1 Cell group analysis reveals changes in upper-layer neurons associated 2 with schizophrenia

3 Rujia Dai^{1,2}, Lulu Chen³, Sihan Liu^{1,9}, Chiung-Ting Wu³, Yu Chen¹, Yi Jiang^{1,4},
4 Jiacheng Dai¹, Qihang Wang¹, Richard Kopp², Guoqiang Yu³, Yue Wang^{3,*}, Chao
5 Chen^{1, 5, 6, 7,*}, Chunyu Liu^{1, 2, 8,*}

6

7 ¹Center for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School
8 of Life Sciences, Central South University, Changsha, China

9 ²Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA

10 ³Department of Electrical and Computer Engineering, Virginia Polytechnic Institute
11 and State University, VA, USA

12 ⁴Department of Molecular Physiology and Biophysics, Vanderbilt University,
13 Nashville, TN, USA

14 ⁵National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central
15 South University, Changsha, China

16 ⁶Hunan Key Laboratory of Animal Models for Human Diseases, Central South
17 University, Changsha, China

18 ⁷Hunan Key Laboratory of Molecular Precision Medicine, Central South University,
19 Changsha, China

20 ⁸School of Psychology, Shaanxi Normal University, Xi'an, China

21 ⁹Institute of Rare Disease, West China Hospital of Sichuan University, Chengdu, Chin
22 a

23 * Corresponding author Contact:

24 Chunyu Liu, liuch@upstate.edu

25 Chao Chen, chenchao@sklmg.edu.cn

26 Yue Wang, yuewang@vt.edu

- 27 **Keywords:** schizophrenia; tissue deconvolution; gene expression; neural cell type;
- 28 prefrontal cortex; eQTL; TWAS

29 **Abstract**

30 Genome-wide association studies (GWAS) of schizophrenia (SCZ) have revealed
 31 over 100 risk loci. We investigated whether these SCZ-associated variants regulate
 32 gene expression by cell type. Using a fully unsupervised deconvolution method, we
 33 calculated gene expression by clusters of estimated cell types (cell-groups, CGs).
 34 Five CGs emerged in the dorsolateral prefrontal cortices (DLPFC) of 341 donors with
 35 and without SCZ. By mapping expression quantitative trait loci (eQTL) per CG, we
 36 partitioned the heritability of SCZ risk in GWAS by CGs. CG-specific expressions and
 37 eQTLs were replicated in both a deconvoluted bulk tissue data set with a different
 38 method and also in sorted-cell expression data. Further, we characterized CG-
 39 specific gene differential expression and cell proportion changes in SCZ brains. We
 40 found upper-layer neurons in the DLPFC to be associated with SCZ based on
 41 enrichment of SCZ heritability in eQTLs, disease-related transcriptional signatures,
 42 and decreased cell proportion. Our study suggests that neurons and related
 43 anomalous circuits in the upper layers of the DLPFC may have a major contribution
 44 to SCZ risk.

45

46

47

48

49

50

51

52

53

54 Introduction

55 Schizophrenia (SCZ) is a complex psychiatric disorder, with approximately 80% of
 56 phenotypic variation explained by genetics¹. In recent years, a surge of genetic
 57 markers associated with SCZ has flowed from genome-wide association studies
 58 (GWAS)². Nonetheless, interpreting the biological meaning of these GWAS signals
 59 remains a challenge. SCZ GWAS loci are located largely within non-coding regions
 60 of the genome, pointing to a mechanistic involvement in gene regulation³. Expression
 61 quantitative trait loci (eQTL) mapping provides an effective means for connecting and
 62 interpreting GWAS results with underlying relevant expression regulation⁴. The
 63 PsychENCODE⁵, CommonMind⁶, Brain Cloud⁷, and UK Brain Expression
 64 Consortium⁸ projects produced large-scale eQTL resources. These data have been
 65 integrated with GWAS in the effort to pinpoint risk genes⁹⁻¹¹.

66
 67 A primary pitfall of current brain eQTL analyses, however, is the indifference to cell-
 68 type specificity¹². All existing brain eQTL resources originate from homogenate bulk
 69 tissues in the brain, a mixture of many cell types or subtypes. Bulk tissue eQTL
 70 mapping offers only averaged effects across the composite cell types. It is therefore
 71 confounded by variations in proportion and gene expression. Reports of cell-type
 72 specificity of eQTLs from blood¹³ provide evidence that at least some common
 73 genetic variants shape gene expression by cell types¹⁴. However, the cell-type-
 74 specific eQTLs in the human brain have not yet been reported.

75
 76 To date, the majority of transcriptomic studies of SCZ have similarly relied on data
 77 derived from bulk tissue, using differential analysis and co-expression analyses^{15,16}.
 78 Such studies bear two major limitations: 1) it is unclear whether genes that are
 79 differentially expressed between individuals with SCZ and controls are attributed to
 80 differences in expression or changes in cell proportion, and 2) it is unknown whether

81 changes in SCZ-related expression occur only within one specific cell type or multiple
82 cell types.

83

84 Researchers investigate cell-type-specific expression using either of two approaches:
85 1) sequencing individual or sorted cells using single-cell or single-nuclei sequencing
86 (scRNA-seq or snRNA-seq), or 2) the computational deconvolution of bulk tissue
87 transcriptomic data. To date, fewer than 50 human brain samples have been
88 sequenced and published using scRNA-seq or snRNA-seq^{17,18}. For the time being,
89 such methods are impractical for most genetic or case-control studies due to the
90 large sample size and associated costs required to achieve adequate statistical
91 power.¹⁹ Alternatively, the computational approach can deconvolute bulk tissue data
92 into cell-type-specific data with sufficient accuracy²⁰.

93

94 The computational deconvolution of bulk tissue data can estimate cell type proportion
95 and cell type-specific expression in either a supervised or an unsupervised manner.
96 Supervised deconvolution relies on an accurate reference, for example, cell type-
97 specific expression from the same species and brain region. However, suitable
98 references are often unavailable, and whether references developed from control
99 samples can be used to deconvolute patient data is uncertain. In contrast, using
100 unsupervised deconvolution sidesteps this problem by estimating cell-type-specific
101 expression and cell proportion without prior information. Unsupervised deconvolution
102 classifies a group of cells algorithmically to differentiate them from other groups.
103 Since these groups may not always reflect biologically homogeneous cell types, we
104 instead use the term cell groups (CGs). The resulting CGs represent an estimated
105 cluster of cell sub-types extractable from bulk tissue. Using unsupervised
106 deconvolution methods, like convex analysis of mixture (CAM),^{21,22} identifies CGs
107 based on gene expression patterns from all tested individuals.

108

109 Some advanced deconvolution methods can *infer* cell-type-specific expression per
110 individual, based upon the methylation²³ or gene expression²⁴ profile from the donor's
111 bulk tissue. Again, such supervised methods rely on suitable references. Ideally,
112 sample-wise and unsupervised deconvolution methods are needed to investigate
113 CG-specific expressions from bulk tissue data in a cost-effective manner.

114

115 We hypothesized that GWAS variants associated with the risk for SCZ regulate gene
116 expression within one or more specific brain cell types. To identify CG-specific
117 expressions at the sample level, we applied the newly-developed method^{25,26}
118 sample-wise convex analysis of mixtures (swCAM) to RNA-seq data from 93
119 postmortem brain samples of donors with SCZ and 248 healthy donors
120 (Supplemental fig.1). An independent postmortem brain dataset of 605 samples was
121 used to replicate the swCAM deconvolution results. Further, we generated CG-
122 specific eQTLs and tested whether the CG-specific eQTLs were enriched for SCZ
123 GWAS signals. To detect SCZ-related transcriptional changes per CG, we performed
124 differential expression analyses, followed by co-expression network and
125 transcriptome-wide association analyses of the deconvoluted expression data. We
126 integrated results from GWAS enrichment, cell proportion, and transcriptomic
127 changes to target SCZ-associated CGs. The processed data and results are
128 accessible on our website (http://lbpq.upstate.edu/module_search/).

129

130 **Results**

131 ***Sample-wise deconvolution of RNA-seq data from bulk brain tissue of donors*** 132 ***with and without SCZ***

133 Five CGs emerged from swCAM deconvolution of bulk tissue data. Their identities
134 were first annotated by enrichment testing of known cell-type marker genes. We

135 calculated the enrichment of marker genes of major brain cell types²⁷ (Supplemental
136 table 1) in the top cell-type differentially expressed genes (ctDEG) for each CG. Top
137 ctDEGs were selected by differential expression analysis of target CG versus the
138 remainder of CGs ($\log_2FC > 2$, False Discovery Rate [FDR] $q < 0.05$, Wilcoxon signed-
139 rank test). The five annotated CGs represented astrocytes, two neuronal groups, a
140 mixture of astrocytes, microglia, and endothelial cells, and oligodendrocytes (p value
141 < 0.05 , Fisher's exact test, Fig. 1A, Supplemental Fig. 2). We also used Expression
142 Weighted Cell Type Enrichment (EWCE) test as a replication for annotating CGs. We
143 applied EWCE to two snRNA-seq data sets^{17,28} from the human frontal cortex and
144 tested to determine if ctDEGs are expressed more highly in a given cell type than
145 that by chance. The EWCE results were consistent with previous CG annotations
146 except for CG1 (Fig. 1B). Because CG1 was enriched for two glial cell subtypes in
147 the EWCE test of two snRNA-seq data sets, we defined CG1 as a mixture of
148 astrocyte and endothelial cells.

