Title

The impact of common variants on gene expression in the human brain: from RNA to protein to schizophrenia risk

Authors

Qiuman Liang[1†], Yi Jiang[1,2†], Annie W. Shieh[3], Dan Zhou[4,5], Rui Chen[5], Feiran Wang[1], Meng Xu[1], Mingming Niu[6], Xusheng Wang[7], Dalila Pinto[8,9,10,11], Yue Wang[12], Lijun Cheng[13], Ramu Vadukapuram[14], Chunling Zhang[15], Kay Grennan[16], Gina Giase[17], The PsychENCODE Consortium[‡], Kevin P White[18], Junmin Peng[6], Bingshan Li[5], Chunyu Liu[1,16,19*], Chao Chen[1,20,21,22*], Sidney H. Wang[3*]

Affiliations

1 Center for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, and Department of Psychiatry, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410000, China.

2 Department of Epidemiology and Biostatistics, Ministry of Education Key Laboratory of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430000, China.

19 3 Center for Human Genetics, The Brown foundation Institute of Molecular Medicine,
20 The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

21 4 School of Public Health and the Second Affiliated Hospital, Zhejiang University School
22 of Medicine, Hangzhou, Zhejiang 310058, China.

23 5 Department of Molecular Physiology and Biophysics, Vanderbilt Genetics Institute,
24 Vanderbilt University, Nashville, TN 37232, USA.

25 6 Department of Structural Biology, Department of Developmental Neurobiology, Center
26 for Proteomics and Metabolomics, St. Jude Children's Research Hospital, Memphis, TN 38105,
27 USA.

28 7 Department of Biology, University of North Dakota, Grand Forks, ND 58202, USA.

29 8 Department of Psychiatry, and Seaver Autism Center for Research and Treatment, Icahn
30 School of Medicine at Mount Sinai, New York, NY 10029, USA.

31 9 Department of Genetics and Genomic Sciences, and Icahn Institute for Data Science and
32 Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

33 10 The Mindich Child Health and Development Institute, Icahn School of Medicine at
34 Mount Sinai, New York, NY 10029, USA.

35 11 Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY
36 10029, USA.

37 12 Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and
38 State University, Arlington, VA 22203, USA.

13 Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL 60637, USA.

14 Department of Psychiatry, The University of Texas Rio Grande Valley, Harlingen, TX 78550, USA.

15 Department of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY 13210, USA.

16 Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY 13210, USA.

17 The Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA.

18 Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117596, Singapore.

19 School of Psychology, Shaanxi Normal University, Xi'an, Shaanxi 710062, China.

20 Furong Laboratory, Changsha, Hunan 410000, China.

21 National Clinical Research Center for Mental Disorders, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410000, China.

22 National Engineering Research Center of Personalized Diagnostic and Therapeutic Technology, Central South University, Changsha, Hunan 410000, China.

† These authors contributed equally to this work.

‡ PsychENCODE Consortium authors and affiliations are listed in the supplementary materials.

* Corresponding author email: **Hsi.Ming.S.Wang@uth.tmc.edu** (S. H. Wang), chenchaor@gmail.com (C. Chen), LiuCh@upstate.edu (C. Liu)

Abstract

The impact of genetic variants on gene expression has been intensely studied at the transcription level and the results have been useful in linking genes to disease risk for complex disorders, such as schizophrenia. On the other hand, the downstream impact of these variants and the molecular mechanisms connecting transcription variation to disease risk are not well understood. To this end, we quantitated ribosome occupancy in prefrontal cortex samples of the BrainGVEx cohort. Together with RNA-Seq and proteomics data from the same cohort, we performed *cis*-Quantitative Trait Locus (QTL) mapping to identify variants associated with transcription (eQTL), ribosome occupancy (rQTL), and protein (pQTL) level. We found that while more than half of pQTL effects originated from transcription variation, less than half of the eQTL effects propagated to the protein level. Leveraging multi-omics QTL results, we found 50 schizophrenia risk genes and further identified driver mechanisms for ten. We also found risk genes that have clear eQTL signals without discernible pQTLs.

