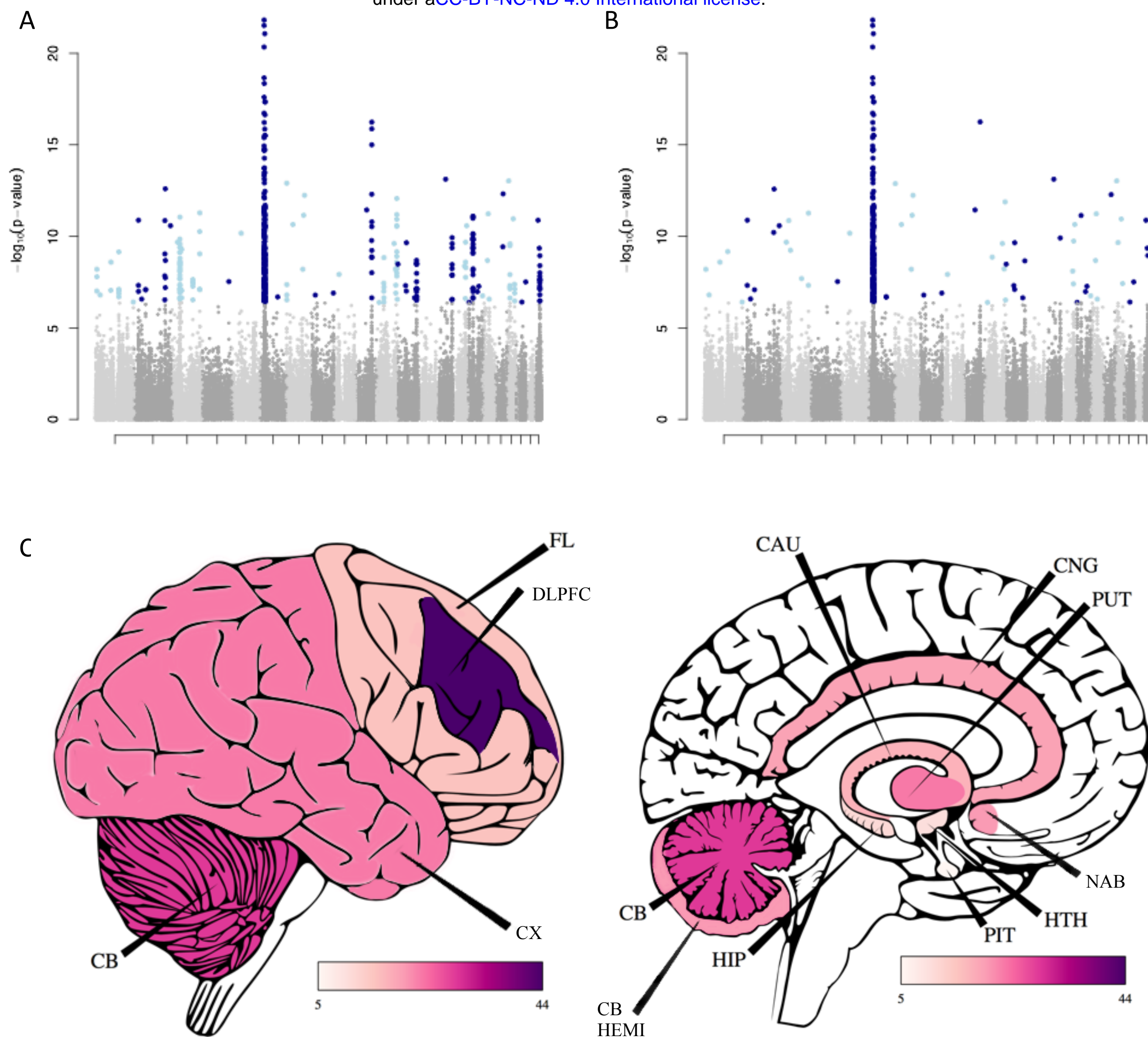


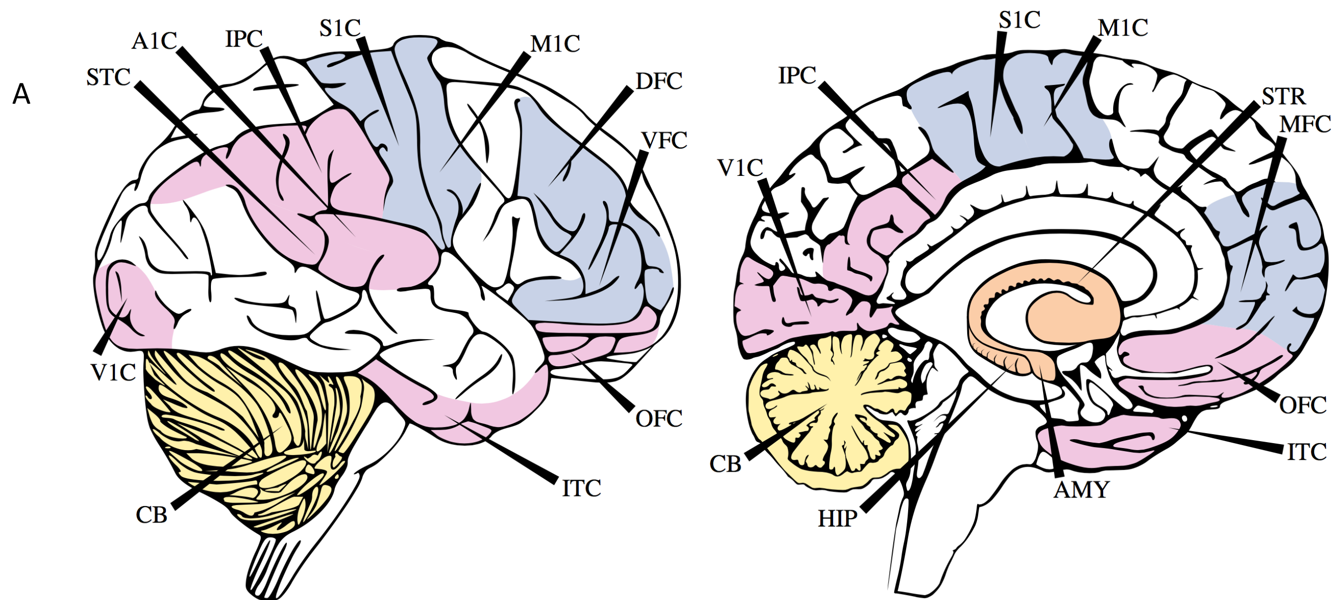
Figure 3: SCZ associations results

A) 413 genes are associated with SCZ across 12 brain tissues

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

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



