

1 **Title: A single-nucleus transcriptome-wide association study implicates novel genes in**
2 **depression pathogenesis**

3
4 Lu Zeng¹, Masashi Fujita¹, Zongmei Gao¹, Charles C. White¹, Gilad S. Green², Naomi Habib²,
5 Vilas Menon¹, David A. Bennett^{3,4}, Patricia A. Boyle^{3,5}, Hans-Ulrich Klein¹, Philip L. De Jager^{1*}

6
7 1. Center for Translational and Computational Neuroimmunology, Department of
8 Neurology, Columbia University Irving Medical Center, New York, NY, USA

9 2. Edmond & Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem,
10 Jerusalem, Israel.

11 3. Rush Alzheimer Disease Center, Rush University Medical Center, Chicago, Illinois, USA

12 4. Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois

13 5. Department of Psychiatry and Behavioral Sciences, Rush University Medical Center,
14 Chicago, Illinois

15 *Corresponding author:
16 Philip L. De Jager, MD PhD
17 Email: pld2115@cumc.columbia.edu

18

19

20

21

22

23

24

25 **Abstract**

26 **Background**

27 Depression is a common psychiatric illness and global public health problem. However, our
28 limited understanding of the biological basis of depression has hindered the development of
29 novel treatments and interventions.

30 **Methods**

31 To identify new candidate genes for therapeutic development, we examined single-nucleus RNA
32 sequencing (snucRNAseq) data from the dorsolateral prefrontal cortex (N=424) in relation to
33 ante-mortem depressive symptoms. To complement these direct analyses, we also used genome-
34 wide association study (GWAS) results for depression (N=500,199) along with genetic tools for
35 inferring the expression of 22,159 genes in 7 cell types and 55 cell subtypes to perform
36 transcriptome-wide association studies (TWAS) of depression followed by Mendelian
37 randomization (MR).

38 **Results**

39 Our single-nucleus TWAS analysis identified 71 causal genes in depression that have a role in
40 specific neocortical cell subtypes; 59 of 71 genes were novel compared to previous studies.
41 Depression TWAS genes showed a cell type specific pattern, with the greatest enrichment being
42 in both excitatory and inhibitory neurons as well as astrocytes. Gene expression in different
43 neuron subtypes have different directions of effect on depression risk. Compared to lower
44 genetically correlated traits (e.g. body mass index) with depression, higher correlated traits (e.g.,
45 neuroticism) have more common TWAS genes with depression. In parallel, we performed
46 differential gene expression analysis in relation to depression in 55 cortical cell subtypes, and we

47 found that genes such as *ANKRD36*, *MADD*, *TAOK3*, *SCAI* and *CHUK* are associated with
48 depression in specific cell subtypes.

49 Conclusions

50 These two sets of analyses illustrate the utility of large snucRNAseq data to uncover both genes
51 whose expression is altered in specific cell subtypes in the context of depression and to enhance
52 the interpretation of well-powered GWAS so that we can prioritize specific susceptibility genes
53 for further analysis and therapeutic development.

54

55 **Introduction**

56 Depression is a common psychiatric illness and is the third leading cause of years lived with
57 disability worldwide (1, 2). Alleviating the burden of this costly disease is an important priority;
58 however, our limited understanding of the biological basis of depression has hindered the
59 development of novel treatments and interventions.

60 Despite the impressive success of the genome-wide association studies (GWAS), there is a
61 substantial gap between the susceptibility variants discovered and understanding how those
62 susceptibility loci contribute to disease onset. Most of the GWAS signals map to non-coding
63 regions and, often, a locus' local linkage disequilibrium (LD) structure does not permit firm
64 conclusions about the identity of the causal variant and its functional effects, which poses a
65 challenge for the identification of risk genes (3) and cell types in which these risk variants may
66 alter gene expression. One approach for addressing this variant-to-function challenge is to use
67 expression quantitative trait loci (eQTL) mapping to characterize the impact of disease-
68 associated regulatory variants on the expression of nearby genes (4).

69 Given the complexity of psychiatric disorders such as depression, disentangling the role of each
70 cell type in the brain is important and requires studies performed at single-cell resolution. In this
71 study, to identify potential causal depression genes in brain, we have leveraged a large set of data
72 on individual nuclei extracted from the dorsolateral prefrontal cortex (DLPFC), an important hub
73 in mood circuits, of 424 older individuals from a collection of prospectively collected brain
74 autopsies of the Religious Orders Study and Rush Memory and Aging Project (ROSMAP) cohort.
75 We integrated this gene expression data set and results of depression GWAS analyses (5), as
76 implemented in function summary-based imputation (FUSION) (6), Summary-data-based

77 Mendelian Randomization (SMR) (7) and colocalization analysis (COLOC) (8) in the 7 major
78 cell types of the DLPFC. In this study, FUSION identifies genes whose cis-regulated gene
79 expression is associated with depression, and SMR tests whether these genes mediate the
80 association between genetic variants and depression, COLOC assesses how replicable of
81 FUSION analysis. FUSION is a suite of tools for performing transcriptome-wide association
82 studies (TWAS), it builds predictive models of the genetic component of a functional phenotype
83 and predicts and tests each component for association with disease using GWAS summary
84 statistics. COLOC is used to perform genetic colocalization analysis of two potentially related
85 phenotypes, to ask whether they share common genetic causal variant(s) in a give region. TWAS
86 and COLOC had similar power under the scenario with a single typed causal variant, but TWAS
87 had superior performance when the causal variant was untyped or in the presence of allelic
88 heterogeneity (6). SMR integrates summary-level data from GWAS with data from eQTL studies
89 to identify genes whose expression levels are associated with a complex trait because of
90 pleiotropy.

91 Furthermore, we also performed differential gene expression analysis within 55 cortical cell
92 subtypes that were present in ≥ 100 participants, discovering additional depression-associated
93 genes expressed only in a subtype of cells, highlighting the importance of cellular context for
94 uncovering many variants which influence gene expression (**Figure 1**). Our analyses illustrate
95 the utility of large snucRNAseq data to uncover both genes whose expression is altered in
96 specific cell subtypes in the context of depression and to enhance the interpretation of well-
97 powered GWAS so that we can prioritize specific susceptibility genes for further analysis.

98

99 **Methods and Materials**

100 **Study participants**

101 All brain specimens were derived from two longitudinal clinico-pathological cohorts: the
102 Religious Orders Study (ROS), or the Memory and Aging Project (MAP). In both cohorts,
103 participants did not have known dementia at the time of enrollment. The participants agreed to
104 receive clinical evaluation each year and donate their brain after death; the two studies were
105 designed and are run by the same group of investigators, with essentially identical ante- and post-
106 mortem phenotypic data collection. Thus, they are designed to be analyzed jointly (9) and are
107 referred to as “ROSMAP”. For this study, 424 participants with both snucRNAseq and whole
108 genome sequence data were retained for analysis. At the time of death, 34% of participants were
109 cognitively non-impaired, 26% were mildly impaired, and 40% were demented. Of the 424
110 participants, 68% were female.

111

112 **ROSMAP depressive symptoms**

113 ROSMAP depressive symptoms were assessed with a modified, 10-item version of the Center
114 for Epidemiologic Studies Depression scale (CES-D). Participants were asked whether or not
115 they experienced each of ten symptoms much of the time in the past week (e.g. could not get
116 going, felt depressed). The score is the total number of symptoms experience, more details can
117 be found in

118 <https://www.radc.rush.edu/docs/var/detail.htm?category=Depression&variable=cesdsum>.

119

120 In this study, we have analyzed differential gene expression based on two depressive symptom
121 scores: the score from the last visit for each participant, and the average depressive symptom

122 score over time calculated for each donor (average visit times=8 (SD= ± 6), follow-up time frame:
123 IQ-25=3.0yrs, median=6.2yrs, IQ-75=11.2yrs).