Summary

78 Attenuation of eQTL effects at the protein level was found prevalent in the human brain
79 and its relevance in schizophrenia risk was investigated.

80 **Introduction**

81 Complex diseases such as neuropsychiatric disorders are multi-factorial with genetic components
82 (*1*, *2*). Large scale Genome-Wide Association Studies (GWAS) have uncovered thousands of
83 disease associated loci, signaling a promising era ahead for causal variant identification (*3*).
84 However, efforts in fine mapping these disease risk loci often narrowed down the underlying
85 causal signal to non-coding regions of the genome (*4–6*). Regulatory variants in such non-coding
86 regions are therefore the prime candidates for driver mutations in disease etiology. Consequently,
87 integrating gene expression information to pinpoint causal variants or to identify risk genes has
88 become a staple of genetic studies of complex diseases (*7*), with multiple consortia efforts
89 facilitating large scale gene expression profiling and regulatory element mapping (*8–10*). Many
90 powerful methods, such as *coloc*, PrediXcan, SMR/HEIDI, to name a few, have also been
91 developed to leverage gene expression information for fine mapping GWAS signals or for
92 identifying underlying risk genes (*11–13*).

93 Schizophrenia (SCZ) is a psychiatric disorder affecting ~1% of the world-wide population (*14*).
94 The heritability of SCZ has been estimated at between 60% to 80% indicating a strong genetic
95 component (*15*, *16*). Accordingly, the most recent schizophrenia GWAS study identified 287
96 significant risk loci and prioritized 120 risk genes using functional genomic data (*17*). The use of
97 RNA-Seq data from the brain was instrumental for risk gene prioritization by Trubetskoy et al.,
98 however, information from downstream gene regulation processes, such as translation rate and
99 protein abundance, was either not utilized or unavailable.

100 Measuring transcriptional changes as a proxy for gene activity has a long history in molecular
101 biology (*18*). In the context of human genetics, buffering of downstream effects of genetic
102 variants impacting gene expression (i.e. an eQTL) has been shown to be prevalent (*19*, *20*). In
103 addition, QTLs specific to protein level have also been reported (*21*). Together, these
104 observations indicated the importance and potential benefits of including downstream datatypes,
105 such as proteomics data, as information sources for fine mapping disease regulatory variants.
106 Indeed recent studies using multi-omics approaches have demonstrated increased power for risk
107 gene identification among other benefits (*22*, *23*). Of note, our recent work on genetic variants
108 associated with protein level in prefrontal cortex of the human brain indicated the extent of
109 contribution from non-synonymous coding variants to changes at protein level, and the utility of
110 these protein QTL variants in prioritizing GWAS risk genes for psychiatric disorders (*24*).

111 Another potential benefit of taking a multi-omics approach for identifying disease risk gene rests
112 in the potential to dissect the fine details of regulatory mechanisms driving the disease-genotype
113 association. Having relevant datasets to illuminate the origin and propagation of genetic impact
114 could potentially arrive at a conclusion of the driver regulatory process for a risk gene. Ribo-seq
115 is a technology that could be used to collect relevant data to fill in the gap between transcript and
116 protein expression. By adapting RNA-Seq to a ribosome footprinting method, ribo-seq provides
117 transcriptome-wide quantification of ribosome occupancy (*25*, *26*), which can serve as a proxy for
118 the amount of active translation produced from each transcript. When analyzed in conjunction
119 with RNA-Seq and quantitative proteomics data, ribo-seq enables identification of translational
120 and post-translational regulatory events (*27*), both major steps defining the human genetics aspect
121 of the central dogma of molecular biology.