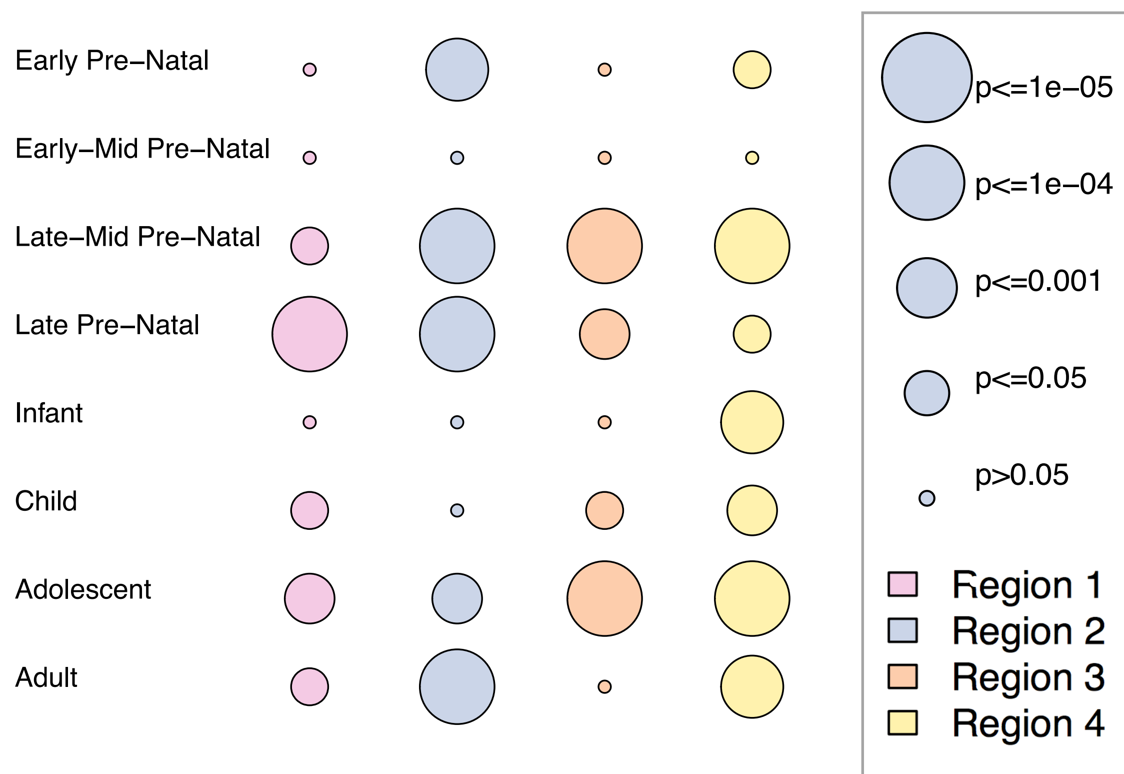


Figure 4: SCZ-associated genes are co-expressed throughout development and across brain regions

A) Brain tissues selected for each of four BRAINSPAN regions. 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 to permuted gene networks to obtain empirical significance levels.