124

125 Linear regression model for depressive symptom-associated gene detection

126 We implemented a linear regression model to identify depressive symptom-associated gene
127 expression in 7 cell types and 55 cell subtypes (Eq. 1).

$$128 Y_{ij} = \beta_j + \alpha_j Depression_i + \gamma_j Age_i + \delta_j Sex_i + \mu_j cogres.path.tdp_i + \delta_j cogng_demog_slope_i + \varepsilon_{ij}$$

129 (1)

130 In our linear regression model, Y_{ij} is the expression level of gene j in donor i ,
131 $Depression_i$ denotes the last visit or average depressive symptom score of donor i , Age_i
132 denotes the age in last visit or age at death of donor i , Sex_i denotes the sex of donor i ,
133 $cogres.path.tdp_i$ is the residual cognition adjusted by amyloid, tau and TDP-43 pathologies of
134 donor i , $cogng_demog_slope_i$ represents the random slope of global cognition of donor i , ε_{ij} is
135 the error term, α_j , γ_j , δ_j , μ_j , δ_j , is the correlation coefficient for each variates. The corresponding
136 correlation and p -values (adjusted with BH (Benjamini & Hochberg) were then calculated for all
137 genes; only FDR value <0.05 were considered as significant depressive symptom-associated
138 genes.

139

140 Whole genome sequencing

141 Whole genome sequencing (WGS) of ROSMAP participants was performed as described
142 previously (10). Briefly, DNA was extracted from brain or blood samples. WGS libraries were
143 prepared using the KAPA Hyper Library Preparation Kit and sequenced on an Illumina HiSeq X
144 sequencer as paired end reads of 150 bp. Reads were mapped to the reference human genome

145 GRCh37 using BWA-mem, and variants were called by GATK HaplotypeCaller. For this study,
146 the VCF file was lifted over to GRCh38 using Picard LiftoverVcf, and only variants that passed
147 the GATK filter (variant quality score-based sensitivity < 99.8%) were used (11). Variant Call
148 Format (VCF) files of GWAS are available at Synapse
149 (<https://www.synapse.org/#!Synapse:syn11724057>).

150

151 Single-nucleus RNA-seq

152 Single-nucleus RNA sequencing (snucRNAseq) of 424 DLPFC was performed as described
153 previously (12). In brief, frozen specimens of DLPFC from ROSMAP participants were obtained
154 from Rush Alzheimer's Disease Center. Gray matter was extracted and dissociated into nuclei
155 suspension. Single-nucleus RNA-seq libraries were constructed using Chromium Single Cell 3'
156 Reagent Kits version 3 (10x Genomics) following the manufacturer's protocol and sequenced
157 using HiSeqX and NovaSeq sequencers (Illumina). FASTQ files were processed using
158 CellRanger (v6.0.0; 10x Genomics) with the "GRCh38-2020-A" transcriptome and the "include-
159 introns" option. Cell calling and ambient RNA removal were performed using the CellBender
160 software. All raw and processed data are available through the AD Knowledge Portal
161 (<https://www.synapse.org/#!Synapse:syn31512863>).

162

163 Cell type classifications

164 Based on the cell-type annotations of our prior work (13), we used a stepwise clustering
165 approach to identify first the major cell types of the DLPFC and then subtypes in each cell type.
166 In the end, we analyzed data organized into 7 major cell types, which were further subdivided
167 into 92 cell subtypes found in the human DLPFC (more details can be found in (12)).

168

169 Pseudo-bulk expression

170 For each cell type and cell subtype, pseudo-bulk UMI count matrices were constructed by
171 extracting UMI counts of the cell (sub)type of interest, aggregating them for each participant,
172 and normalizing them by sequencing depth. Pseudo-bulk UMI counts were normalized by using
173 the trimmed mean of M-values (TMM) method of edgeR (14), and \log_2 of counts per million
174 mapped reads (CPM) were computed using the voom function of limma (version 3.44.3) (15).
175 Low expression genes ($\log_2\text{CPM}<2.0$) were filtered out. Batch effects were corrected using
176 ComBat (16). Expression levels were quantile normalized. Pseudo-bulk expression of cell
177 subtypes was quantified by the same method.□

178

179 *cis*-eQTL mapping

180 *Cis*-eQTLs were identified by linear regression with 1 Mbp of the transcription start site of each
181 measured gene (gene expression derived using $\log_2\text{CPM}$), as implemented in Matrix eQTL (ver.
182 2.3) (17), adjusting for top 3 genotype PCs, top 30 expression PCs, age, sex, post-mortem
183 interval, study (ROS or MAP), and total number of genes detected in each participant. Multiple
184 hypothesis correction was performed using a two-step method. Gene-wise p -values were
185 computed by applying Bonferroni correction to the smallest nominal p -value of each gene with
186 the number of tested SNPs for the gene. The threshold for statistical significance of eGenes was
187 set to the false discovery rate (FDR) $<5\%$, where FDR was computed from gene-wise p -values
188 using the Benjamini-Hochberg method. Statistical significance of eSNPs were judged by
189 nominal p -values, and its threshold was set to the largest nominal p -value of gene-SNP pair that
190 had FDR $<5\%$. More details can be found in our recent paper (12).

191

192 GWAS summary statistics

193 We used depression GWAS summary statistics from 500,199 participants of European descent
194 by Howard et.al (5) that did not include 23andMe participants, of these, 361,315 were from the
195 UK Biobank and 138,884 were from the Psychiatric Genetics Consortium (18). Approximately
196 34% of participants had depression.

197

198 Five other diseases/traits GWAS summary statistics were used to investigate the specificity of
199 the depression TWAS results. Including three psychiatric disorders: bipolar disorder (BD)
200 ($N_{\text{cases}}=41,917$; $N_{\text{controls}}=371,549$) (19), schizophrenia (SCZ) ($N_{\text{cases}}=76,755$; $N_{\text{controls}}=243,649$) (20)
201 and neuroticism ($N=390,278$) (21); body-mass-index (BMI) ($N\sim 700,000$) (22) and waist-to-hip
202 ratio adjusted for BMI (WHRadjBMI) ($N=484,680$) (23).

203

204 Expression network analysis

205 Gene expression network was constructed by WGCNA (24), an R package that implements a
206 computationally optimized procedure for weighted gene coexpression network analysis. Gene
207 expression values were normalized and covariate-adjusted (i.e., study index, postmortem interval
208 and total genes detected). For each gene expression module, we performed module-trait
209 correlation analysis between module eigengene and 29 phenotypic traits (e.g. last visit/average
210 depressive symptom score, amyloid, *APOE* genotype, braak stage). To do this, we corrected age
211 for sex, sex for age, age and sex for the other 27 traits, and education was corrected as an extra
212 covariate for cognition-related traits (cogng_demog_slope). To account for multiple hypotheses,
213 FDR corrected p-value threshold ($\text{FDR}\leq 0.05$) was used to define significant trait-module

214 relationships. Similar analysis was applied to other five cell types: astrocytes, excitatory neurons,
215 inhibitory neurons, oligodendrocytes and OPCs. We also performed correlation analysis between
216 each module eigengene and the frequency of cell subtypes within each major cell type.