122 As a part of the consortium efforts to improve our understanding of the genetic basis of
123 neuropsychiatric disorders (*28*), we generated multiple data modalities that included SNP

124 genotyping, RNA-Seq, ribo-seq, and proteomics of postmortem cortical tissue samples of the
125 BrainGVEX cohort, which altogether covered multiple expression levels from transcript to
126 protein. In conjunction with the quantitative proteomics and transcriptome profiling results that
127 we previously published (*24*, *29*), here we report our findings by integrating ribo-seq data to serve
128 as our operational definition for the level of protein translation across the transcriptome. Our
129 results reveal regulatory properties of common variants in the human brain and their utility in
130 identifying the regulatory processes driving disease risk for schizophrenia. Moreover, our results
131 demonstrated clear advantages of using multiple data modalities for studying disease risk
132 mechanisms.

133 **Results**

134 *Measuring transcriptome-wide ribosome occupancy level in prefrontal cortex of adult human*
135 *brain to quantify the level of protein translation*

136 To investigate regulatory impact of genetic variants in the prefrontal cortex of the human brain,
137 we performed ribosome profiling on 211 prefrontal cortex samples from the BrainGVEX
138 collection. In total we collected ~62 billion ribosome profiling reads. Consistent with the expected
139 ribosome footprint size, we found the average insert size of our ribo-seq libraries to range
140 between 27.4 and 29.5 nucleotides. Similar to prior published studies (*30*), we found on average
141 74 % of unwanted reads from ribosomal RNA, tRNA, and snoRNA, which contributed no
142 information to the translation of protein coding genes. After removing these unwanted reads, we
143 obtained an average of 30.3 million uniquely mapped informative reads per sample (inter-quartile
144 range: 20.5 ~ 37.6 million reads). When focusing our analysis on the informative reads, we found
145 our ribo-seq data to have ~3 times higher proportion of exon reads than that of the total RNA-Seq
146 data collected from the same individuals using an rRNA-depletion method (*29*). When visualized

147 in aggregate across annotated coding genes, we found our ribo-seq data to show strong sub-codon
148 periodicity at the expected positions (Fig. S1). High proportion of exon reads and strong sub-
149 codon periodicity reflects the enrichment of footprints from ribosomes actively engaged in
150 translating mature mRNA and indicates the quality of the dataset.

151 *Multi-omics cisQTL mapping identified candidate regulatory variants and revealed translational*
152 *and post-translational attenuation of eQTL effects*

153 To identify variants associated with inter-individual expression differences we perform cis-QTL
154 mapping for each data type independently. Using the full dataset (i.e. 416 RNA-Seq, 193 ribo-seq,
155 and 266 mass spec, which are coupled with the corresponding genotype data), we identified
156 11,857 eQTL (out of 16,540 genes we deem sufficiently quantitated), 3,086 rQTL (out of 25,506
157 genes), and 704 pQTL (out of 7,911 genes) at FDR < 0.1. The majority of the eQTL identified
158 here are replicated in the prefrontal cortex RNA-Seq data from the GTEx consortium (Table S1).
159 Intriguingly, we found large differences in the number of QTL between datatypes, suggesting that
160 some of the eQTL effects do not propagate all the way to the protein level. However, the
161 differences in the number of genes tested between datatypes make the comparison challenging to
162 interpret. To better compare genetic regulations captured by multi-omics QTLs derived from
163 multiple data modalities, we identified 175 samples with 5,080 genes that were sufficiently
164 quantitated across all three datatypes. Using this unified dataset, we found 2,140 eQTL, 840 rQTL,
165 and 246 pQTL at FDR<0.1 (Fig. 1A). Similar to the results derived from the full dataset, using
166 the unified dataset we found fewer significant QTLs as we move downstream the central dogma
167 of molecular biology.