149

150 To test the expression similarity between estimated CGs and corresponding cell
151 types, we performed a Spearman correlation test between CG expressions and
152 snRNA-seq data from the human frontal cortex²⁸. We corrected the batch effects
153 between the tested data prior to the correlation test. The three CGs that only
154 enriched for single cell types showed high correlations with corresponding cell types
155 (averaged $\rho_{\text{neuron_CG2}} = 0.83$, $\rho_{\text{neuron_CG3}} = 0.81$, $\rho_{\text{oligodendrocyte_CG5}} = 0.86$, all three p
156 values < 0.05 , Supplemental Fig. 3) The averaged correlation coefficient for two glial
157 mixture CGs was 0.55 (SD=0.15).

158

159 To further differentiate the two neuronal clusters, marker gene-sets of standard
160 neuronal subtypes were applied. We initially suspected that CG2 and CG3
161 represented inhibitory and excitatory neurons but were dissuaded because there was
162 not convincing statistical differentiation provided by marker genes of these neuron

163 types (Fisher's exact test $p > 0.05$, Supplemental Fig. 4). We then speculated that
 164 CG2 and CG3 may reflect the origin of specific cortical layers. With only two detected
 165 neuronal CGs and six cortical layers, we took a quantitative approach in allocating
 166 gene expression to gradient cortical layers: 770 *upper-layer (upperL) genes* were
 167 expressed most abundantly in the uppermost layer (layer 1) steadily decreasing by
 168 layer with the least expression in the layer 6, while 875 *deep-layer (deepL) genes*
 169 were distributed in the opposite direction ($FDR < 0.05$, linear regression). *UpperL* and
 170 *deepL genes* corresponded to the CG2 and CG3 clusters respectively (Fisher's exact
 171 test, $p_{upperL} < 2.2e-16$, $p_{deepL} = 7.90e-10$, Fig. 1C).

172
 173 The expression data of manually isolated neurons from each of the six layers of the
 174 middle temporal cortex²⁹ was used to validate the identity of neuronal CGs. We
 175 calculated the CG-specific difference for each gene by comparing the expression of
 176 CG2 and CG3 individually to that of the other CGs. Layer-specific differences were
 177 also calculated for each gene by comparing the neuronal expression of one specific
 178 layer to that of the other layers. CG-specific differences and layer-specific differences
 179 were then tested for correlation. CG3-specific differences were highly correlated to
 180 layer-specific differences of neurons from layers 1 through 3, while CG3-specific
 181 differences were poorly correlated to layer-specific differences of neurons from layers
 182 4 through 6. In contrast, correlations for CG2-specific differences ran in the opposite
 183 direction (Supplemental Fig. 5). We also performed an EWCE test on this snRNA-
 184 seq data. The EWCE results showed CG2 enriched for neurons from layer 4
 185 ($FDR < 0.05$) and layer 5 ($FDR < 0.05$) and CG3 enriched for neurons from layer 2
 186 ($FDR = 0.006$), layer 3 ($p = 0.04$, $FDR = 0.40$), and layer 4 ($FDR = 0.002$, Fig. 1D). Based
 187 on these tests, we defined CG2 as representing deep-layer neurons (deepN) and
 188 CG3 as upper-layer neurons (upperN). At this juncture, the gene expression profiles
 189 for the five CGs were now garnered from the bulk brain tissue of donors with and
 190 without SCZ, i.e., upperN, deepN, a mixture of astrocytes and endothelial cells

191 (ast_endo), oligodendrocytes (oli), and a mixture CG of astrocytes, microglia, and
192 endothelial cells (mic_ast_endo) (Fig. 1E).

193

194 Results of this swCAM deconvolution analysis were replicated with another sample-
195 wise deconvolution method, bMIND³⁰. We applied bMIND to our tested data and
196 identified expressions of five CGs with similar composition to those identified by
197 swCAM. We used the Spearman correlation test to examine the similarity between
198 CG-specific expressions estimated by swCAM and bMIND. The correlation
199 coefficients were 0.85 ± 0.03 ($FDR < 0.05$) across corresponding CGs (Supplemental
200 Fig. 6A). The showed that our deconvolution results were robust to different methods.

201

202 swCAM deconvolution results were replicated on a second RNA-seq dataset of 605
203 brain samples³¹. We detected five CGs in the replication dataset. (Supplemental Fig.
204 7). Through marker gene enrichment tests and EWCE test, we found the five CGs in
205 these data were upperN (FDR of layer 2 neuron < 0.05 , p value of layer 1 neuron =
206 0.03), deepN (FDR of layer 4 neuron and layer 5 neuron < 0.05), astrocyte
207 (FDR < 0.05), a mixture of microglia and endothelial cells (FDR < 0.05) and a CG
208 showing oligodendrocyte expression with non-significant p value ($p = 0.14$). The
209 dominant CGs such as upperN and deepN CGs were consistently identified.

210

211 **Generation of eQTLs for CGs**

212 eQTL mapping was used to examine the genetic regulation of CG-specific
213 expression. We included 72 additional brain samples from individuals with bipolar
214 disorder (BD) to increase the sample size. Since we regressed out the affection
215 status, the inclusion of BD patients did not interfere with the eQTL mapping results.
216 The same five CGs were still detected by swCAM deconvolution after adding the BD

217 samples. After removing known and hidden covariates in the CG-specific
218 transcriptome, we focused on local eQTLs (cis-eQTLs), searching for any
219 associations between individual expression of genes with common variants within ± 1
220 Mb of the gene body region. In total, we identified 1,878,815 genome-wide significant
221 eQTLs (FDR $q < 0.05$). After subtracting shared eQTL across all CGs, we observed
222 more eQTL signals in neuronal CGs (both of upperN and deepN) than in glial CGs
223 (Fig. 2A). The number of single nucleotide polymorphisms of eQTL (eSNPs) of
224 neuronal CGs ($n=302,608$) was more than that of eSNPs of oligodendrocyte CGs
225 ($n=80,229$). No eQTL survived multiple testing corrections in mixture CGs of
226 astrocyte and endothelial cell data. We annotated the genomic regions for eSNPs
227 with Ensembl annotations using ANNOVAR software³². The majority of eSNPs were
228 located in intergenic and intronic regions within non-coding RNAs (Fig. 2C).

229

230 To examine the replication rate of eQTLs of the three CGs enriched for singular cell
231 types (deepN, upperN, and oli), we compared our results to a sorted-cell study³³,
232 which profiled the gene expression and genotype of neurons ($n = 42$) and
233 oligodendrocytes ($n = 36$). Using the same eQTL mapping pipeline described earlier,
234 we calculated the eQTLs using these replication data. From this small sorted-cell
235 dataset, we could identify 644 eQTLs for neurons and 190 eQTLs for
236 oligodendrocytes (FDR $q < 0.05$). We used π_1 from Storey's method³⁴ to calculate the
237 proportion of replicated eQTLs. We applied this method to the nominal p-values of
238 replicated eQTLs in sorted cell populations, uncorrected for multiple comparisons.
239 We found that upperN ($\pi_1=0.56$) and deepN ($\pi_1=0.55$) had greater reproducibility
240 than oli ($\pi_1=0.35$). Because π_1 is the estimated proportion of the true alternative
241 hypothesis, it has loose criteria for power and p-values. We used replication rate to
242 quantify the replications, which was calculated by the ratio of replicated eQTLs
243 among eQTLs in sorted populations with the criteria FDR <0.05 . We found 29% of the
244 neuron eQTLs were specifically replicated in neuronal CGs and 15% of the

245 oligodendrocyte eQTLs were replicated in all CGs (Supplemental Fig. 8).

246

247 We compared our eQTL results with CG-specific eQTLs identified by bMIND. We
248 found our results were well replicated by bMIND, with the $\pi_1 > 0.98$ for all CGs
249 (Supplemental Fig. 6B). We also compared our results with eQTLs from bulk brain
250 tissue samples from BrainGVEX. CG-specific eQTLs showed high reproducibility in
251 the BrainGVEX dataset ($\pi_1 > 0.97$ in four CGs, $\pi_1 = 0.92$ in mic_ast_endo, Fig. 2B).

252

253 ***Neuronal CGs and a mixture CG containing astrocyte and endothelial cells***
254 ***explain the risk heritability of SCZ***

255 To associate brain cell type with SCZ, we used CG-specific eQTLs to assess the
256 SCZ GWAS risk heritability. We hypothesized that if a CG is associated with SCZ,
257 their eQTLs will explain greater GWAS heritability. To test this hypothesis, we applied
258 stratified linkage disequilibrium (LD) score regression³⁵ (sLDSC) on GWAS summary
259 statistics and CG-specific eQTLs to partition risk heritability from summary statistics.
260 The SNP-based heritability analysis showed neuronal CGs and a CG mixture
261 containing astrocyte and endothelial cells significantly enriched for risk heritability of
262 SCZ (Fig. 3A). The upper-layer neuron CG explained 19.24% of heritability with 14%
263 SNPs (adjusted p value=2.97e-04, Fig. 3B). The deep-layer neuron CG explained
264 19.03% of heritability with 14% SNPs (adjusted p value=4.81e-04). Mixture CG of
265 astrocyte and endothelial cells explained 12% of heritability with 8% of SNPs
266 (adjusted p value=1.04e-03). These results suggest the regulatory regions in the
267 neuronal CGs and a mixture CG of astrocyte and endothelial cells have a higher
268 proportion of SCZ common variants than that in oligodendrocyte CG and a CG
269 mixture containing microglia, astrocyte, and endothelial cells. We also compared
270 enrichment score in bulk tissue eQTLs and CG-specific eQTLs. We found the

271 enrichment scores at bulk level (1.29 ± 0.07) and cell levels (1.42 ± 0.17) were very
272 close (Fig. 3A).