217

218 eQTL integrative analysis

219 We used pseudo-bulk RNA-seq data and genotypes from ROSMAP (424 brain subjects) to
220 impute the *cis* genetic component of expression into the depression GWAS summary statistics
221 (5). The complete TWAS pipeline is implemented in the FUSION (ver. Oct. 1, 2019) suite of
222 tools (6). The details steps implemented in FUSION are as follows. First, we estimated the
223 heritability of gene expression and stopped if not significant. We estimated using a robust
224 version of GCTA-GREML (25), which generates heritability estimates per feature as well as the
225 likelihood ratio test P value. Only features that have a heritability of $P < 0.05$ were retained
226 for TWAS analysis. Second, the expression weights were computed by modeling all *cis*-SNPs
227 (± 1 Mb from the transcription start site) using best linear unbiased prediction, or modeling
228 SNPs and effect sizes with Bayesian sparse linear mixed model, least absolute shrinkage and
229 selection operator, Elastic Net and top SNPs (6, 26). A cross-validation for each of the desired
230 models were then performed. Third, a final estimate of weights for each of the desired models
231 was performed and the results were stored. The imputed unit is treated as a linear model of
232 genotypes with weights based on the correlation between SNPs and expression in the training
233 data while accounting for linkage disequilibrium (LD) among SNPs. To account for multiple
234 hypotheses, FDR corrected p -value threshold ($FDR \leq 0.05$) was used to define significant TWAS
235 associations. snucRNA-sequencing from 55 cell subtypes (sample size ≥ 100) was also imputed
236 for TWAS analysis (**Figure S1**).

237
238 SMR (ver. 1.03) (7) was used to test whether depression TWAS-significant genes (from the
239 FUSION approach) were associated with depression via their *cis*-regulated brain transcriptomics.
240 We used the ROSMAP genotype, our pseudo-bulk cell types eQTL results and the Howard et al.
241 depression GWAS summary statistics to perform SMR, the conservative unadjusted $P < 0.05$
242 from heterogeneity in dependent instruments (HEIDI) was used to suggest that the presence of
243 linkage likely influences the main SMR findings. Similarly, we performed FUSION, SMR and
244 HEIDI in bipolar disorder, schizophrenia, neuroticism, BMI and WHRadjBMI.

245
246 **Colocalization analysis**
247 The COLOC package (version 5.1.0) was applied to test the approximate Bayes factor (ABF)
248 colocalization hypothesis, which assumes a single causal variant. Under ABF analysis, the
249 association of a trait with a SNP is assessed by calculating the posterior probability (value from 0
250 to 1), with the value of 1 indicating the causal SNP. In addition, the ABF analysis has 5
251 hypotheses, where, PP.H0.abf indicates there is neither an eQTL nor a GWAS signal at the loci;
252 PP.H1.abf indicates the locus is only associated with the GWAS; PP.H2.abf indicates the locus is
253 only associated with the eQTL; PP.H3.abf indicates that both the GWAS and eQTL are
254 associated but to a different genetic variant; PP.H4.abf indicates that the eQTL and the GWAS
255 are associated to the same genetic variant. With the posterior probability of each SNP and aiming
256 to find the casual variants between the GWAS and eQTL, we focused on extracting the PP.H4
257 value for each SNP in our study. □ □

258 For depression GWAS (5), we used the reported lead SNPs of 102 loci. For each locus, we
259 searched for the eSNPs that are within 500 Kb of the lead SNP, and listed eGenes that were

260 paired with the eSNP. We then obtained the eGenes *cis*-eQTL output around the lead eSNP
261 within 1 Mbp window size. In addition, we extracted GWAS summary statistics around the
262 reported 102 lead SNP. At last, we conducted COLOC for respective pair of eGene-eQTL and
263 eSNP-GWAS for each cell type. □

264

265 Function annotation for cell subtype-level depression TWAS genes

266 The DAVID tool (27) was used to perform GO annotation. Up-/down-regulated depression
267 TWAS gene lists were submitted to DAVID by choosing GO_FAT and KEGG pathway terms to
268 describe the overrepresented functional terms. The threshold for overrepresented GO terms was
269 set to FDR<0.05.

270

271 Cross-trait LD score regression

272 LDSC (28) bivariate genetic correlations attribute to genome-wide SNPs (r_g) were estimated
273 across seven human diseases/traits from published GWASs as we mentioned above. We used LD
274 scores from the 'eur_w_ld_chr' file available from
275 <https://alkesgroup.broadinstitute.org/LDSCORE>, computed using 1000 Genomes Project (29)
276 Europeans as a reference panel as previously described (30). Adjusting for the number of traits
277 tested, the FDR *P*-value threshold (FDR<0.05) was used to define significant genetic correlations.

278

279 Results

280 Differentially expressed depressive symptom genes

281 The Religious Order Study (ROS) and the Memory and Aging Project (MAP) are two
282 longitudinal studies of brain aging with annual neuropsychiatric evaluations and prospective
283 autopsy; all participants in the two studies are non-demented at the time of enrollment. The two
284 studies were designed and are run by a single set of investigators, and they were designed to be
285 analyzed jointly. As a result, we refer to them as ROSMAP. The ROSMAP team has assessed
286 the depressive symptoms of their participants over time, we assessed the relation of each cell
287 subtype to depressive symptom scores. Specifically, we evaluated two related outcomes: (1) the
288 depressive symptom score at the last visit prior to death using a modified, 10-item version of the
289 Center for Epidemiologic Studies Depression scale (CES-D), and (2) the average depressive
290 symptom score over the participant's time in the study (IQ-25=3.0yrs, median=6.2yrs, IQ-
291 75=11.2yrs) (see Methods, **Figure S2**). The latter outcome may help to better capture the
292 participant's history of depression while the former outcome may best capture the state near the
293 time of death. A linear regression model was then used to identify differentially expressed
294 depressive symptom genes in 55 cell subtypes with pseudo-bulk RNA-seq measures (we limit
295 these analyses to those subtypes that have >10 cells in ≥ 100 participants), adjusting for age, sex,
296 and two manifestations of AD: slope of global cognitive function and residual cognition adjusted
297 for AD pathologies. We adjusted for these two cognition-related traits due to their significant
298 correlation with the last visit/average depressive symptom scores (**Supplementary Table S1**).

299
300 By using the last visit depressive symptom score and an FDR<0.05, we identified 8 depressive
301 symptom-associated genes in astrocytes, neurons or oligodendrocytes (**Table 1**). For example,
302 TAO Kinase 3 (*TAOK3*) (FDR=0.01) (**Figure 2**), which showed downregulated gene expression
303 with depressive symptom in oligodendrocyte subtype 12. The protein encoded by this gene is a

304 serine/threonine protein kinase that activates the p38/MAPK14 stress-activated MAPK cascade
 305 but inhibits the basal activity of the MAPK8/JNK cascade. A previous study found individuals
 306 suffering from bipolar disorder and schizophrenia showed that a microdeletion that affects
 307 *TAOK3* (and *PEBP1*) is present in schizophrenia patients (31). Further, a GWAS analysis
 308 suggested that *TAOK3* may contribute to neurodevelopmental disorders, at least in schizophrenia
 309 (32). Another gene, MAP kinase-activating Death Domain protein (*MADD*) showed increased
 310 expression with depressive symptom in inhibitory neuron subtype 11 (FDR=0.03). This gene
 311 plays a survival-promoting role against TNF mediated apoptosis (33), and it has been found
 312 associated with post-traumatic stress disorder (34).

313

314 **Table 1: Results of the differentially expressed genes identified from 55 cell subtypes.**

315 Asterisks indicate genes whose cis-regulated brain mRNA levels were also associated with
 316 depression based on a transcriptome-wide association study (TWAS) of depression that
 317 integrated the depression GWAS (N=500,199) with snucRNAseq transcriptomic and genetic
 318 data (N=424).