168 A challenge in comparing the numbers of QTLs identified from each datatype rest in the fact that
169 not all true effects were identified. Tests replicating QTLs identified from one datatype in the

170 other datatypes can better capture the proportion of genetic effects shared between QTL types.
171 We performed replication tests using pi1 estimates from the qvalue method (*31*). Overall, we
172 found high proportion of QTLs replicated in other datatypes (Fig. 1B). However, when
173 considering the replication rate with the direction of genetic information flow, we found
174 asymmetric replication rates with the downstream datatypes to replicate less than the upstream
175 datatypes. More specifically, we found 90.4% rQTLs were replicated at the transcript level, but
176 only 65.3% eQTLs were replicated at the ribosome occupancy level. Moreover, while 76.8% of
177 pQTLs were replicated at the transcript level, only 24.1% of eQTLs were replicated at the protein
178 level (Fig. 1B). The lower percentage of eQTL and rQTL replicated at the protein level indicates
179 potential effect attenuation (i.e. either the inter-individual variation in gene expression becomes
180 smaller and therefore harder to detect or a lack of such effect in the downstream datatype).
181 Interestingly, in addition to the effect attenuation at the protein level, which was previously
182 reported for lymphoblastoid cell lines (LCLs) (*27*), here, using brain samples, we found further
183 asymmetry between eQTL and rQTL replication, indicating effect attenuation at the level of
184 translation. Similar asymmetry in proportion replicated between upstream and downstream
185 datatypes were observed when using a direction-aware cutoff-based approach at a wide range of
186 significance cutoffs used to define replication rates (Fig. S2).

187 While our replication tests revealed a trend of effect attenuation for eQTL variants in the
188 downstream phenotypes (Fig. 1B), these same observations could alternatively be explained by
189 differences in statistical power between technologies. An independent piece of evidence that is
190 not sensitive to measurement precision is needed to reach a solid conclusion. Using eQTLs
191 independently identified from prefrontal cortex samples by the CommonMind Consortium (*32*),
192 we avoided the ascertainment bias for large effect size eQTLs identified from our dataset and can
193 therefore directly compare effects size of eQTL variants between the three datatypes in our
194 dataset. A similar approach was successfully implemented to address this power issue in prior

195 work in LCLs (*27*). Using the 2,263 CMC eQTL that were also quantitated in our dataset, we
196 found the eQTL effect on the transcript expression level to be significantly larger than its effect
197 on ribosome occupancy level (per allele log2 fold differences: RNA 0.755 +/- 0.008 vs. ribo
198 0.680 +/- 0.010), which was in turn significantly larger than its effect on protein level (per allele
199 log2 fold differences: 0.640 +/- 0.012) (Fig. 1C, Wilcoxon rank sum test $P < 0.005$ for all
200 pairwise comparisons). By focusing on the effect size of independently identified eQTL, our
201 results strongly support the presence of downstream mechanisms attenuating eQTL effect both at
202 the ribosome occupancy level (translationally) and at the protein level (post-translationally).
203 Moreover, for the CommonMind Consortium eQTL, we found translational regulation to account
204 for more effect size reduction than post-translational regulation (Fig. 1C).

205 *Identifying datatype specific QTLs and their signal colocalization with schizophrenia GWAS*

206 The prevalent effect size reduction of eQTL raised the question of the relevance of these genetic
207 regulation at the organismal level. Because most of cellular tasks are executed by machineries
208 mainly composed of proteins, the genetic regulatory effects not reaching the protein level, either
209 attenuated by translational or post-translational regulation, could be less likely to have an impact
210 on organismal traits. To answer this question, we set out to investigate the relevance of datatype
211 specific QTL genes in SCZ. Consistent with prior reports (*33*), using our full dataset we found
212 significant proportion of SCZ heritability to be mediated by gene expression. By performing
213 mediated expression score regression (MESC) (*33*) on summary statistics from the Trubetskoy et
214 al. SCZ GWAS (*17*), we found our eQTL to mediate 7.04%, rQTL to mediate 2.96%, and pQTL
215 to mediate 1.99% of schizophrenia heritability.