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

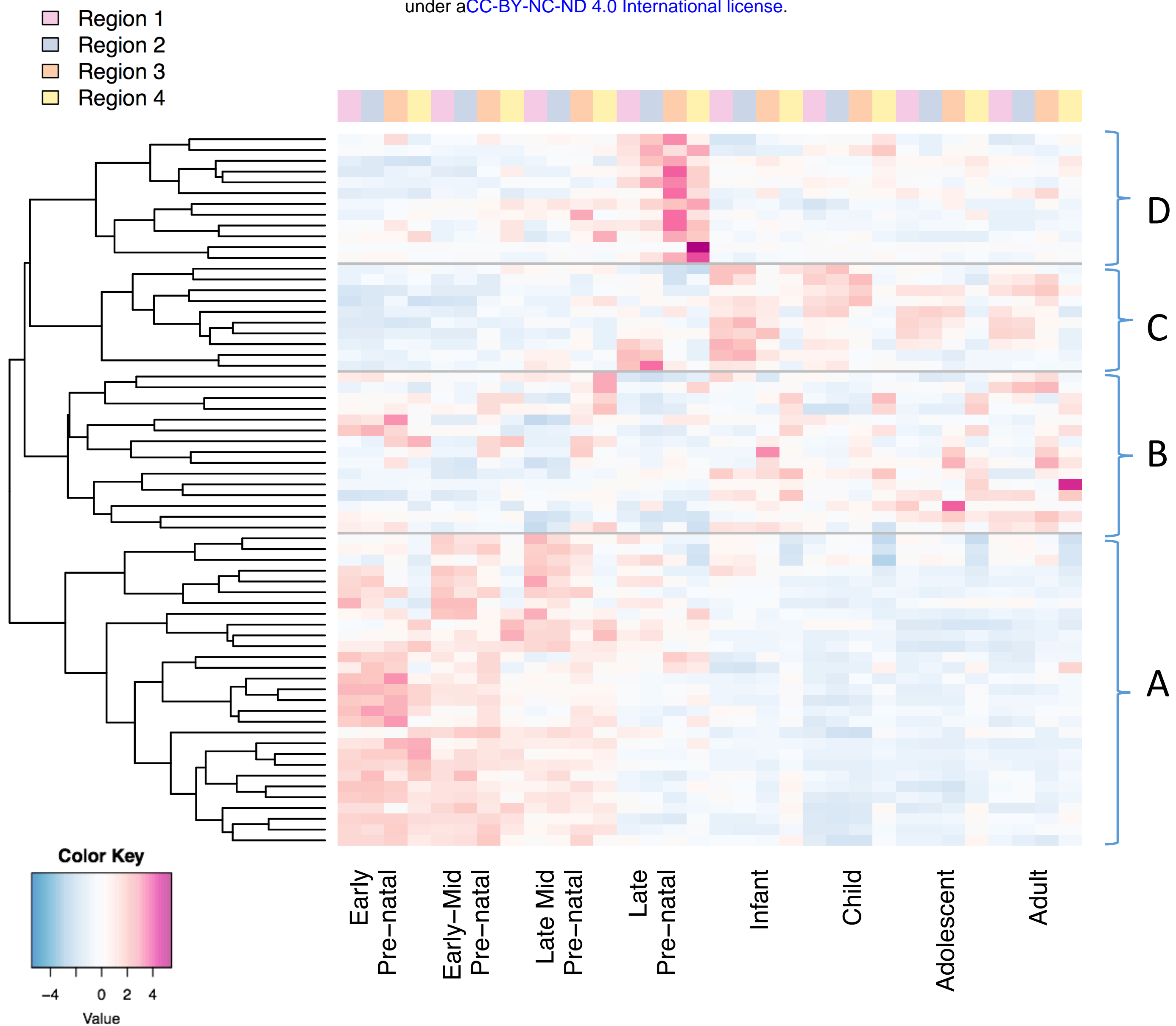


Figure 5: Gene expression patterns for SCZ-associated genes cluster into four groups, relating to distinct spatiotemporal expression.
Brain regions are shown in figure 5a.

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

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: SCZ-associated genes

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

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DMC1	CMC	DLPFC	7.229	4.48E-10	0.000	0.130	1.139
RAD51D	**CMC**	**DLPFC**	7.612	2.11E-09	0.000	0.111	1.117
RERE	CMC	DLPFC	2.847	6.32E-09	0.000	0.036	1.037
PCCB	CMC	DLPFC	-0.044	2.05E-08	0.054	-0.010	0.990
CLCN3	CMC	DLPFC	0.141	2.96E-08	0.005	0.010	1.010
ATG101	**CMC**	**DLPFC**	8.086	4.90E-08	0.007	0.695	2.005
JRK	CMC	DLPFC	0.032	1.25E-07	0.091	0.010	1.010
PTPRU	CMC	DLPFC	-0.077	1.60E-07	0.016	-0.010	0.990
MARCKS	**CMC**	**DLPFC**	0.398	2.05E-07	0.001	0.015	1.015
TCF4	GTEX	Anterior Cingulate Cortex	-0.059	5.22E-13	0.051	-0.013	0.987
DGKD	GTEX	Anterior Cingulate Cortex	-0.937	2.63E-11	0.001	-0.022	0.979
C1QTNF4	GTEX	Anterior Cingulate Cortex	-0.173	1.37E-09	0.010	-0.017	0.983
PITPNA	GTEX	Anterior Cingulate Cortex	-0.243	1.77E-07	0.002	-0.010	0.990