Gene	Chromosome	coefficient	p-value	FDR	Cell subtype
Last visit depressive symptom score					
<i>ADHFE1</i>	8	0.20	8.43×10^{-6}	0.02	Ast.5
<i>DCUN1D2</i>	13	0.20	7.85×10^{-6}	0.02	Ast.5
<i>CHUK*</i>	10	-0.14	2.13×10^{-6}	0.03	Exc.4
<i>AC005225.2</i>	14	-0.14	2.20×10^{-6}	0.02	Inh.6
<i>MADD</i>	11	0.14	6.76×10^{-6}	0.03	Inh.11
<i>MEF2A</i>	15	0.13	6.87×10^{-6}	0.03	Oli.2

<i>CLASP2</i>	3	0.16	6.27×10^{-5}	0.04	Oli.7
<i>TAOK3*</i>	12	-0.14	4.29×10^{-6}	0.02	Oli.12
Average depressive symptom score					
<i>CRADD</i>	12	-0.18	2.20×10^{-6}	0.02	Exc.7
<i>CCDC6*</i>	10	0.33	1.18×10^{-6}	7.26×10^{-3}	Exc.10
<i>AC005225.2</i>	14	-0.17	4.01×10^{-6}	0.03	Inh.6
<i>C11orf58*</i>	11	0.19	8.73×10^{-6}	0.03	Inh.9
<i>ZCCHC17</i>	1	0.179	6.71×10^{-6}	0.03	Inh.11
<i>TBC1D5</i>	3	-0.18	7.84×10^{-6}	0.01	Inh.12
<i>SCAI*</i>	9	0.19	4.03×10^{-6}	0.01	Inh.12

319

320 Another 7 depressive symptom-associated genes were identified in neurons when using the

321 average depressive symptom score over the study period (**Table 1**), including Coiled-Coil

322 Domain Containing 6 (*CCDC6*) ($FDR=7.26 \times 10^{-3}$) and Suppressor of Cancer Cell Invasion (*SCAI*)

323 ($FDR=0.01$) (**Figure 2**). *CCDC6* encodes a coiled domain-containing protein, which it has been

324 reported as a causal gene for the psychiatric and neurodegenerative diseases at both the mRNA

325 and protein levels in bulk cortex data (35). *SCAI* was found downregulated in several human

326 tumors, and decreased levels of *SCAI* are tightly correlated with increased invasive cell migration

327 (36). It was also found to be modestly associated with major depressive disorders (MDD) (37).

328 We did not pursue pathway analyses given the small number of associated genes, which are not

329 adequate to support such analyses. Thus, while we are likely still underpowered for a differential

330 gene expression analysis, we identified a small number of intriguing candidates, including two

331 neuronal genes that have been implicated in prior studies of bulk cortical data. To be thorough,

332 we also assembled a list of 109 statistically robust reported associations from bulk cortical RNA
333 data in prior studies (38-40), and we evaluated them for altered expression in our data: we
334 detected no significant differential expression. However, this is not surprising, since depression
335 is a polygenic trait, arising from the influence of multiple loci with small individual effects, and
336 our sample size remains modest.

337

338 To complement the single gene analyses, we repeated our analyses after reducing the
339 dimensionality of the data: using WGCNA, we established modules of co-expressed genes in
340 each cell type and derived an eigengene measure for each module (**Figure S3A-C,**
341 **Supplementary Table S2**). However, these analyses returned no association to either the last
342 measure or the average measure of depressive symptom (**Figure S3B**), suggesting that, as with
343 GWAS, we will need substantially larger studies to be properly powered for identifying
344 alterations in gene expression relative to depression. Finally, we also assessed the frequency of
345 107 cell subtypes that we defined in these data (12) in relation to the two outcomes, but we did
346 not find any significant associations either. Thus, while direct association analyses in our
347 moderately sized study for depression yielded only a small number of results, this data resource
348 is well-powered for other analyses – such as expression quantitative trait locus (eQTL) mapping
349 - that can be leveraged to powerfully interrogate the genetic architecture of the disease.

350

351 **A Transcriptome-Wide Association Study (TWAS) of depression**

352 We integrated our snucRNAseq-based eQTL results (12) with the latest depression GWAS (5)
353 results to perform a TWAS of depression using the FUSION pipeline (6): the expression level of
354 a subset of genes can be inferred from single nucleotide polymorphism (SNP) data, and these

355 genetic instruments can be deployed into any relevant collection of genome-wide genotype data.
356 Using over 1.6 million human brain transcriptomes derived from individual nuclei extracted
357 from the DLPFC of 424 older individuals (**Figure S4**) of European descent from the ROSMAP
358 cohorts (41), we have previously mapped *cis*-eQTLs in 7 major cell types, and 55 cell subtypes
359 which have data from more than 100 participants (12). Genes that had significant single
360 nucleotide polymorphism (SNP)-based heritability ($p < 0.05$) were retained in the TWAS analysis
361 (**Supplementary S3, Figure S5**). Depression GWAS summary statistics were from 500,199
362 participants of European descent, who were not 23andMe, Inc. participants. TWAS identified the
363 greatest number of associated genes in excitatory neurons: 99 genes displayed association of
364 inferred gene expression with depression ($FDR < 0.05$) (**Figure 3A and Supplementary Table**
365 **S4**). Associations in other cell types include: 68 genes in inhibitory neurons, 67 genes in
366 astrocytes, 51 genes in oligodendrocytes, 33 genes in oligodendrocyte progenitor cells (OPCs),
367 28 genes in microglia, and 2 genes in endothelial cells. We note that our dataset had relatively
368 few endothelial cell nuclei in each participant, limiting our power to map *cis*-eQTLs in that cell
369 type; in general, the less frequent cell types such as microglia also had fewer genes whose
370 expression we could infer (12).

371
372 To further evaluate whether gene expression mediates the association between genetic variants
373 and depression for each of the genes identified in one of the cell types in the FUSION analyses,
374 we performed Summary-data-based Mendelian Randomization (SMR) (7). Results from SMR
375 demonstrated that 67/99 genes in excitatory neurons, 36/68 genes in inhibitory neurons, 34/67
376 genes in astrocytes, 30/51 genes in oligodendrocytes, 16/33 in OPC, 15/28 in microglia reached a
377 nominally significant level ($p < 0.05$). Next, we used the heterogeneity independent instrument

378 (HEIDI) test to distinguish pleiotropy/causality from linkage for genes in each of the 7 cell types.
379 HEIDI results suggested that 11/99 genes in excitatory neurons, 6/68 genes in inhibitory neurons,
380 4/67 genes in astrocytes, 4/51 genes in oligodendrocytes, 2/33 genes in OPC, 3/18 genes in
381 microglia were consistent with either pleiotropy or causality ($p > 0.05$) (**Supplementary Table**
382 **S5**). In **Figure 3C**, we illustrate a Manhattan plot of the 99 depression TWAS genes identified in
383 excitatory neurons, highlighting the results of novel candidates as well as SMR and HEIDI
384 results.

385
386 Next, we conducted colocalization analysis (COLOC) to assess whether altered gene expression
387 may be the mechanism for a particular risk allele or haplotype. Using a genome-wide association
388 study of major depressive disorder (MDD) and our list of eGenes, we find evidence of
389 colocalization ($PP.H4 > 0.5$) for 43 eGenes among the 102 MDD loci that we interrogated (**Figure**
390 **3D, Supplementary Table S6**). As expected, excitatory and inhibitory neurons harbor the most
391 implicated target genes. We note that while many loci have unambiguous cell type-specific
392 effects (i.e., *TMTC2* or *ZKSCAN7*), *AL139147.1* has an effect shared by four cell types (Inh, Exc,
393 Ast, Oli and OPC). In addition, the COLOC analysis is helpful in supporting the FUSION
394 analysis: 65% (28/43) of TWAS genes also showed colocalized effects in MDD. Taken together,
395 our results from FUSION, SMR and HEIDI, and COLOC analysis suggested that there is an
396 enrichment of depression genes in neurons and astrocytes.

397
398 We also applied FUSION, SMR and HEIDI on the 55 cell subtypes of single-nucleus RNA-seq
399 for identifying depression genes (**Supplementary Table S7-S9**). Consistent with results at the
400 cell type level, we identified the greatest number of depression TWAS genes in excitatory

401 neurons, especially subtype 4, 12 and 3 (**Supplementary Table S7**). Among the 1,068
402 depression TWAS genes identified from 55 cell subtypes, 167 of them were found shared with
403 252 genes in 7 cell types (**Supplementary Table S10**). The high but not complete overlap of
404 depression TWAS genes (66.3%) identified from cell types and cell subtypes indicates that using
405 cell subtype-level snucRNAseq data can provide additional gene signals whose expression is
406 altered in specific cell subtypes.