216 Because regulation of a gene is often modulated by multiple genetic variants, to evaluate the
217 consequence of overall *cis*-QTL impact on gene expression we use PrediXcan to estimate

218 aggregate genetic regulatory effects for each gene. To distinguish between genes that have QTL
219 effects shared across multiple datatypes and genes that have datatype-specific QTL effects, for
220 each datatype we build PrediXcan models each with or without regressing out the other datatypes
221 and compute the correlation between the imputed expression from the two models. We termed
222 this correlation R_c. To identify R_c^2 values that reflect significant sharing between datatypes, we
223 permuted sample labels to emulate conditions of no real correlation between the three datatypes in
224 order to create empirical null distributions. Using both a false discovery rate (FDR), which is
225 calculated based on the empirical null, and an effect size cutoff based on R_c^2, we defined a set of
226 shared QTL genes and three sets of datatype-specific QTL genes. At 10% FDR, we defined
227 shared QTL genes by further requiring the R_c^2 to be smaller than 0.5. Using these criteria, from
228 the 633 genes that passed the minimum PrediXcan criteria for being included in this analysis (see
229 details in methods), we identified 164 shared QTL genes that have QTL effects shared between at
230 least two datatypes. For genes that failed to reject null at the 10% FDR cutoff (i.e. potentially
231 datatype-specific) we further set a conservative $R_c^2 > 0.9$ cutoff to define datatype-specific QTLs
232 for those QTLs that does not change after regressing out effects from other datatypes. Using these
233 criteria, we found 74 esQTL (mRNA-specific QTL) genes, 46 rsQTL (ribo-specific QTL) genes,
234 and 85 psQTL (protein-specific QTL) genes (Fig. S3, Table S2).

235 To investigate the relevance of datatype-specific QTL genes in schizophrenia, we performed
236 summary statistics-based signal colocalization between QTL signals and SCZ GWAS signals.
237 Using *coloc* with default prior (*11*), at a posterior probability cutoff of 70% for the signal
238 colocalization hypothesis, we found the esQTL of *CCDC117* and the rsQTL of *RAB15* to
239 colocalize with SCZ GWAS signals at loci *22q12.1* and *14q23.3*, respectively (Fig. 2). On the
240 other hand, for shared QTL genes, we found the eQTL of *RASA3*, the eQTL of *PCCB*, and both
241 the eQTL and the rQTL of *YWHAB* to colocalize with SCZ GWAS signals at loci *13q34*, *3q22.3*,
242 and *20q13.12*, respectively. Although we identified two datatype-specific QTLs that strongly

243 colocalized with SCZ GWAS, the low number of QTL-GWAS colocalization identified from this
244 approach limited our power in subsequent analyses to determine the overall relevance of datatype-
245 specific QTL in SCZ. To investigate more broadly the relevance of attenuated eQTLs in SCZ risk,
246 we next took a complementary approach by first identifying risk genes from each datatype
247 separately and then investigate the relevant regulatory processes driving these disease risks.

248 *Functional genomics identification of Schizophrenia risk genes*

249 Regulatory variants are prime candidates for mechanistic underpinning of GWAS signals. Given
250 the percentage of SCZ heritability mediated by gene regulation found in our dataset, here we
251 expand our effort to identify risk genes for SCZ using S-PrediXcan (*34*) based on our unified
252 multi-omic dataset. At 5% family-wise error rate, we found 37, 20, and 8 schizophrenia risk genes,
253 respectively from RNA-Seq, ribo-seq, and mass spectrometry data (Fig. 3A; note that the color
254 code of this figure will become relevant in the next results section). Among them only three genes,
255 *NEK4*, *KHK*, and *DARS2* were consistently identified as SCZ risk genes from all three datatypes
256 (Fig. 3B). The majority (76%) of the risk genes were identified only from one of the three
257 datatypes. This limited sharing in risk gene identification between datatypes is in clear contrast to
258 the amount of signal sharing found between QTL types (Fig. 1B). It would appear that while QTL
259 signals are often shared between datatypes, different datatypes have different power in explaining
260 GWAS signals at different loci.