273

274 ***Decreased proportion of upper-layer neuron CGs from donors with SCZ***

275 To determine if the proportion of each CG was altered in the brains of patients with
276 SCZ, we estimated the proportion of CGs for each sample using CAM. The
277 proportion of upperN was reduced by 2.76% in patients with SCZ compared with
278 controls, and the mixture CG of mic_ast_endo increased by 0.8% in patients with
279 SCZ (Wilcoxon signed-rank test, p value < 0.05 , Fig. 4A). Other CGs did not show
280 significant change. The proportional change in upper-layer neurons was statistically
281 significant in a permutation test ($p < 0.001$, Supplemental Fig. 9).

282

283 Based on our testing for genetic enrichment, we hypothesized that the changes in
284 CG proportion could arise from genetic risk variants. We used two strategies to test
285 this hypothesis. First, we performed QTL mapping using CG proportions as the
286 phenotype, but we did not detect any significant CG proportion-QTL signals. Second,
287 we used polygenic risk score (PRS) as a measure of genetic risk burden for each
288 individual and tested the correlation between PRS and CG proportions. We observed
289 weak negative correlations between SCZ PRS and the upperN proportion and the p
290 value did not survive multiple testing ($r = -0.1$, p value $= 0.04$, FDR $= 0.1$, Fig. 4B).

291

292 ***CG-specific differential gene expression in SCZ***

293 We further analyzed CG-specific differential expression comparing individuals with
294 and without SCZ using Wilcoxon signed-rank test with significance by FDR $q < 0.05$,
295 (Supplemental Table 2). We compared our results with differential expression results

296 from bMIND and observed significant consistency (Fisher's exact test p value < 0.05 ,
 297 Supplemental Fig. 6C). We found substantial overlap in differentially expressed
 298 genes (DEG) across five CGs, with 330 upregulated genes and 344 downregulated
 299 genes shared across five CGs (Supplemental Fig. 10). The number of shared DEGs
 300 across all CGs was greater than the number of CG-specific DEGs, except for
 301 upregulated genes in CG mic_ast_endo, with 367 genes. Approximately 33% of the
 302 genes ($n=14,865$, Gencode v19) were affected in at least one CG in the SCZ cohort.
 303 The two neuronal CGs possessed greater transcriptional changes than the three glial
 304 CGs, (fold change: 0.20 in neuronal CGs versus 0.09 in glial related CGs, p
 305 value $< 2.2e-16$, Kolmogorov–Smirnov test). We performed permutation tests to
 306 determine the empirical probability of transcriptional changes in CGs. We found that
 307 the fold changes observed in all CGs were significant ($p < 0.001$). Larger fold changes
 308 in neuronal CGs than that in the glia CG were also significant in permutations
 309 ($p < 0.001$, Supplemental Fig. 11).

310

311 We used Gene Ontology (GO) enrichment to annotate the function of DEGs (Table
 312 1), focusing on the functions of dysregulated genes in neuronal groups. A total of
 313 1,537 downregulated DEGs were detected in the upperN. These were enriched for
 314 protein folding (FDR $q = 3.20e-09$) and antigen processing (FDR $q = 8.01e-05$). The
 315 1,407 downregulated DEGs within deepN were enriched for positive regulation of cell
 316 proliferation (FDR $q = 2.24e-06$) and trans-synaptic signaling (FDR $q = 4.10e-06$).
 317 Upregulated genes in both the upperN ($n=229$) and deepN ($n=223$) GCs were
 318 enriched for carboxylic acid breakdown (FDR $q_{\text{deepN}} = 1.04e-06$, FDR $q_{\text{upperN}} = 6.93e-$
 319 08).

320

321 ***Transcriptome-wide association analysis identified SCZ-associated genes in all*** 322 ***CGs***

323 To determine which genes had CG-specific dysregulated expression driven by
324 genetics, we performed a TWAS analysis using the CG-specific eQTLs. We identified
325 51 SCZ risk genes across 43 genetic loci (adjusted p value<0.05, Supplemental
326 Table 3). We compared our TWAS results to the 193 SCZ risk genes identified in bulk
327 tissue by TWAS in Gandal et al¹⁶. Twelve genes in our results had been previously
328 identified in Gandal's results. Our analysis attributed these 12 genes to all 5 CGs:
329 deepN (*CYP17A1-AS1*, *AS3MT*, *CSPG4P11*, *EMB*), upperN (*CYP17A1-AS1*, *AS3MT*,
330 *CSPG4P11*, *EMB*, *GATAD2A*, *NAGA*, *RP11-73M18.8*, *TYW5*), ast_endo (*RP11-*
331 *73M18.8*, *CYP17A1-AS1*, *EMB*), oli (*SNX19*, *AC011330.5*, *CSPG4P11*, *EMB*), and
332 mic_ast_endo (*BTN3A2*, *TOM1L2*, *CSPG4P11*). Additionally, 39 CG-specific TWAS
333 genes are novel to bulk tissue-based TWAS.

334
335 Out of the 51 TWAS genes, 17 genes are in deepN and 30 in upperN. Eight genes
336 existed in both subtypes (*CYP17A1-AS1*, *AS3MT*, *PIGB*, *GOLGA6L9*, *CSPG4P11*,
337 *CNPPD1*, *EMB*, *ENDOG*). Twenty-two genes were uniquely identified in upperN
338 (*DIABLO*, *C12orf65*, *RP11-73M18.8*, *STRC*, *AP3B2*, *WHAMM*, *SLX1B*, *ARL17B*,
339 *GATAD2A*, *RPS5*, *ADAM15*, *RP1-130H16.16*, *WBP2NL*, *NAGA*, *ALMS1*, *TYW5*,
340 *ITIH4*, *TMEM110*, *SFMBT1*, *NDUFAF2*, *BTN2A1*, *DNPH1*). Nine genes were
341 identified in the deepN only (*CWF19L2*, *MORN3*, *TIE1*, *RBBP5*, *CBR3*, *RFTN2*,
342 *RP11-10L12.4*, *REEP2*, *STK4*). Twenty-four TWAS genes are specific to glial CGs (9
343 from ast_endo, 10 from oli and 5 from mic_ast_endo).

344

345 ***Co-expression network analysis refined CG-specific biological processes*** 346 ***associated with SCZ***

347 To investigate the coordinated expression changes in each CG, we performed
348 weighted gene correlation network analysis³⁶ (WGCNA) on CG-specific expressions
349 (Supplemental Table 4). For this study, we identified *CG-specific modules*, a group of
350 genes showing concordant variation across samples in one specific CG. Modules of
351 upperN and deepN were highly consistent. In total, we identified eight CG-specific
352 modules out of the 149 detected modules. Seven of the eight showed either
353 differential expression of eigengenes in SCZ or enrichment of SCZ GWAS signals.
354 These seven modules were defined as “SCZ-associated modules” (Table 2).

355
356 The neuronal CGs shared five of the seven SCZ-associated modules (Fig. 5). The
357 eigengenes of two modules showed upregulated expression in SCZ: neuron_ME28
358 ($\log_2FC=0.02$, $FDR=6.8e-03$) and neuron_ME30 ($\log_2FC=0.03$, $FDR=6.3e-07$).
359 These two were enriched in the pathways of neural stem cell maintenance and
360 carbohydrate transmembrane transport ($FDR\ q<0.05$). The other three modules
361 showed only enrichment of GWAS signals (neuron_ME32, neuron_ME27, and
362 neuron_ME29) and were enriched within pathways of regulation of actin filament
363 polymerization, regulation of dendritic spine morphogenesis, and SWI/SNF
364 superfamily-type complex, respectively ($FDR\ q<0.05$).

365
366 The other two SCZ-associated modules were specific to glial CGs. In the mixed CG
367 of astrocytes and endothelial cells, we identified ast_endo_ME28, enriched for SCZ
368 GWAS signals, and involved with the RNA splicing pathway ($FDR\ q<0.05$). In the
369 oligodendrocyte CG-related module (oli_ME14), the eigengene was downregulated in
370 SCZ ($\log_2FC=-0.01$, $FDR=0.04$). Member genes in this oli_ME14 were involved in
371 neuron ensheathment ($FDR\ q<0.05$).

372

373 Discussion

374 Combining CG-specific eQTL analysis and GWAS can reveal *how* (the regulation of
375 gene expression) and *where* (within which CGs) variants contribute to disease risk.
376 Recent studies have employed this approach identifying cell-type-specific eQTLs in
377 blood that have heightened risk heritability for immune-related diseases^{37,38}. However,
378 cell-type-specific eQTLs for human brains are not available and are difficult to obtain.
379 To bridge genetic findings of SCZ to brain cell types, we generated CG, a proxy of
380 cell type, using an unsupervised deconvolution method. We then mapped CG-
381 specific eQTLs using computationally partitioned gene expression specific to each
382 CG per individual. Based on GWAS data, we found enrichment for risk heritability in
383 the eQTLs of upper-layer DLPFC neuronal CGs (upperN). Furthermore, we found
384 that the upperN CGs also displayed dramatic gene expression changes and altered
385 cell proportions in the individuals with SCZ.

386

387 Our results are consistent with previous bulk tissue studies of human and cell-type-
388 specific studies of mice in implicating neurons in the risk for SCZ. In a case-control
389 study of bulk brain tissues, Gandal et al. showed the neuron-specific co-expression
390 module to be downregulated in SCZ¹⁶. Skene et al. calculated the neuronal specificity
391 of the human-homologous gene in mouse scRNA-seq data and correlated this
392 specificity to risk heritability enrichment within the genetic region³⁹. They found that
393 four neuron subtypes in the mouse brain (i.e., medium spiny neurons, pyramidal cells
394 in hippocampal CA1, pyramidal cells in the somatosensory cortex, and cortical
395 interneurons) are associated with SCZ. Our study expands upon those findings,
396 confirming that a subtype of neurons, the group of upper-layer neurons in human
397 DLPFC, are associated with SCZ.