407
408 Overall, our results showed that the effect directions of genetically regulated expression on
409 depression risk are largely consistent across cell subtypes within the same cell type (**Figure 4A**).
410 For example, the increased gene expression of *SLC12A5* was found associated with increased
411 depression risk across cell subtypes of excitatory neurons and inhibitory neurons. A previous
412 study integrating depression GWAS with bulk brain and blood eQTL data found *SLC12A5*
413 expression showed evidence of a genetically predicted effect on depression and neuroticism (42),
414 so we refined this robust association by highlighting the relevant cell type. By contrast, *NEGR1* in
415 excitatory neuron subtype 3 (Exc.3) and 4 (Exc.4) showed opposite effects for genetically
416 regulated expression of depression risk: in Exc.3, the increased gene expression of *NEGR1* is
417 associated with increased depression risk, while in subtype 4, the increased gene expression of
418 *NEGR1* associated with decreased depression risk (**Figure 4A**). *NEGR1* is a cell adhesion
419 molecule expression in neurons, and its expression in primary hippocampal neurons was found to
420 be significantly reduced in antidepressant-treated rats (43). Similar patterns of opposite
421 directions of effect can be seen with *ZFYVE21* and *B3GLCT* in excitatory neuron subtypes, and
422 *TMEM161B-AS1* in inhibitory neuron subtypes (**Figure 4A**). This is intriguing, suggesting that

423 gene expression related to functional genetic variation can be differentially modulated among
424 different subtypes of neurons.

425

426 In addition, we inspected the functional similarity and difference among genes associated with
427 depression by using the DAVID tool (27). Depression TWAS genes were divided into 491 with
428 up- and 449 with down-regulation with respect to the effect of depression risk allele, and genes
429 showed inconsistent effect directions across cell subtypes were excluded for this analysis. GO
430 annotation and pathway analysis demonstrated differential functional enrichment, among up-
431 regulated depression TWAS genes (**Figure 4B**): they were characterized by being involved in
432 neuronal projection (false discovery rate (FDR= 1.01×10^{-7}), synapse function (FDR= 1.09×10^{-6}),
433 protein serine/threonine kinase activity (FDR= 2.18×10^{-3}), behavior (FDR= 4.39×10^{-3}), or bipolar
434 disorder and schizophrenia (FDR=0.03). On the other hand, down-regulated depression TWAS
435 genes were found to be enriched in fewer functions (**Figure 4C**), for example cytosol
436 (FDR= 1.51×10^{-4}), nervous system development (FDR= 5.21×10^{-4}), nucleoplasm
437 (FDR= 4.77×10^{-3}) and anchoring junction (FDR=0.01) (a full list of functional annotations is
438 provided in **Supplementary Table S11**). These results suggest that depression variants where
439 the risk allele increases gene expression may be more likely to influence neuronal function at the
440 synapse while the loci containing alleles that diminish expression will need further evaluation.

441

442 **Identifying novel depression genes**

443 To extract novel depression genes identified from our analysis, we compiled a list of 176 unique
444 genes reported from previous studies (18, 44, 45), which all used a TWAS approach in bulk
445 RNAseq data. Comparing these known depression TWAS genes with our FUSION, SMR and

446 HEIDI results, we found in 7 cell types, 34/252 (13.5%) of depression TWAS genes
447 (hypergeometric test $p < 1.00 \times 10^{-16}$), and 14/43 (32.6%) of depression COLOC genes were
448 common with the known candidates, while this number reduced to 5/28 (17.9%) when using the
449 SMR and HEIDI results (hypergeometric test $p = 0.32$). When it comes to the 55 cell subtypes, we
450 found 67/1,068 (6.3%) of the depression TWAS genes (hypergeometric test $p < 1.00 \times 10^{-16}$), and
451 9/51 (17.6%) of TWAS-significant SMR and HEIDI genes (hypergeometric test $p < 3.20 \times 10^{-03}$)
452 are shared with the known gene list from bulk RNA studies. Thus, combining cell types and cell
453 subtypes results, we identified 59 unique novel depression genes based on FUSION, SMR and
454 HEIDI analysis (**Supplementary Table S12**). This represents an important step towards
455 enabling a cell type-specific downstream dissection of genes implicated in susceptibility to
456 depression.

457

458 **Cell type similarity and specificity of depression TWAS genes**

459 We next investigated the specificity of depression genes by displaying the extent to which they
460 are shared among the different cell types (**Figure 3B**). The results showed a cell type-specific
461 pattern, for example, 60/99 (60.6%) depression genes in excitatory neurons were specific to this
462 cell type, and only 14 of the 99 genes were found to be shared with inhibitory neurons. This
463 result can be explained by (i) a target gene that is expressed only in one cell type, (ii) a target
464 gene that is expressed in multiple cell types but where the genetic variant affects only one cell
465 type. A similar trend was observed in the results of the cell subtype analysis (**Figure S6**).

466

467 **Trait similarity and specificity of the depression TWAS genes**

468 To understand the trait specificity of the depression TWAS results, we performed a TWAS for
469 neuroticism, bipolar disorder (BD), schizophrenia (SCZ), body-mass-index (BMI) and waist-to-
470 hip ratio (WHR) adjusted BMI (WHRadjBMI) (**Figure 5A**). These traits were chosen because
471 they have a range of estimated genetic correlation with depression: 0.69 for neuroticism, 0.45 for
472 BD, 0.32 for SCZ, 0.11 for BMI and WHRadjBMI (**Figure 5B**). We expected that traits with
473 evidence of higher genetic correlation would have more TWAS results in common. Next, we
474 performed TWAS analysis on GWAS summary statistics for each trait (see Methods) and
475 pseudo-bulk RNA-seq in seven cell types. Combining 7 cell types together, the TWAS of
476 neuroticism identified 521 genes, the TWAS of BD identified 574 unique genes, the TWAS of
477 SCZ identified 1,318 unique genes, the TWAS of BMI identified 57 unique genes and the
478 TWAS of WHRadjBMI identified 3,033 genes (**Supplementary Table S13**).

479
480 In addition, we applied SMR and HEIDI on the TWAS-significant genes to remove genes with
481 an SMR $p < 0.05$, a HEIDI $p > 0.05$ or in cases where a HEIDI P value was unable to be
482 determined. These tests allow us to focus on genes with evidence that their genetically regulated
483 gene expression mediates their association with the trait of interest and to remove genes likely to
484 be the result of linkage disequilibrium. After considering findings from FUSION, SMR and
485 HEIDI, we identified 70 genes in neuroticism, 103 genes in BD, 214 genes in SCZ, 5 genes in
486 BMI and 378 genes in WHRadjBMI that likely contribute to these traits by modulating the brain
487 transcriptome (**Supplementary Tables S13**). Unsurprisingly, 7 of 70 (10.00%) of the
488 neuroticism genes, 2 of 103 (1.94%) of the BD genes, 5 of 214 (2.34%) of the SCZ genes were
489 also identified by the depression TWAS, which reflects the genetic correlation among these traits.

490 By contrast, 0 of 5 (0.00%) of the BMI genes and 3 of 378 (0.79%) of the WHRadjBMI genes
491 overlapped with the 28 depression TWAS-significant genes.

492
493 Additionally, our results showed a trait-specific pattern for our analyses of inferred gene
494 expression, for example, 15/28 (53.6%) of depression genes, 159/214(74.3%) of SCZ genes,
495 323/378 (85.4%) of WHRadjBMI genes were found specific to itself (**Figure 5C**).