261 Among the 50 risk genes we identified using S-PrediXcan (i.e. the union between the risk genes
262 identified from each of the three datatypes), 25 has previously been reported in GWAS studies as
263 either the mapped genes or as one of the nearby genes under the GWAS peak (Table S3) (*17*, *35*–
264 *37*). Of these prior reported genes, 12 matched the mapped genes while the remaining 13
265 nominated an alternative candidate gene for each GWAS locus (Note that for some of these 13

266 loci, the original GWAS signal was mapped to an intergenic region). Comparing our results to
267 other published SCZ risk gene identification studies, we found 8 matches to the risk genes
268 identified by Giambartolomei et al., which used RNA-Seq and DNA methylation data from
269 prefrontal cortex of the human brain (*22*) and 11 matches to the 120 prioritized SCZ risk genes
270 from Trubetskoy et al., which used colocalization with eQTL and Hi-C data (*17*). On the other
271 hand, for 18 of our 50 risk genes, we found no match to the known risk gene list, which we
272 compiled based on prior GWAS and functional genomics risk gene identification studies (*17*, *22*,
273 *35–37*). Similarly, using SZDB2.0, a database aggregating SCZ risk genes from the literature, we
274 found 40% of our risk genes are novel findings (*38*). These "no match" novel risk genes have
275 relatively weak SCZ GWAS signals and are therefore challenging to identify without the help
276 from our multi-omics QTL dataset. For example, we found strong colocalization between a
277 modest schizophrenia GWAS signal at *2p23.3* and all three types of molecular QTLs of the gene
278 *KHK* (Fig. 3D). Intriguingly, the *KHK* pQTL is in opposite direction of the *KHK* eQTL and rQTL
279 (Fig. 3C) indicating a linked post-translational process regulating the protein level in addition to
280 the transcriptional regulation.

281 *Analyses of multi-omics dataset reveal regulatory mechanisms of schizophrenia risk genes*

282 While PrediXcan is a powerful tool for GWAS risk gene identification, it does not control for
283 potential horizontal pleiotropy (*39*). To this end, we perform two-sample Mendelian
284 Randomization (MR) with Egger regression to replicate the risk genes we identified using S-
285 PrediXcan. Egger regression includes an intercept term, which can be used to evaluate the level of
286 horizontal pleiotropy (*40*). For each risk gene we first use LD-clumping (*41*) to identify top SNPs
287 from independent signals as strong genetic instruments (*42*). We then test for the causal
288 relationship between gene regulation (i.e. QTL signal) and SCZ (i.e. GWAS). We used an
289 operational definition of a causal effect based on the MR test results (see methods). At 10% FDR,

290 of the 65 gene-by-datatype combinations (i.e. a total of 50 risk genes including some discovered
291 from more than one datatype), 75% (49) passed the MR-Egger test and 25% (16) failed. Of those
292 25% that failed the MR-Egger test, 81% (13 out of 16) failed because of horizontal pleiotropy
293 identified by the Egger intercept test (Table S4). A total of 39 genes were replicated in at least one
294 of the three datatypes. Similar, but stronger, causal effect were observed when using SuSiE (the
295 Sum of Single Effects) (*43*) as an alternative approach to define instrument SNPs (Fig. S4).
296 However, when using SuSiE for selecting instrument variables, fewer genes were tested because
297 some of the risk genes have no fine mapped QTL SNPs according to the SuSiE parameters we
298 explored.