398

399 Our study provides genetic support for an association between the upper-layer
400 neurons of the DLPFC and SCZ. This is supported by the enrichment of GWAS
401 signals, dysregulated gene expression, and decreased proportions within the upperN
402 CGs. Previous research has shown upper-layer neurons of the DLPFC are over-
403 represented in humans compared with other species,⁴⁰ and to be actively engaged in
404 the delayed activity of working memory⁴¹, the impairment of which is a major feature
405 of SCZ. Some phenotypes of upper-layer neurons within the DLPFC have been
406 reported to be associated with SCZ in terms of cell shape⁴², dendritic
407 abnormalities^{43,44}, and candidate mRNA expression⁴⁵. However, no genetic evidence
408 that we are aware of to date has pinpointed upper-layer neurons as having a role in
409 SCZ risk. In this study, we observed that GWAS risk variants contribute to SCZ by
410 altering gene expression within the upper-layer neurons. Our coexpression network
411 analysis and TWAS analysis further narrowed down specific genes and pathways.
412 Therefore, we suspect that previously reported phenotypic changes⁴²⁻⁴⁶ within the
413 upper-layer neurons of patients with SCZ may be rooted in these genetic variants
414 and their related gene expression regulation.

415

416 Our findings of SCZ-associated changes in upper-layer neurons were validated by
417 multiple lines of experiments from other studies. In a recent snRNA-seq study of
418 cortical layers from seven SCZ patients and 11 controls, the selective vulnerability of
419 supragranular layer neurons was revealed in SCZ in terms of dramatic transcriptomic
420 changes and cell proportion changes⁴⁷. A decreased proportion of upper-layer neuron
421 subtypes was also observed in mouse model SCZ-predisposing 22q11.2
422 microdeletion⁴⁸ and human postmortem brains^{49,50}. Moreover, the functional
423 mechanism of a differential expressed gene in upper-layer neuron CYFIP1 (log2FC=
424 0.13, FDR=0.0004) has been discovered in the mouse cortex. It has been reported
425 that CYFIP1 deficiency causes upper-layer neurons to present in deep layers and

426 deep-layer neurons to present in upper layers⁵¹. Moreover, the changes in upper-
 427 layer neurons also have been noted in association with another psychiatric disorder,
 428 autism, based on snRNA-seq study¹⁸. In summary, this study strengthened the
 429 findings of upper-layer neuron dysfunction in SCZ and provided more detailed cell-
 430 subtype evidence based on a relatively large brain collection. It provides strong
 431 support to previous studies of animal models, cell models, and brain transcriptome
 432 studies with limited samples.

433

434 Our findings suggest the SCZ-associated changes in upper-layer neurons may occur
 435 in neuron development. Neuron development in human brain includes proliferation,
 436 differentiation, neuron migration and synaptic formation. This process is coupled with
 437 substantial transcriptome changes⁵². We found that the DEGs in upperN were
 438 enriched in the pathways of cell proliferation and trans-synaptic signaling, which has
 439 previously been suggested as involved in SCZ risk. For example, the disrupted
 440 proliferation of neurons has been reported in primate and murine models of SCZ⁵³.
 441 Five TWAS genes in upperN are related to neuronal development as well: *AS3MT*,
 442 *CSPG4P11*, *ADAM15*, *STRC*, and *WHAMM*. For example, *AS3MT* participates in
 443 neural stem cell differentiation and is a risk gene for SCZ⁵⁴. *CSPG4P11* is also an
 444 eQTL gene in fetal brains, and its expression has also been linked to SCZ risk^{55,56}.
 445 *ADAM15*, *STRC*, and *WHAMM* function in cell-matrix adhesion, which is linked to
 446 neurodevelopment and SCZ^{57,58}. Therefore, our findings support the hypothesis that
 447 the disturbed gene expression in neurodevelopment is involved in SCZ molecular
 448 pathology. While our study cannot provide causality of these genes in SCZ pathology,
 449 future functional experiments in cells and useful animal models could potentially
 450 confirm the functional mechanism of genes.

451

452 While swCAM yielded reproducible unsupervised deconvolution results, particularly
 453 for neuronal CGs, it also has its limitations. The CGs estimated by swCAM represent

groupings of cells with similar transcriptional profiles. Correspondingly, swCAM has relatively poor resolution in separating rare cell types. Glial subtypes with small compositions, such as microglia, have been particularly challenging. As single-cell data are accumulating, a hybrid model integrating a suitable reference may boost swCAM's resolution. Despite these constraints, swCAM enhances the merit of existing bulk-tissue data by deriving expression for major CGs. With this method, it is possible to determine whether disease-related gene expression is a feature of specific CGs or a pattern of composite CGs. Furthermore, gene networks and genetic regulatory relationships can emerge that may help to ferret out hidden clues to the molecular mechanisms of complex psychiatric disorders.

Conclusion

Of five clusters of cell types estimated by gene expression *in silico*, the upper-layer neurons within the DLPFC display distinctive enrichment of risk heritability and transcriptional changes in the brain tissue of donors with SCZ. Furthermore, decreased proportion of upper-layer neurons relative to all cell groups was observed within the SCZ brain. The uniquely dysregulated genes and pathways associated with upper-layer neurons, particularly those related to neurodevelopment, should be targeted in future experimental studies and quantitative analysis of SCZ to further understand the molecular mechanism of SCZ.

Materials and methods

Brain tissues

The PsychENCODE/BrainGVEX project has 427 postmortem DLPFC samples, including healthy controls, BD and SCZ patients. Brain samples originated from the

Stanley Medical Research Institute (SMRI) and the Banner Sun Health Research
Institute (BSHRI). Detailed information can be found at
<https://www.synapse.org/#!/Synapse:syn4590909>.

RNA sequencing

Total RNA was isolated from brain tissue at the University of Illinois at Chicago and at
the University of Chicago with the Qiagen miRNeasy mini kit. Quality of RNA samples
was measured with an Agilent Bioanalyzer RNA 6000 Nano assay kit and all samples
had a RIN score of 5.5 or greater. All total RNAs were processed into stranded,
rRNA-depleted libraries for sequencing in an Illumina HiSeq2000. Libraries were
triplexed per lane to reach 40 million paired-end reads per library. Details of library
preparation and RNA sequencing can be found in our previous study¹⁶.

Data processing

Data processing included three steps: reads processing, quantification, and post-
quantification processing. Fastq files underwent adapter removal using cutadapt
(<https://pypi.python.org/pypi/cutadapt>), after which the resulting adapter-trimmed
fastq files were checked for quality using FastQC
(<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). A subset of 10,000 reads
estimated the insert mean size and standard deviation for use with Tophat. STAR
was used to align trimmed reads to the GENCODEv19 reference (modified to include
artificial ERCC RNA ExFold spike-in sequences). BAM files were sorted in samtools
(v1.3). Expression level was then calculated using RSEM (v1.2.29). Quality control
metrics were calculated using RNA-SeQC (v1.1.8), featureCounts (v1.5.1),
PicardTools (v1.128), and Samtools (v1.3.1). The quantified fragments per kilobase

of transcript per million mapped reads (FPKM) were used for downstream analysis. Mitochondrial genes, pseudoautosomal genes, and genes with FPKM fewer than 1 in more than 75% of samples were dropped. We calculated the distance between samples and removed samples with a z-score of normalized connectivity to other samples lower than -2. After filtering, 341 samples and 14865 genes were retained for subsequent analyses. Linear regression removed the effect of covariates including age, sex, RIN, PMI, brain bank, batches, and principal components of sequencing statistics (seqPC). The seqPCs was composed of the top 29 principal components analyzed from the sequencing statistics. Covariates were selected by multivariate adaptive regression splines (MARs).

513

514 ***Genotype processing***

Genotypes were called from three platforms, including 277 samples by whole genome sequencing, 257 samples by Psych v1.1 beadchips, and 137 samples by Affymetrix 5.0 450K. Genotypes were imputed by Michigan Imputation Server using HRC (r1.1 2016) EUR samples as references from each platform. Variants with R squared less than 0.3 or HWE (Hardy-Weinberg Equilibrium) p-value less than 1e-6 were removed. We used DRAMS software⁵⁹ to correct mixed-up data IDs for the three platforms based on relationships across various omics data. We had 251 samples genotyped by whole genome sequencing and beadchip. The genotypes called from the two platforms were highly consistent ($r = 0.986$). Genotypes of the three platforms were combined, with discordant genotypes marked as missing. Polygenic risk scores were calculated using PRSice (2.1.4) with GWAS summary data from Ripke et al² as base data set.

527 **Sample-specific deconvolution**

528 To isolate CGs *in-silico*, we developed a novel sample-wise deconvolution technique
 529 (swCAM²⁶), applying it to the processed RNA-seq data. The original CAM framed the
 530 deconvolution of mixture expression to solve the blind source separation problem
 531 (sBSS), which was accomplished by nuclear norm regularization. The basic
 532 assumption in the deconvolution was

$$533 \quad X=AS+E \quad (1)$$

534 which is the mixture sample being the weighted sum of the CG-specific expression S.