496

497 **Discussion**

498 In this study, we prioritized genes expressed in specific neocortical cell types and subtypes that
499 contribute to the pathogenesis of depression to accelerate the prioritization of new therapeutic
500 targets. We analyzed data from the DLPFC, a region of the brain that is an important hub for
501 mood-related circuits. We used multiple complementary approaches, including differential gene
502 expression analysis as well as methods integrating our data with GWAS results using FUSION,
503 SMR and HEIDI, and COLOC. Overall, leveraging the known genetic architecture of depression,
504 we identified 71 causal genes in depression and have identified those cell types in which the
505 functional consequences are occurring. 59 of these genes are novel compared to previous studies,
506 for example, *ANKRD36*, *PBRM1* and *FAM120AOS*. We found depression TWAS genes showed
507 a cell type-specific pattern, with the greatest enrichment being in neurons and astrocytes, in
508 comparison to Alzheimer's disease where we have reported an excess of microglial genes (46).
509 Depression TWAS genes were found to be more often shared with TWAS genes for traits that
510 are more genetically correlated (e.g., neuroticism) with depression when compared with traits
511 with lower genetic correlation (e.g., BMI).

512

513 Five genes (*MADD*, *TAOK3*, *SCAI*, *C11orf58* and *CHUK*) were both identified by the TWAS
514 and differential expression analysis. *MADD* plays a survival-promoting role against TNF
515 mediated apoptosis (33), it has been found associated with post-traumatic stress disorder (34). A
516 previous study found individuals suffering from bipolar disorder and schizophrenia showed that
517 a microdeletion that affects *TAOK3* (and *PEBPI*) is present in schizophrenia patients (31), and a
518 GWAS analysis suggested that *TAOK3* alone may contribute to neurodevelopmental disorders, at
519 least in schizophrenia (32). *SCAI* was found downregulated in several human tumors, and
520 decreased levels of *SCAI* are tightly correlated with increased invasive cell migration (36). It was
521 also found modest associated with major depressive disorders (37). *C11orf58* was found
522 associated with bipolar disorder in whole brain expression (47), and *CHUK* plays a key role in
523 the negative feedback of NF-kappa-B canonical signaling to limit inflammatory gene activation,
524 where the defective expression of NF- kappa-B has been proposed to play a role in the
525 development of depression (48).

526
527 In recent years, the target cell types in depression pathophysiology expanded from excitatory
528 neurons to include inhibitory interneurons (49) and non-neuronal cells (38, 50, 51). Previous
529 studies have reported that the greatest depression-associated differential gene expression
530 occurred in deep layer excitatory neurons and immature oligodendrocyte precursor cells (OPCs),
531 which contributed almost half (47%) of all changes in gene expression (52). The density and
532 form of cell abnormalities (in astrocytes, microglia, or oligodendrocytes) play an important role
533 in psychiatric disorders, including BD, depression, and SCZ (53). Here, we found depression,
534 SCZ and BD TWAS genes showed the greatest enrichment in excitatory neurons, followed by
535 inhibitory neurons, astrocytes, oligodendrocytes, OPC and microglia. In terms of cell subtypes,

536 excitatory neurons 3,4 and 12 showed the greatest enrichment of depression genes, as well as
537 inhibitory neurons 2, 6 and 15. Our results further confirmed the pre-eminent role of neurons in
538 depression but also highlights that at least some of the functional consequences of risk variants
539 are likely to be widely distributed among cortical cell subtypes, and our analyses provide
540 evidence that certain risk variants and gene expression may have opposite functional effects on
541 depression in different neuron subtypes. These data support the idea that there is functional
542 heterogeneity among neurons, and that many additional cell subtype resolved analyses will be
543 needed to fully map the functional consequences of depression susceptibility loci.

544
545 Despite of an excess of common TWAS genes between depression and other psychiatric
546 disorders (neuroticism, SCZ and BP) when compared to less genetically correlated traits such as
547 BMI and WHRadjBMI. We found >53% of TWAS genes were only identified in depression and
548 74.3% of TWAS genes were specific to SCZ, indicating genetic regulated gene expression
549 distinctions between psychiatric disorders. These findings could help with development of
550 treatment biomarkers targeting each specific psychiatric disorder.

551
552 In summary, this study illustrates the utility of large snucRNAseq data to uncover both genes
553 whose expression is altered in specific cell subtypes in the context of depression and to enhance
554 the interpretation of well-powered GWAS so that we can prioritize specific susceptibility genes
555 for further analysis.

556

557 **Acknowledgements and Disclosures**

558 The study was supported by NIH grants P30AG10161, P30AG72975, R01AG15819,
559 R01AG17917, U01AG46152, U01AG61356, RF1 AG057473, U01 AG072572. ROSMAP
560 resources can be requested at <https://www.radc.rush.edu>. The content is solely the responsibility
561 of the authors and does not necessarily represent the official views of the National Institutes of
562 Health.

563
564 LZ and PLD conceived and designed the study; LZ performed the research; MF, ZG, CCW and
565 GSG provided data and code contributed to the analyses; HUK, DAB, PAB, NH and VM
566 participated in the discussion and interpretation of the results; LZ and PLD wrote the paper; LZ
567 created the figures which were edited by PLD; All the authors reviewed and revised the paper.

568

569 Reference

- 570 1. Friedrich MJ. Depression Is the Leading Cause of Disability Around the World. *JAMA*.
571 2017;317(15):1517.
- 572 2. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence,
573 and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for
574 the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1545-602.
- 575 3. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark
576 road from association to function. *Am J Hum Genet*. 2013;93(5):779-97.
- 577 4. Gamazon ER, Segre AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas
578 of gene regulation across 44 human tissues to inform complex disease- and trait-associated
579 variation. *Nat Genet*. 2018;50(7):956-67.
- 580 5. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide
581 meta-analysis of depression identifies 102 independent variants and highlights the importance
582 of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343-52.
- 583 6. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for
584 large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48(3):245-52.
- 585 7. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary
586 data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*.
587 2016;48(5):481-7.
- 588 8. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al.
589 Bayesian test for colocalisation between pairs of genetic association studies using summary
590 statistics. *PLoS Genet*. 2014;10(5):e1004383.

- 591 9. Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and
592 findings from the rush Memory and Aging Project. *Curr Alzheimer Res.* 2012;9(6):646-63.
- 593 10. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. A multi-omic atlas of
594 the human frontal cortex for aging and Alzheimer's disease research. *Sci Data.* 2018;5:180142.
- 595 11. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et
596 al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices
597 pipeline. *Curr Protoc Bioinformatics.* 2013;43:11 0 1- 0 33.
- 598 12. Fujita M, Gao Z, Zeng L, McCabe C, White CC, Ng B, et al. Cell-subtype specific effects of
599 genetic variation in the aging and Alzheimer cortex. *bioRxiv.* 2022:2022.11.07.515446.
- 600 13. Anael Cain MT, Cristin McCabe, Idan Hekselman, Charles C. White, Gilad Green, Orit
601 Rozenblatt-Rosen, Feng Zhang, Esti Yeger-Lotem, David A. Bennett, Hyun-Sik Yang, Aviv Regev,
602 Vilas Menon, Naomi Habib, Philip L. De Jager. Multi-cellular communities are perturbed in the
603 aging human brain and with Alzheimer's disease. *bioRxiv.* 2020.
- 604 14. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
605 expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-40.
- 606 15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
607 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*
608 2015;43(7):e47.
- 609 16. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data
610 using empirical Bayes methods. *Biostatistics.* 2007;8(1):118-27.
- 611 17. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.
612 *Bioinformatics.* 2012;28(10):1353-8.
- 613 18. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al.
614 Genome-wide association analyses identify 44 risk variants and refine the genetic architecture
615 of major depression. *Nat Genet.* 2018;50(5):668-81.
- 616 19. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-
617 wide association study of more than 40,000 bipolar disorder cases provides new insights into
618 the underlying biology. *Nat Genet.* 2021;53(6):817-29.
- 619 20. Trubetskoy V, Pardinas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al.
620 Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature.*
621 2022;604(7906):502-8.
- 622 21. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis
623 of genome-wide association studies for neuroticism in 449,484 individuals identifies novel
624 genetic loci and pathways. *Nat Genet.* 2018;50(7):920-7.
- 625 22. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis
626 of genome-wide association studies for height and body mass index in approximately 700000
627 individuals of European ancestry. *Hum Mol Genet.* 2018;27(20):3641-9.
- 628 23. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis
629 of genome-wide association studies for body fat distribution in 694 649 individuals of European
630 ancestry. *Hum Mol Genet.* 2019;28(1):166-74.
- 631 24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network
632 analysis. *BMC Bioinformatics.* 2008;9:559.
- 633 25. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait
634 analysis. *Am J Hum Genet.* 2011;88(1):76-82.