FXR1	GTEX	Caudate Basal Ganglia	0.439	5.40E-12	0.001	0.017	1.017
ZDHHC1	GTEX	Caudate Basal Ganglia	0.354	5.36E-08	0.001	0.011	1.012
PDE4D	GTEX	Cerebellar Hemisphere	0.365	6.81E-11	0.001	0.013	1.013
DRD2	GTEX	Cerebellar Hemisphere	-0.182	2.47E-10	0.004	-0.012	0.988
PITPNM2	GTEX	Cerebellar Hemisphere	-0.065	2.21E-09	0.028	-0.011	0.989
RINT1	GTEX	Cerebellar Hemisphere	0.086	6.32E-09	0.016	0.011	1.011
SRMS	GTEX	Cerebellar Hemisphere	-0.440	3.08E-08	0.001	-0.011	0.989
SETD6	GTEX	Cerebellar Hemisphere	-0.043	1.05E-07	0.054	-0.010	0.990
APOPT1	GTEX	Cortex	-0.074	1.24E-10	0.026	-0.012	0.988
VSIG2	GTEX	Cortex	-0.092	6.01E-09	0.013	-0.011	0.989
SDCCAG8	GTEX	Cortex	-0.069	3.88E-07	0.002	-0.003	0.997
PIK3C2A	**GTEX**	Cortex	-0.040	4.04E-07	0.365	-0.024	0.976
AS3MT	GTEX	Frontal Cortex	0.594	5.65E-17	0.001	0.017	1.017
FOXN2	**GTEX**	**Hippocampus**	-0.250	2.65E-07	0.021	-0.036	0.964
RASIP1	**GTEX**	**Nucleus Accumbens Basal Ganglia**	0.055	3.80E-08	0.034	0.010	1.010
TCF23	**GTEX**	**Nucleus Accumbens Basal Ganglia**	-0.076	4.83E-08	0.019	-0.010	0.990
TTC14	GTEX	**Nucleus Accumbens Basal Ganglia**	-0.089	4.84E-08	0.013	-0.010	0.990
TYW5	GTEX	Putamen Basal Ganglia	-0.080	2.63E-13	0.035	-0.015	0.985