535 In the classic deconvolution method, S is a common matrix for all mixture samples.

536 However, each sample may have its sample-specific S_i :

$$537 \quad S_i = \bar{S} + \Delta S_i \quad (i=1, \dots, M) \quad (2)$$

538 The associated sample-specific BSS model is given by

$$539 \quad x_i = a_i (\bar{S} + \Delta S_i) + e_i \quad (3)$$

540 If some samples have a group pattern in a particular source or some molecules'
 541 expression are more highly correlated in one particular source, the associated
 542 samples may share a similar pattern of ΔS_i , leading to the following swCAM objective
 543 function:

$$544 \quad \min \frac{1}{2} \sum_{i=1}^M \|x_i - a_i(\bar{S} + \Delta S_i)\|_2^2 + \lambda \sum_{k=1}^K \|T_k\|_* \quad (4)$$

545 Where T_k consists of the kth column in all ΔS_i , representing between-sample
 546 variability in source K; $\|T_k\|_*$ is the nuclear norm of T_k . λ is the regularization
 547 parameter of nuclear norms. We solved equation (4) using quadratic programming
 548 and the alternating direction method of multipliers (ADMM). Code of swCAM can be
 549 accessed at <https://github.com/Lululuella/swCAM>.

550 We used CAM to estimate the CG proportion in an unsupervised manner. The
 551 processed expression data was the input of CAM. The minimum description length
 552 (MDL) algorithm was used in CAM²¹ to automatically determine the number of CGs
 553 (K). MDL is a widely-adopted and consistent information theoretic criterion in model

554 selection. The shortest MDL denotes the best model. Shortest MDL was shown when
555 K=5 for the data. The proportions of five CGs and processed RNA-seq data were
556 entered into swCAM to infer the CG-specific expression for each sample. To obtain
557 the top ctDEGs and thereby differentiate the CGs, Wilcoxon signed-rank test
558 determined the difference of expression in each CG ($\log_2FC > 2$, $FDR < 0.05$).

559

560 ***Processing of single-cell datasets***

561 Three single-cell datasets provided further confirmation of CG identities^{28,29,60}. We
562 downloaded raw counts from frontal cortices and used the Seurat package (v2.0) for
563 data processing⁶¹. We filtered out the following: genes expressing in less than half of
564 the smallest CGs, cells with unique feature counts (i.e., $> 5,000$ and < 200), and cells
565 with mitochondrial counts $> 5\%$. The data were then normalized by library size and
566 log-transformed. The cell identities from the original studies were used for cell
567 clustering. The averaged expression was calculated to represent gene expressions
568 for the specific cell type.

569

570 ***Processing the ROSMAP dataset for replication***

571 The synapse website (10.7303/syn3388564) provided a replication dataset with 640
572 postmortem human DLPFCs from the Religious Orders Study and Memory and Aging
573 Project (ROSMAP)³¹. Gene expression was measured by RNA-seq. The downloaded
574 raw expression matrix was quantified in FPKM. After quality control, the data retained
575 19,144 genes and 605 samples. Expression was $\log_2$ -transformed and normalized
576 by quantile normalization. The confounder effects from age, sex, diagnosis,
577 sequencing batch, study, ancestry, education year, and PMI were removed by linear
578 regression.

579

580 ***Cell group identification***

581 We used two strategies to annotate the CGs; gene sets enrichment and correlation
 582 with scRNA-seq data. In the identification of five general CGs, the marker genes of
 583 five major cell types were collected from Zhang et al²⁷ and the marker genes for
 584 inhibitory neurons and excitatory neurons were collected from Lake et al⁶³. Fisher's
 585 exact test was used to test the enrichment of marker genes in CGs. Meanwhile,
 586 Spearman correlation testing with scRNA-seq datasets^{28,60} was used to measure
 587 similarities between CG expression and that of known cell types. To identify genes
 588 expressed in gradient cortical layers, we downloaded RNA-seq data across human
 589 DLPFC layers from the He et al. study⁶⁴. Gene expression was set as a linear
 590 regression function for DLPFC layers (from layer 1 to layer 6). Genes with coefficient
 591 >0 and FDR <0.05 were named "deep-layer genes." Those with coefficient <0 and
 592 FDR <0.05 were named "upper-layer genes". scRNA-seq data used for correlation
 593 testing with neuronal CGs was downloaded from [https://portal.brain-map.org/atlas-](https://portal.brain-map.org/atlas-and-data/rnaseq)
 594 [and-data/rnaseq](https://portal.brain-map.org/atlas-and-data/rnaseq), which provided neuronal expression for the multiple layers of
 595 human middle temporal gyri²⁹. Spearman correlation test was performed on layered
 596 neuronal expression and that of CG2 and CG3.

597

598 Expression Weighted Cell Type Enrichment⁶⁵ (EWCE) test was used as a replication
 599 of CG annotations. The purpose of EWCE is to determine the chance of a target
 600 gene list having higher expression in a specific cell type than a random gene list. The
 601 ctDEGs identified in each CG were used as target lists. Two snRNA-seq data sets
 602 from frontal cortex^{17,28} and one snRNA-seq data set from six cortical layers²⁹ were
 603 used as tested data. The test was repeated 10000 times. The p values were
 604 corrected by the Benjamini-Hochberg method.

605

606 ***CG-specific differential expression***

607 Differential expression analysis was conducted in each CG with the Wilcoxon rank-
608 sum test. The p values were corrected by FDR. Genes with FDR q value <0.05 were
609 identified as differentially expressed genes. To calculate the empirical probability of
610 observed fold changes, we conducted a permutation test as follows. In each CG, we
611 permuted the case/control labels and calculated the fold changes of case-control
612 differential expression. The permutation test was repeated 1000 times.

613

614 ***Pathway analysis***

615 Gene ontology enrichment tests for biological processes, molecular function, and
616 cellular components were performed using gProfileR⁶⁶, with p-values FDR-corrected.
617

618 ***CG-specific eQTL mapping***

619 To identify CG-specific eQTLs in the brain, we performed cis-eQTL mapping using
620 package FastQTL⁶⁷. eQTL mapping was performed independently for each CG. We
621 used CG-specific expression estimations from swCAM as phenotype data. The
622 phenotype data were gene expression matrixes for five CGs (five matrixes formatted
623 in genes * samples). A cis-window was defined for genes in each CG as 1 mega
624 base up- and down-stream of the gene body. Each CG expression matrix was
625 adjusted for hidden covariates using PEER⁶⁸ with FDR-corrected p-values. eQTLs
626 with FDR <0.05 were retained.

627

628 To compare CG-specific eQTLs with sorted-cell eQTLs and bulk brain eQTLs, we

629 used Storey's Qvalue package³⁴ and replication rate. The proportion of true
630 associations (π_1) was estimated for significant CG-specific eQTLs in the sorted-cell
631 eQTLs/BrainGVEX DLPFC eQTLs. With the distribution of corresponding p values for
632 the overlapping eQTLs, we calculated π_0 , i.e., the proportion of true null associations
633 based on distribution. Then, $\pi_1 = 1 - \pi_0$ estimated the lowest bound for true-positive
634 associations. The replication rate is the ratio of replicated eQTLs with $FDR < 0.05$
635 among all significant eQTLs in sorted cell populations.

636 ***Heritability estimation with GWAS summary statistics***

637 Partitioned heritability was measured through LD Score Regression v1.0.0⁶⁹,
638 identifying enrichment of GWAS summary statistics among CG-specific eQTLs by
639 accounting for LD. CG-specific eQTL annotation files were created by eSNPs
640 detected in each CG. The annotation file was produced by marking all HapMap3
641 SNPs that fell within the eQTL annotations. LD-scores were calculated for SNPs
642 within the file (LD window of 1cM). The LD reference panel was downloaded from
643 1000 Genomes European Phase Three. SNP LD-scores were interwoven in the
644 computation of partitioned heritability (proportion of h^2). Enrichment for each
645 annotation file was calculated by the proportion of heritability explained by each
646 annotation file divided by the proportion of SNPs falling in that annotation file.
647 Enrichment p values were then corrected by multiple testing.

648

649 ***Estimation and case-control comparison of cell proportions***

650 To compare the CG composition in SCZ patients and healthy controls, we used CAM
651 to estimate the proportion of each CG. Wilcoxon rank-sum test was used to compare
652 the CG proportions between SCZ patients and healthy controls. FDR correction was
653 performed on the p values. To determine the empirical probability of proportion

654 changes in upper-layer neuron CG, we permuted the case/control labels and
655 calculated the case-control proportion changes in upper-layer neuron CG. This
656 process was repeated 1000 times.

657

658 ***Transcriptome-wide association study***

659 We performed TWAS analysis on the GWAS summary statistics⁷⁰ and the CG-
660 specific eQTLs using the TWAS/FUSION package⁹. To identify risk genes with
661 evidence of genetic control, we used genome-wide complex trait analysis to estimate
662 cis-SNP heritability h^2_g (± 1 MB window around gene body). We identified genes with
663 significant h^2_g (nominal p-value < 0.05), which were used to calculate the SNP-based
664 predictive weights per gene. Using the FUSION package, we calculated five-fold
665 cross-validation of five models of expression prediction and evaluated the prediction
666 models for accuracy. The five models are best cis-eQTL, best linear unbiased
667 predictor, Bayesian sparse linear mixed model [BSLMM], and elastic-net regression,
668 LASSO regression. The model with the largest cross-validation R^2 was chosen for
669 downstream association analyses. TWAS statistics were calculated using different
670 cell-group weights, LD SNP correlations from the 1000 Genomes European Phase 3
671 reference panel. TWAS association p values were FDR-corrected and were corrected
672 for multiple CGs.