- 635 26. Gusev A, Mancuso N, Won H, Kousi M, Finucane HK, Reshef Y, et al. Transcriptome-wide
636 association study of schizophrenia and chromatin activity yields mechanistic disease insights.
637 *Nat Genet.* 2018;50(4):538-48.
- 638 27. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for
639 functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic*
640 *Acids Res.* 2022.
- 641 28. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of
642 genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-41.
- 643 29. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A
644 global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
- 645 30. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning
646 heritability by functional annotation using genome-wide association summary statistics. *Nat*
647 *Genet.* 2015;47(11):1228-35.
- 648 31. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al. High
649 frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron.* 2011;72(6):951-63.
- 650 32. Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M, et al. Diverse types of
651 genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci.*
652 2012;15(12):1723-8.
- 653 33. Kurada B, Li LC, Mulherkar N, Subramanian M, Prasad KV, Prabhakar BS. MADD, a splice
654 variant of IG20, is indispensable for MAPK activation and protection against apoptosis upon
655 tumor necrosis factor-alpha treatment. *J Biol Chem.* 2009;284(20):13533-41.
- 656 34. Zhang Z, Meng P, Zhang H, Jia Y, Wen Y, Zhang J, et al. Brain Proteome-Wide Association
657 Study Identifies Candidate Genes that Regulate Protein Abundance Associated with Post-
658 Traumatic Stress Disorder. *Genes (Basel).* 2022;13(8).
- 659 35. Wingo TS, Liu Y, Gerasimov ES, Vattathil SM, Wynne ME, Liu J, et al. Shared mechanisms
660 across the major psychiatric and neurodegenerative diseases. *Nat Commun.* 2022;13(1):4314.
- 661 36. Brandt DT, Baarlink C, Kitzing TM, Kremmer E, Ivaska J, Nollau P, et al. SCAI acts as a
662 suppressor of cancer cell invasion through the transcriptional control of beta1-integrin. *Nat Cell*
663 *Biol.* 2009;11(5):557-68.
- 664 37. Mostafavi S, Battle A, Zhu X, Potash JB, Weissman MM, Shi J, et al. Type I interferon
665 signaling genes in recurrent major depression: increased expression detected by whole-blood
666 RNA sequencing. *Mol Psychiatry.* 2014;19(12):1267-74.
- 667 38. Pantazatos SP, Huang YY, Rosoklija GB, Dwork AJ, Arango V, Mann JJ. Whole-
668 transcriptome brain expression and exon-usage profiling in major depression and suicide:
669 evidence for altered glial, endothelial and ATPase activity. *Mol Psychiatry.* 2017;22(5):760-73.
- 670 39. Darby MM, Yolken RH, Sabunciyar S. Consistently altered expression of gene sets in
671 postmortem brains of individuals with major psychiatric disorders. *Transl Psychiatry.*
672 2016;6(9):e890.
- 673 40. Mahajan GJ, Vallender EJ, Garrett MR, Challagundla L, Overholser JC, Jurjus G, et al.
674 Altered neuro-inflammatory gene expression in hippocampus in major depressive disorder.
675 *Prog Neuropsychopharmacol Biol Psychiatry.* 2018;82:177-86.
- 676 41. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders
677 Study and Rush Memory and Aging Project. *J Alzheimers Dis.* 2018;64(s1):S161-S89.

- 678 42. Korologou-Linden R, Leyden GM, Relton CL, Richmond RC, Richardson TG. Multi-omics
679 analyses of cognitive traits and psychiatric disorders highlights brain-dependent mechanisms.
680 *Hum Mol Genet.* 2021.
- 681 43. Carboni L, Pischedda F, Piccoli G, Lauria M, Musazzi L, Popoli M, et al. Depression-
682 Associated Gene *Negr1-Fgfr2* Pathway Is Altered by Antidepressant Treatment. *Cells.* 2020;9(8).
- 683 44. Wingo TS, Liu Y, Gerasimov ES, Gockley J, Logsdon BA, Duong DM, et al. Brain proteome-
684 wide association study implicates novel proteins in depression pathogenesis. *Nat Neurosci.*
685 2021;24(6):810-7.
- 686 45. Li X, Su X, Liu J, Li H, Li M, and Me Research T, et al. Transcriptome-wide association
687 study identifies new susceptibility genes and pathways for depression. *Transl Psychiatry.*
688 2021;11(1):306.
- 689 46. Hansen DV, Hanson JE, Sheng M. Microglia in Alzheimer's disease. *J Cell Biol.*
690 2018;217(2):459-72.
- 691 47. McCarthy MJ, Liang S, Spadoni AD, Kelsoe JR, Simmons AN. Whole brain expression of
692 bipolar disorder associated genes: structural and genetic analyses. *PLoS One.*
693 2014;9(6):e100204.
- 694 48. Napetschnig J, Wu H. Molecular basis of NF-kappaB signaling. *Annu Rev Biophys.*
695 2013;42:443-68.
- 696 49. Northoff G, Sibille E. Why are cortical GABA neurons relevant to internal focus in
697 depression? A cross-level model linking cellular, biochemical and neural network findings. *Mol*
698 *Psychiatry.* 2014;19(9):966-77.
- 699 50. Nagy C, Suderman M, Yang J, Szyf M, Mechawar N, Ernst C, et al. Astrocytic
700 abnormalities and global DNA methylation patterns in depression and suicide. *Mol Psychiatry.*
701 2015;20(3):320-8.
- 702 51. Edgar N, Sibille E. A putative functional role for oligodendrocytes in mood regulation.
703 *Transl Psychiatry.* 2012;2:e109.
- 704 52. Nagy C, Maitra M, Tanti A, Suderman M, Theroux JF, Davoli MA, et al. Single-nucleus
705 transcriptomics of the prefrontal cortex in major depressive disorder implicates
706 oligodendrocyte precursor cells and excitatory neurons. *Nat Neurosci.* 2020;23(6):771-81.
- 707 53. Writing Committee for the Attention-Deficit/Hyperactivity D, Autism Spectrum D,
708 Bipolar D, Major Depressive D, Obsessive-Compulsive D, and Schizophrenia EWG, et al. Virtual
709 Histology of Cortical Thickness and Shared Neurobiology in 6 Psychiatric Disorders. *JAMA*
710 *Psychiatry.* 2021;78(1):47-63.

711

712

713

714

715

716

717

718

719

720

721

722

723 **Figure legends**

724 **Figure 1. Study overview**

725 We leveraged a large set of data derived from individual nuclei extracted from the DLPFCs of
726 424 genotyped older individuals. We integrated gene expression data and results of depression
727 GWAS analyses to perform transcriptome-wide association analysis (TWAS), mendelian
728 randomization (MR) and colocalization analysis (COLOC) in the 7 major cell types and 55 cell
729 subtypes of the DPLFPC. Differential gene expression analysis was applied to discover
730 additional depression associated genes expressed only in a subtype of cells.

731

732 **Figure 2. Example of depressive symptom-associated gene expression changes in cell**
733 **subtypes**

734 **A,B,C)** Scatterplot shows depression associated *TAOK3* gene expression changes in
735 oligodendrocyte subtype 12, astrocytes subtype 1 and excitatory neurons subtype 10, using the
736 last visit of depressive symptom score.