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SNX19	GTEX	Putamen Basal Ganglia	0.031	1.31E-12	0.179	0.013	1.013
CIART	GTEX	Putamen Basal Ganglia	0.090	6.78E-10	0.017	0.012	1.012
SH2D7	GTEX	Putamen Basal Ganglia	0.096	7.89E-09	0.013	0.011	1.011
DGUOK	**GTEX**	Putamen Basal Ganglia	0.255	8.26E-08	0.002	0.011	1.011
C12orf76	GTEX	Putamen Basal Ganglia	0.031	2.27E-07	0.095	0.010	1.010
LRRC37A	**GTEX**	Putamen Basal Ganglia	-0.035	2.69E-07	0.076	-0.010	0.991
AC005841.1	GTEX	Pituitary	0.162	3.28E-09	0.005	0.011	1.011
RPS17	GTEX	Pituitary	0.035	4.03E-08	0.082	0.010	1.010
MHC Region:							
BTN1A1	GTEX	Caudate Basal Ganglia	-0.2606	1.6666E-22			
VARS2	GTEX	Anterior Cingulate Cortex	0.0747019	7.4821E-15			
HIST1H3H	GTEX	Putamen Basal Ganglia	-1.105982	3.2236E-10			
NUDT3	GTEX	Nucleus Accumbens Basal Ganglia	0.10378753	6.546E-09			

bioRxiv preprint doi: https://doi.org/10.1101/222596; this version posted November 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: Significantly enriched pathways and gene sets

Analysis	Gene Set	Comp P	FDR P
Hypothesis driven	FMRP-targets	$1.96x10^{-08}$	$3.097x10^{-06}$
	BP denovo CNV	$7.92x10^{-08}$	$6.257x10^{-06}$
	HIGH LOF intolerant	$5.86x10^{-05}$	0.00309
Agnostic	Increased spleen iron level	$2.72x10^{-08}$	0.000245
	Decreased IgM level	$6.80x10^{-07}$	0.00307
	Condensed chromosome	$1.99x10^{-06}$	0.00598
	Chromosome	$2.80x10^{-06}$	0.00632
	Abnormal spleen iron level	$6.79x10^{-06}$	0.00765
	Mitotic Anaphase	$6.39 x10^{-06}$	0.00765
	Mitotic Metaphase and Anaphase	$5.13 x10^{-06}$	0.00765
	Resolution of Sister Chromatid Cohesion	$5.82 x10^{-06}$	0.00765
	Increased liver iron level	$1.03 x10^{-05}$	0.0103
	Separation of Sister Chromatids	$1.28 x10^{-05}$	0.0115
	Regulation of Rab GTPase activity	$1.78 x10^{-05}$	0.0123
	Regulation of Rab protein signal transduction	$1.78 x10^{-05}$	0.0123
	Protein phosphorylated amino acid binding	$1.75x10^{-05}$	0.0123
	Chromosome	$2.57x10^{-05}$	0.0165
	Hexosaminidase activity	$3.47x10^{-05}$	0.0174
	Abnormal learningmemoryconditioning	$3.11x10^{-05}$	0.0174
	Abnormal liver iron level	$3.47x10^{-05}$	0.0174
	Mitotic Prometaphase	$2.99x10^{-05}$	0.0174
	M Phase	$3.70x10^{-05}$	0.0176
	Positive regulation of Rab GTPase activity	$5.93x10^{-05}$	0.0232
	Rab GTPase activator activity	$5.93x10^{-05}$	0.0232
	Protein phosphatase type 2A regulator activity	$5.24x10^{-05}$	0.0232
	Replicative senescence	$5.44x10^{-05}$	0.0232
	Condensed nuclear chromosome	$7.11x10^{-05}$	0.0267
	Ubiquitin-specific protease activity	0.000104	0.0335
	Ras GTPase activator activity	$9.61x10^{-05}$	0.0335
	Metabolism of porphyrins	0.000103	0.0335
	Kinetochore	0.000103	0.0335
	Decreased physiological sensitivity to xenobiotic	0.000127	0.0381
	Antigen Activates B Cell Receptor Leading to Generation of Second Messengers	0.000124	0.0381
	Phosphoprotein binding	0.000146	0.0424
	Abnormal dorsal-ventral axis patterning	0.000152	0.0429