673

674 ***Gene co-expression analysis***

675 We constructed the gene co-expression network for each CG using Weighted Gene
676 Co-Expression Network Analysis (WGCNA) on deconvoluted CG-specific expression
677 data. A correlation matrix was calculated for the genes, which was then weighted
678 after a scale-free topology was approximated. The following weighted powers were

679 chosen for the respective CGs: 12 for ast_endo and deepN, 14 for upperN,
680 mic_ast_endo, and oli. The signed adjacency, biweight midcorrelation, and
681 blockwiseModules parameters were selected to build the network. Other parameter
682 settings included the following: mergeCutHeight= 0.1, minModuleSize= 30,
683 pamStage=FALSE, and deepSplit=4.

684

685 ***Associating the gene modules with SCZ***

686 Two strategies were used to test the association between gene modules and SCZ.
687 First, we constructed a linear model in which the dependent variable is the module
688 eigengene across all samples and the independent variable is the disease state for
689 the samples (FDR-adjusted p value<0.05). Second, we tested the correlation
690 between GWAS significance and the module membership (kME), which is the
691 expression correlation between each module member gene and the module
692 eigengene. GWAS significance of each module member gene was calculated by
693 MAGMA v1.06⁷¹ using the European subset of the 1000 Genomes as a reference
694 panel for LD. An annotation step was performed first in which GWAS significant
695 SNPs were mapped to genes, based on the presence of a GWAS significant SNP in
696 the region between a gene's start and stop sites +/- 10kb. A competitive gene-set
697 analysis was then performed using sets defined by gene co-expression modules.
698 Resulting p values were FDR-corrected for multiple comparisons. Modules meeting
699 any of the two criteria above were defined as "SCZ-associated modules".

700

701 ***Gene module preservation test***

702 To test the similarity among the CG-specific modules, we tested preservation among
703 the modules detected per CG using Z_{summary} . The network-based preservation test

704 generated two types of statistics: 1) density-based preservation statistics to
705 determine whether the nodes in the reference network were still highly connected in
706 the test network; 2) connectivity-based preservation statistics to determine whether
707 the connectivity pattern in the reference network was preserved in test network. We
708 applied Z_{summary} test by Peter Langfelder et al⁷², in the formula

$$Z_{\text{summary}} = \frac{Z_{\text{density}} + Z_{\text{connectivity}}}{2}$$

709 Z_{density} summarizes density preservation statistics, and $Z_{\text{connectivity}}$ summarizes
710 connectivity-based statistics.

711

712 **Data availability**

713 Access to the raw BrainGVEX data is controlled by the NIMH Repository and
714 Genomics Resources (NRGR), <https://www.nimhgenetics.org/>. Instructions can be
715 found in the PsychENCODE Knowledge Portal:
716 <https://www.synapse.org/#!Synapse:syn4921369>. Deconvoluted data can be reached
717 at http://lbpg.upstate.edu/module_search/. Source data are provided with this paper.

718

719 **Code availability**

720 Scripts used in the data analysis for this manuscript can be found at GitHub:
721 <https://github.com/RujiaDai/CellSpecificAnalysis>. Code for swCAM can be found at
722 <https://github.com/Lululuella/swCAM>.

723 **Acknowledgments**

724 This work was supported by NIH grants 1U01 MH103340-01, 1R01ES024988, and
725 New York State Empire Innovation Program (to C.L.) and National Natural Science
726 Foundation of China (NSFC) grants 31970572 and 31871276, National Key Plan for

727 Scientific Research and Development of China grant 2016YFC1306000, and
 728 Innovation-Driven Project of Central South University grants 2020CX003 (to C.C.).
 729 We thank the PsychENCODE consortium and other data contributors (De Jager et al,
 730 Habib et al, Zhang et al, Darmanis et al, Lake et al, Mendizabal et al, Hodge et al, He
 731 et al) for their data support. We thank James Liu for contributions in web site design
 732 and data visualization. We gratefully acknowledge the families of the brain donors,
 733 without whom this work would not have been possible.

734

735 **Author contributions**

736 Data processing: R.D., S.L., Y.J. Bioinformatics analysis: R.D., S.L., Y.C., J.D.
 737 Algorithm development: L.C., C.W., G.Y., Y.W. Web design: R.D., Q.W. The
 738 conception of study design: R.D., S.L., C.C., C.L. Writing of the manuscript: R.D.
 739 Substantive revisions of the manuscript: R.D., C.L., C.C., Y.W., R.K.

740

741 **Competing interests**

742 The authors have no competing financial interests to declare.

743

744 **Additional information**

745 This work was honored with the Hugh Gurling Memorial Awards by the International
 746 Society of Psychiatric Genetics at XXVI World Congress of Psychiatric Genetics.

747

748

749

750

751

752

753

754

755

756

757

758

759 **Figures and tables**

760

761 **Fig. 1 Cell-group-specific expression for each individual.** (A) Expression of brain cell-type

762 marker genes per individual estimated cell groups (CGs). Oli: oligodendrocytes, Mic:

763 microglia, Ast: astrocytes, Endo: endothelial cells. (B) Annotation of CGs with Expression

764 Weighted Cell Type Enrichment (EWCE) test. Two snRNA-seq data sets from human frontal

765 cortex were used. The black asterisks denote enrichment p value <0.05. (C) Gradients of

766 gene expression across cortical layers in deconvoluted neuronal CGs. The red column

767 denotes upper-layer genes (upperL genes), which express in a decreasing gradient from

768 layers 1 to 6, and the blue column denotes deep-layer genes (deepL genes), which express in

769 an increasing gradient from layers 1 to 6 (FDR<0.05, Spearman correlation). (D) EWCE

770 results showed neuronal CGs enriched in neurons of different cortical layers. The black
 771 asterisks denote enrichment p value <0.05 and the yellow asterisk denotes nominal p value
 772 <0.05. (E) tSNE plot of deconvoluted groups, color-coded by the annotated CG. Each dot
 773 denotes one deconvoluted cell within a sample. Ast_endo: mixture CG of astrocyte and
 774 endothelial cells, deepN: deep-layer neuron CG, upperN: upper-layer neuron CG,
 775 mic_ast_endo: mixture CG of microglia, astrocyte, and endothelial cells, oli: oligodendrocyte
 776 CG.

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

Fig. 2 Summary of cell-group-specific eQTLs. (A) The number of eQTLs and eSNPs detected in each cell group (CG) (FDR corrected p value<0.05). (B) The similarity between CG-specific eQTLs, bulk brain tissue eQTLs, and single-cell eQTLs, evaluated by π_1 values. (C) Genomic annotation of eSNPs. Ast_endo: mixture CG of astrocyte and endothelial cells, deepN: deep-layer neuron CG, upperN: upper-layer neuron CG, mic_ast_endo: mixture CG of microglia, astrocyte, and endothelial cells, oli: oligodendrocyte CG.

Fig. 3 Genetic risk heritability enrichment in cell groups. GWAS heritability was partitioned in five estimated cell groups (CGs). The SNPs of CG-specific eQTLs (FDR corrected p value<0.05) were cataloged in annotated files in sLDSC. (A) the SCZ risk heritability enrichment in different categories (calculated by the proportion of heritability (h2) / proportion of SNPs) and (B) the h2 explained by different categories. The asterisk denotes a corrected p value <0.05 in the enrichment test. ast_endo: mixture CG of astrocyte and endothelial cell, deepN: deep-layer neuron CG, upperN: upper-layer neuron CG, mic_ast_endo: mixture CG of microglia, astrocyte, and endothelial cells, oli: oligodendrocyte CG.

Fig. 4 Proportion per cell group and association with SCZ. (A) Proportions of the five cell groups (CGs) for donors with and without SCZ, estimated with the CAM method. The asterisk denotes corrected p-values of the Wilcoxon rank-sum test <0.05. (B) Spearman correlation between polygenic risk score (PRS) and CG proportion. The red asterisk denotes corrected p-values of correlation test <0.05 and the yellow asterisk denotes nominal p value <0.05. Ast_endo: mixture CG of astrocyte and endothelial cells, deepN: deep-layer neuron CG, upperN: upper-layer neuron CG, mic_ast_endo: mixture CG of microglia, astrocyte, and endothelial cells, oli: oligodendrocyte CG.

Fig. 5 Neuron-specific co-expression modules related to SCZ. Top 10 genes ordered by module membership (kME, i.e., correlation between gene expression and module eigengene) are shown for each neuron-group-specific module.

855 **Table 1 Summary of cell group-specific differential gene expression**

Cell group	#Upreg- ulated gene	#Downreg- ulated gene	Mean (log ₂ fold- change) ± sd	Most prominent functional pathways of upregulated genes	Most prominent functional pathway of downregulated genes
ast_endo	950	1660	0.11±0.18	carboxylic acid catabolic process	defense response
deepN	1254	1297	0.22±0.26	organic acid metabolic process	positive regulation of cell proliferation
upperN	1304	1882	0.26±0.25	carboxylic acid catabolic process	protein folding
mic_ast_en do	1197	1083	0.11±0.18	regulation of cellular component size	positive regulation of immune system process
oli	747	1445	0.14±0.18	modulation of chemical synaptic transmission	positive regulation of immune system process
Ast_endo: mixture CG of astrocytes and endothelial cells, deepN: deep-layer neuron CG, upperN: upper-layer neuron CG, mic_ast_endo: mixture CG of microglia, astrocyte, and endothelial cells, oli: oligodendrocyte CG.					