737 **D,E,F)** Scatterplot shows depression associated *SCAI* gene expression changes in inhibitory
738 neurons subtype 12, 2 and 10, using the average depressive symptom score over time.

739

740 **Figure 3. Depression genes identified from pseudo-bulk RNA-seq**

741 **A)** Number of depression TWAS genes identified from 7 cell types.

742 **B)** Number of depression TWAS genes that were unique to or shared between 7 cell types.

743 **C)** Statistical significance and effect directions of excitatory neurons TWAS of depression. Each
744 dot represents a gene. Positive and negative y coordinates show that transcript abundance was
745 associated with increased and decreased risk of depression, respectively. Novel and known
746 candidates for depression risk protein-coding genes are colored black and gray, respectively.
747 Genes in red color are TWAS genes filtered with SMR $p < 0.05$ & HEIDI $p > 0.05$. Red dashed
748 lines show genes with TWAS $FDR < 0.05$.

749 **D)** Heatmap reports the posterior probabilities of the H4 hypothesis (PP.H4) of the COLOC
750 method, which assumes GWAS and eQTL share a single causal SNP. Rows report overlap for
751 individual genes and SNP pair; columns report PP.H4 score in each of our cell types. The color
752 of each cell is based on the code found to the right of each panel; the darker color denotes higher
753 confidence that the same variant influences susceptibility and gene expression in that cell type.
754 Grey cells indicate that the gene was not an eQTL target in that cell type. Top bar chart shows
755 the number of colocalized eGenes with high confidence ($PP.H4 > 0.5$) in each cell type. Cells
756 highlighted with blue color indicates that the same gene was also detected by cell type-level
757 TWAS.

758

759 **Figure 4. Cell subtype-level TWAS of depression.**

760 **A)** Subtype-level pseudo-bulk expression and GWAS summary statistics of Howard et al., 2019
761 were used as input. Non-gray elements show 30 significant TWAS genes ($FDR < 0.05$), and their
762 colors represent effect directions and p-values. In row names, novel and known candidates for

763 depression risk genes are colored black and gray, respectively. Red asterisk indicates that the
764 same gene was also detected by cell-type level TWAS. Grey color represents the genes which are
765 not significant in TWAS analysis (TWAS FDR>0.05). White color denotes genes that did not
766 detected in the corresponding cell subtype. Genes were ordered alphabetically from bottom to
767 top.

768 **B)** GO terms and KEGG pathways enriched in the up-regulated genes associated with depression.

769 **C)** GO terms and KEGG pathways enriched in the down-regulated genes associated with
770 depression.

771 **Figure 5. Trait specificity of depression TWAS genes filtered with SMR $p<0.05$ & HEIDI**
772 **$p>0.05$**

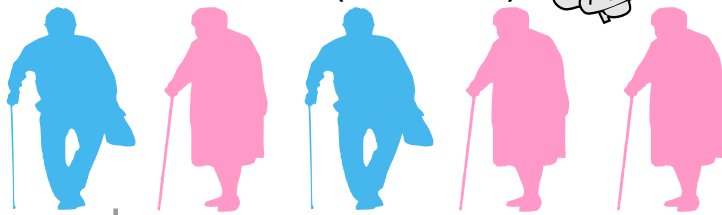
773 **A)** Number of TWAS, SMR and HEIDI genes identified from 7 cell types of pseudo-bulk RNA-
774 seq associated with six traits (bipolar disorder (BD), body mass index (BMI), major depressive
775 disorder (MDD), neuroticism, schizophrenia (SCZ), and waist-to-hip ratio adjusted for BMI
776 (WHRadjBMI)).

777 **B)** Genetic correlation estimated between depression and other diseases/traits. The areas of the
778 squares represent the absolute value of corresponding genetic correlations. After FDR correction
779 for 36 tests at 5% significance level, genetic correlations estimates that are significantly different
780 from zero are marked with a asterisk (* $0.01<FDR\leq 0.05$; ** $0.001<FDR\leq 0.01$;
781 *** $FDR\leq 0.001$). The blue color denotes a positive genetic correlation, the red color represents
782 a negative genetic correlation.

783 **C)** Number of TWAS, SMR and HEIDI genes that were unique to or shared between other
784 diseases/traits.

785

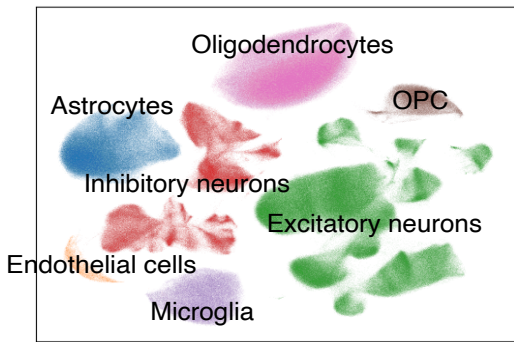
DLPFCs (n = 424)



medRxiv preprint doi: <https://doi.org/10.1101/2023.03.27.23286844>; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Single-nucleus RNA-seq

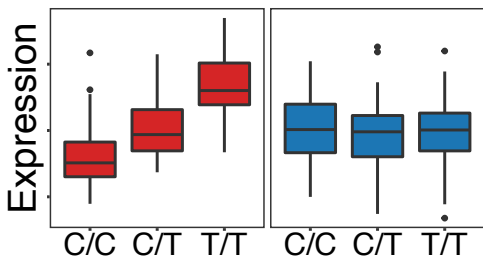
7 Cell types & 55 cell subtypes



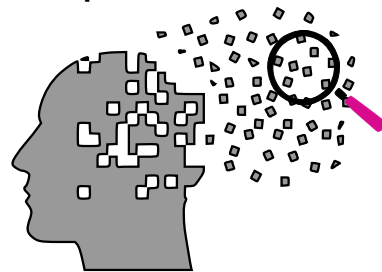
Genotyping

Pseudo-bulk gene expression

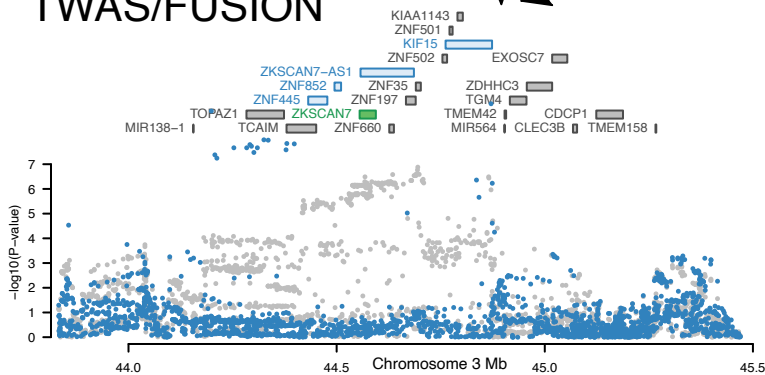
Cell type- & subtype-specific cis-eQTL



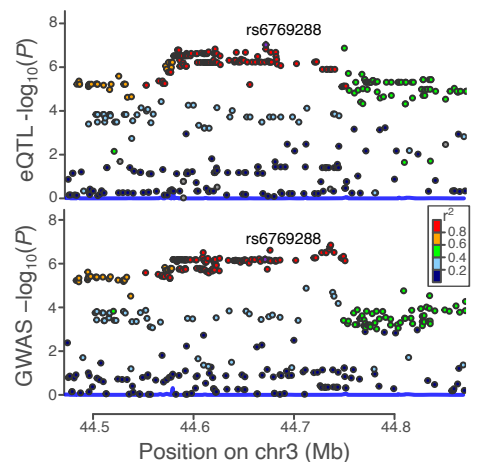
GWAS sumstats for Depression



TWAS/FUSION



COLOC



SMR & HEIDI