856

857

858

859

860

861

862

863

864

865

866

867

868

869 **Table 2 Cell group-specific co-expression modules**

Cell group	Module name	Trait association	GWAS enrichment	Representative GO term (FDR<0.05)	Top 3 hub genes
ast_nedo	Ast_endo_ME28	-	SCZ2014, Education, IBD	RNA splicing	SEC62, PPIG, PNN
neuron	neuron_ME28	SCZ↑	SCZ2014, Education	neuronal stem cell population maintenance	TSC22D1, ENTPD3, SYNE1
	neuron_ME29	-	SCZ2014, Education	SWI/SNF superfamily-type complex	WDR48, STOML2, OS9
	neuron_ME30	SCZ↑	SCZ2014, Education, Neuroticism	carbohydrate transmembrane transport	FKBP5, SORT1, WIPF3
	neuron_ME32	-	SCZ2014, Education	negative regulation of actin filament polymerization	GRIA4, CUX2, EPHX4
	neuron_ME27	-	SCZ2014, Education, Neuroticism, BD2016	regulation of dendritic spine morphogenesis	RIMS3, NGEF, LZTS3
mic_ast_endo	mic_ast_endo_ME25	-	-	Defense response to virus	RSAD2, DDX58, PSME2
oli	oli_ME14	SCZ↓	-	Ensheatment of neuron	RNASE1, GPR37, PLP1

Ast_endo: mixture CG of astrocyte and endothelial cells, deepN: deep-layer neuron CG, upperN: upper-layer neuron CG, mic_ast_endo: mixture CG of microglia, astrocyte, and endothelial cells, oli: oligodendrocyte CG.

870

871 Supplemental materials

872

873 **Supplemental Fig. 1 Overview of the study.** Using the sample-wise convex analysis of
874 mixtures (swCAM) deconvolution method, we extracted cell-group-specific (CG-specific)
875 expression from RNA-seq data from bulk brain tissue of 341 samples (SCZ=93, CTL=248). To
876 annotate the identified CGs, maker gene enrichment and correlation testing with scRNA-seq
877 data were used. To discover SCZ-associated cell types and genes, we performed CG-specific
878 eQTL mapping, GWAS heritability enrichment, CG-specific differential expression, CG-
879 specific co-expression, and transcriptome-wide association study. These analyses offer a
880 framework for discerning cell subtypes and underlying genes and pathways associated with
881 SCZ. We found upper-layer neuron CGs in human DLPFC to be associated with SCZ.

882

883

884 **Supplemental Fig. 2** Cell-group enrichment of brain cell marker genes. Enrichment
885 of the top cell-type differentially expressed genes (ctDEGs) in estimated cell groups
886 (CGs) ($\log_2FC > 2$, FDR < 0.05) for human brain cell marker genes. Color denotes
887 $\log_{10}$ -scaled p values of enrichment for significant associations and text in cell
888 denotes odds ratio (Fisher's exact test).

889

890

891

892

893

894

895

896

897

898

Supplemental Fig. 3 Correlations between estimated cell-group-specific (CG-specific) expression and (A) single-cell expression (Darmanis et al. 2015) and (B) single-nuclei expression (Habib et al. 2017). The Spearman correlation test was used. The number and the color both denote the correlation coefficient. ASC: astrocytes, END: endothelial cells, MG: microglia, ODC: oligodendrocytes, OPC: oligodendrocyte precursor cells, EX: excitatory neurons, IN: inhibitory neurons.

907

908 **Supplemental Fig. 4** Expression of marker genes of the inhibitory and excitatory
 909 neuronal subtypes in deconvoluted neuronal cell groups (CGs). Ex: excitatory
 910 neurons; In: inhibitory neurons. The data of subtypes was from a single-nucleus
 911 study of human brains (Lake et al, 2017).

912

913

914

915

916

917

918

919 **Supplemental Fig. 5** Spearman correlation between expressions of estimated
 920 neuronal cell groups (CGs) and expressions of neurons from different layers of the
 921 middle temporal gyrus.

922

923

Supplemental Fig. 6 Comparison with an alternative deconvolution method bMIND. (A) Spearman correlations between CG-specific expressions estimated from swCAM (rows) and bMIND (columns). (B) Replication of CG-specific eQTLs by bMIND. (C) Comparison of SCZ differentially expressed genes identified by swCAM and bMIND.

929

930 **Supplemental Fig. 7** Deconvolution of replication dataset from ROSMAP project. (A) Marker
 931 gene expression in estimated cell groups (CGs). (B) To annotate the identity of CGs, an
 932 EWCE test with snRNA-seq data from human cortex was conducted. (C) Expressions of
 933 upper-layer and deep-layer genes in estimated CGs. (D) To annotate the identity of neuronal
 934 CG1 and CG5, EWCE test with snRNA-seq data of neurons from cortical layers was
 935 conducted. Ast: astrocytes, End: endothelial cells, Mic: microglia, Oli: oligodendrocytes, Opc:
 936 oligodendrocyte precursor cells, Ex: excitatory neurons, In: inhibitory neurons.

937

938 **Supplemental Fig. 8** Replication of CG-specific eQTLs in eQTLs from sorted cell

939 populations. (A) The degree of replication is indexed by Π_1 and (B) replication rate.

940 Π_1 is the proportion of true alternative hypothesis of replicated eQTLs in sorted

941 population. Replication rate is the ratio of replicated eQTLs over eQTLs in sorted

942 populations (FDR<0.05).

943

944

945

946 Proportion change of upper-layer neurons
 947 **Supplemental Fig. 9** Permutation test of proportion changes observed in upper-layer
 948 neurons. The sample labels were shuffled and the proportion changes were
 949 calculated from two randomly selected groups. This process was permuted 1000
 950 times. The black line is the distribution of proportion changes in permutations and the
 951 red line is the observed proportion changes.

upregulated

downregulated

952

953 **Supplemental Fig. 10** Overlap of differentially expressed genes detected in each

954 cell group. The differential expression of genes in patients with SCZ was calculated

955 with Wilcoxon signed-rank test (p value <0.05) per cell group (CG). The left-side bars

956 denote the number of differentially expressed genes in each CG. The upper bars

957 denote the intersection size between sets of differentially expressed genes. Dark

958 connected dots on the bottom panel denote which substrates are considered for each

959 intersection.

960

Supplemental Fig. 11 Permutation test of fold changes of differentially expressed genes in schizophrenia. The sample labels were shuffled and fold changes were calculated from two randomly selected groups. This process was repeated 1000 times. The black lines are the distribution of permuted fold changes and the red lines are the observed fold changes in each CG.

Supplemental table 1 Collected brain cell marker genes

Supplemental table 2 Differential expressed genes in SCZ per cell group

Supplemental table 3 TWAS genes in SCZ per cell group

Supplemental table 4 Coexpression-module member genes and their module membership in each cell group

Supplemental table 5 Source data used in figures and tables

978 References

- 979 1 Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a
980 meta-analysis of twin studies. *Archives of general psychiatry* **60**, 1187-1192,
981 doi:10.1001/archpsyc.60.12.1187 (2003).
- 982 2 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108
983 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).
- 984 3 Richards, A. L. *et al.* Schizophrenia susceptibility alleles are enriched for alleles that affect
985 gene expression in adult human brain. *Molecular psychiatry* **17**, 193-201,
986 doi:10.1038/mp.2011.11 (2012).
- 987 4 Myers, A. J. *et al.* A survey of genetic human cortical gene expression. *Nature genetics* **39**,
988 1494-1499, doi:10.1038/ng.2007.16 (2007).
- 989 5 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the
990 human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).
- 991 6 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for
992 schizophrenia. *Nature neuroscience* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).
- 993 7 Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome
994 illuminate schizophrenia pathogenesis. *Nature neuroscience* **21**, 1117-1125,
995 doi:10.1038/s41593-018-0197-y (2018).
- 996 8 Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of
997 the human brain. *Nature neuroscience* **17**, 1418-1428, doi:10.1038/nn.3801 (2014).
- 998 9 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies.
999 *Nat Genet* **48**, 245-252, doi:10.1038/ng.3506 (2016).
- 1000 10 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait
1001 gene targets. *Nature genetics* **48**, 481-487, doi:10.1038/ng.3538 (2016).
- 1002 11 Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
1003 transcriptome data. *Nat Genet* **47**, 1091-1098, doi:10.1038/ng.3367 (2015).
- 1004 12 Brown, C. D., Mangravite, L. M. & Engelhardt, B. E. Integrative modeling of eQTLs and cis-
1005 regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS*
1006 *Genet* **9**, e1003649, doi:10.1371/journal.pgen.1003649 (2013).
- 1007 13 van der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs
1008 and co-expression QTLs. *Nat Genet* **50**, 493-497, doi:10.1038/s41588-018-0089-9 (2018).
- 1009 14 Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-
1010 dependent manner. *Science* **325**, 1246-1250, doi:10.1126/science.1174148 (2009).
- 1011 15 Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders
1012 parallels polygenic overlap. *Science* **359**, 693-697, doi:10.1126/science.aad6469 (2018).
- 1013 16 Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia,
1014 and bipolar disorder. *Science* **362**, doi:10.1126/science.aat8127 (2018).
- 1015 17 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-
1016 337, doi:10.1038/s41586-019-1195-2 (2019).
- 1017 18 Velmeshev, D. *et al.* Single-cell genomics identifies cell type-specific molecular changes in
1018 autism. *Science* **364**, 685-689, doi:10.1126/science.aav8130 (2019).
- 1019 19 Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate and
1020 Winner's Curse in eQTL studies. *Nucleic acids research* **46**, e133, doi:10.1093/nar/gky780
1021 (2018).
- 1022 20 Shen-Orr, S. S. & Gaujoux, R. Computational deconvolution: extracting cell type-specific
1023 information from heterogeneous samples. *Current opinion in immunology* **25**, 571-578,
1024 doi:10.1016/j.coi.2013.09.015 (2013).
- 1025 21 Wang, N. *et al.* Mathematical modelling of transcriptional heterogeneity identifies novel
1026 markers and subpopulations in complex tissues. *Sci Rep* **6**, 18909, doi:10.1038/srep18909
1027 (2016).
- 1028 22 Chen, L. *et al.* debCAM: a bioconductor R package for fully unsupervised deconvolution of
1029 complex tissues. *Bioinformatics* **36**, 3927-3929, doi:10.1093/bioinformatics/btaa205 (2020).
- 1030 23 Rahmani, E. *et al.* Cell-type-specific resolution epigenetics without the need for cell sorting or
1031 single-cell biology. *Nature communications* **10**, 3417, doi:10.1038/s41467-019-11052-9

- (2019).
- 24 Wang, J., Devlin, B. & Roeder, K. Using multiple measurements of tissue to estimate subject- and cell-type-specific gene expression. *Bioinformatics*, doi:10.1093/bioinformatics/bt619 (2019).
- 25 Chen, L. Mathematical Modeling and Deconvolution for Molecular Characterization of Tissue Heterogeneity. *Doctoral dissertation*, doi:<http://hdl.handle.net/10919/96553> (2020).
- 26 Lulu Chen, C.-H. L., Chiung-Ting Wu, Robert Clarke, Guoqiang Yu, Jennifer E. Van Eyk, David M. Herrington, Yue Wang. Sample-wise unsupervised deconvolution of complex tissues. *bioRxiv*, doi:<https://doi.org/10.1101/2021.01.04.425315> (2021).
- 27 Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37-53, doi:10.1016/j.neuron.2015.11.013 (2016).
- 28 Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955-958, doi:10.1038/nmeth.4407 (2017).
- 29 Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-68, doi:10.1038/s41586-019-1506-7 (2019).
- 30 Jiebiao Wang, K. R., Bernie Devlin. Bayesian estimation of cell-type-specific gene expression per bulk sample with prior derived from single-cell data. *bioRxiv*, doi:<https://doi.org/10.1101/2020.08.05.238949> (2020).
- 31 De Jager, P. L. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data* **5**, 180142, doi:10.1038/sdata.2018.142 (2018).
- 32 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164, doi:10.1093/nar/gkq603 (2010).
- 33 Mendizabal, I. *et al.* Cell type-specific epigenetic links to schizophrenia risk in the brain. *Genome biology* **20**, 135, doi:10.1186/s13059-019-1747-7 (2019).
- 34 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-9445, doi:10.1073/pnas.1530509100 (2003).
- 35 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
- 36 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
- 37 Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715 e1716, doi:10.1016/j.cell.2018.10.022 (2018).
- 38 Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519-523, doi:10.1126/science.1249547 (2014).
- 39 Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nature genetics* **50**, 825-833, doi:10.1038/s41588-018-0129-5 (2018).
- 40 Marin-Padilla, M. The mammalian neocortex new pyramidal neuron: a new conception. *Front Neuroanat* **7**, 51, doi:10.3389/fnana.2013.00051 (2014).
- 41 Finn, E. S., Huber, L., Jangraw, D. C., Molfese, P. J. & Bandettini, P. A. Layer-dependent activity in human prefrontal cortex during working memory. *Nature neuroscience* **22**, 1687-1695, doi:10.1038/s41593-019-0487-z (2019).
- 42 Cullen, T. J. *et al.* Anomalies of asymmetry of pyramidal cell density and structure in dorsolateral prefrontal cortex in schizophrenia. *The British journal of psychiatry : the journal of mental science* **188**, 26-31, doi:10.1192/bjp.bp.104.008169 (2006).
- 43 Garey, L. J. *et al.* Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia. *Journal of neurology, neurosurgery, and psychiatry* **65**, 446-453, doi:10.1136/jnnp.65.4.446 (1998).
- 44 Glantz, L. A. & Lewis, D. A. Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. *Archives of general psychiatry* **57**, 65-73, doi:10.1001/archpsyc.57.1.65 (2000).
- 45 Joshi, D., Catts, V. S., Olaya, J. C. & Shannon Weickert, C. Relationship between somatostatin and death receptor expression in the orbital frontal cortex in schizophrenia: a postmortem brain mRNA study. *NPI schizophrenia* **1**, 14004, doi:10.1038/npschz.2014.4 (2015).
- 46 Benes, F. M., Davidson, J. & Bird, E. D. Quantitative cytoarchitectural studies of the cerebral cortex of schizophrenics. *Archives of general psychiatry* **43**, 31-35,

- doi:10.1001/archpsyc.1986.01800010033004 (1986).
- 47 Mykhailo Y. Batiuk, T. T., Shenglin Mei, Rasmus Rydbirk, Viktor Petukhov, Dora Sedmak, Erzsebet Frank, Virginia Feher, Nikola Habek, Qiwen Hu, Anna Igolkina, Lilla Roszik, Ulrich Pfisterer, Zdravko Petanjek, Istvan Adorjan, Peter V. Kharchenko, Konstantin Khodosevich. Selective vulnerability of supragranular layer neurons in schizophrenia. *bioRxiv*, doi:<https://doi.org/10.1101/2020.11.17.386458> (2020).
- 48 Fenelon, K. *et al.* The pattern of cortical dysfunction in a mouse model of a schizophrenia-related microdeletion. *J Neurosci* **33**, 14825-14839, doi:10.1523/JNEUROSCI.1611-13.2013 (2013).
- 49 Lewis, D. A. Development of the prefrontal cortex during adolescence: insights into vulnerable neural circuits in schizophrenia. *Neuropsychopharmacology* **16**, 385-398, doi:10.1016/S0893-133X(96)00277-1 (1997).
- 50 Benes, F. M., Vincent, S. L. & Todtenkopf, M. The density of pyramidal and nonpyramidal neurons in anterior cingulate cortex of schizophrenic and bipolar subjects. *Biol Psychiatry* **50**, 395-406, doi:10.1016/s0006-3223(01)01084-8 (2001).
- 51 Yoon, K. J. *et al.* Modeling a genetic risk for schizophrenia in iPSCs and mice reveals neural stem cell deficits associated with adherens junctions and polarity. *Cell Stem Cell* **15**, 79-91, doi:10.1016/j.stem.2014.05.003 (2014).
- 52 Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248-268, doi:10.1016/j.neuron.2015.12.008 (2016).
- 53 Selemon, L. D. & Zecevic, N. Schizophrenia: a tale of two critical periods for prefrontal cortical development. *Translational psychiatry* **5**, e623, doi:10.1038/tp.2015.115 (2015).
- 54 Li, M. *et al.* A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. *Nature medicine* **22**, 649-656, doi:10.1038/nm.4096 (2016).
- 55 O'Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome biology* **19**, 194, doi:10.1186/s13059-018-1567-1 (2018).
- 56 Hauberg, M. E. *et al.* Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression. *American journal of human genetics* **100**, 885-894, doi:10.1016/j.ajhg.2017.04.016 (2017).
- 57 Hattori, T. *et al.* DISC1 regulates cell-cell adhesion, cell-matrix adhesion and neurite outgrowth. *Molecular psychiatry* **15**, 778, 798-809, doi:10.1038/mp.2010.60 (2010).
- 58 Hirota, Y. & Nakajima, K. Control of Neuronal Migration and Aggregation by Reelin Signaling in the Developing Cerebral Cortex. *Frontiers in cell and developmental biology* **5**, 40, doi:10.3389/fcell.2017.00040 (2017).
- 59 Jiang, Y. *et al.* DRAMS: A tool to detect and re-align mixed-up samples for integrative studies of multi-omics data. *PLoS computational biology* **16**, e1007522, doi:10.1371/journal.pcbi.1007522 (2020).
- 60 Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 7285-7290, doi:10.1073/pnas.1507125112 (2015).
- 61 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
- 62 Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-315, doi:10.1093/bioinformatics/btg405 (2004).
- 63 Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nature biotechnology* **36**, 70-80, doi:10.1038/nbt.4038 (2018).
- 64 He, Z. *et al.* Comprehensive transcriptome analysis of neocortical layers in humans, chimpanzees and macaques. *Nature neuroscience* **20**, 886-895, doi:10.1038/nn.4548 (2017).
- 65 Skene, N. G. & Grant, S. G. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front Neurosci* **10**, 16, doi:10.3389/fnins.2016.00016 (2016).
- 66 Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research* **35**,

1146 W193-200, doi:10.1093/nar/gkm226 (2007).
1147 67 Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL
1148 mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479-1485,
1149 doi:10.1093/bioinformatics/btv722 (2016).
1150 68 Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of
1151 expression residuals (PEER) to obtain increased power and interpretability of gene expression
1152 analyses. *Nat Protoc* **7**, 500-507, doi:10.1038/nprot.2011.457 (2012).
1153 69 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in
1154 genome-wide association studies. *Nature genetics* **47**, 291-295, doi:10.1038/ng.3211 (2015).
1155 70 Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic
1156 address, d. r. v. e., Bipolar, D. & Schizophrenia Working Group of the Psychiatric Genomics, C.
1157 Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell*
1158 **173**, 1705-1715 e1716, doi:10.1016/j.cell.2018.05.046 (2018).
1159 71 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set
1160 analysis of GWAS data. *PLoS computational biology* **11**, e1004219,
1161 doi:10.1371/journal.pcbi.1004219 (2015).
1162 72 Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and
1163 reproducible? *PLoS computational biology* **7**, e1001057, doi:10.1371/journal.pcbi.1001057
1164 (2011).