299 A key strength of using a multi-omic QTL approach to identify GWAS risk genes rests in the
300 possibility of further narrowing down the potential regulatory mechanisms. To this end, we
301 further examined the likely driver mechanisms for the 39 replicated SCZ risk genes using one-
302 sample Mendelian Randomization to infer causality between QTL types. We focused our analysis
303 on independently testing for causal effect between neighboring QTL types following the direction
304 of information flow of the central dogma (i.e. RNA -> ribo, and ribo -> protein). Here we used
305 fine-mapped QTL SNPs identified from the exposure datatype (i.e. the upstream datatype) as
306 instrument variable for one-sample MR analysis. Among the 39 two-sample MR replicated SCZ
307 risk genes, we found 10 genes with significant causal effects both from eQTL to rQTL (i.e. the
308 upstream pathway) and from rQTL to pQTL (i.e. the downstream pathway) (Fig. 3A dark red
309 color datapoints). Significant causal effects detected from both pathways indicates
310 transcriptionally regulated protein level differences as the potential driver for these risk genes in
311 SCZ etiology (see an example in Fig. 4A, Fig. 4B). On the other hand, 20 and 9 replicated risk
312 genes have either significant causal effects only in one of the two pathways (Fig. 3A blue color
313 datapoints) or have no significant causal effects (Fig. 3A orange color datapoints), respectively.

314 Of note, for the 20 risk genes that we found only one of the two pathways show significant causal
315 relationships, all of which were only significant in the upstream pathway (RNA -> ribo). This
316 asymmetry is reminiscent of the eQTL effect attenuation described in the prior sections. However,
317 a failed test could indicate either a lack of effect or a lack of statistical power. As an attempt to
318 distinguish between these two scenarios, we directly assessed the effect size, the noise level, and
319 instrument strength of the one-sample MR test results. When comparing the 10 genes that showed
320 significant causal relationships for both pathways to the 20 genes that only showed significant
321 causal relationships in the upstream pathway, we found significantly smaller effect in the
322 downstream pathway for the 20 genes that failed the test (Fig. S5), no significant difference were
323 found in standard error (Fig. S5); however, we found the QTL variants used as instrument for the
324 downstream pathway among the 20 genes that failed the MR test for the downstream pathway to
325 be significantly weaker than the QTL variants used as instrument for the 10 genes that passed the
326 MR test for both pathways (Fig. S5). This observation indicates that for many of these 20 genes,
327 the weak instrument effect likely led to an underestimated effect size and consequently these
328 genes failed the MR test. In other words, some of these genes could also operate through a
329 transcriptionally regulated protein level in schizophrenia disease etiology.

330 On the other hand, for the 9 genes that we do not find any significant causal relationships in either
331 of the two pathways, when compared to the 10 genes that we found significant causal effect from
332 both pathways, we saw significantly smaller effect size but no significant difference in standard
333 error nor in instrument strength (Fig. S6). These results indicate that comparing to the 10 genes
334 that passed both pathways, these 9 genes failed the Mendelian Randomization tests either because
335 they have much smaller causal effects (i.e. false negatives) or they have no real causal
336 relationships between datatypes. In other words, some of these genes could be datatype-specific
337 risk genes. Case in point, we found *PAK6* to have similar patterns in p value distributions between

338 eQTL and SCZ GWAS (Fig. 4D) but no clear QTL signal in either ribo-seq or quantitative mass
339 spectrometry data (Fig. 4C, Fig. 4D).

340 **Discussion**

341 Using a panel of postmortem prefrontal cortex samples, we found clear evidence of post-
342 transcriptional attenuation of eQTL effects in the human brain. Many of the differences found
343 between individual brain transcriptomes were not present between individual brain proteomes.
344 This observation echoes earlier work in HapMap lymphoblastoid cell lines (*27*) and extends the
345 prior conclusion from *in vitro* cell lines to complex human tissues. Importantly, distinct from the
346 earlier work in lymphoblastoid cell lines, which found translation to mostly track with
347 transcription, we found clear attenuation of eQTL effect in ribo-seq data indicating that
348 translational regulatory processes are involved in eQTL effect attenuation in the human brain.
349 Prevalent translational attenuation of variant impact on transcriptional gene expression levels has
350 previously been reported in budding yeast (*44*, *45*). However, results from follow up studies (*46*)
351 appear to present an inconsistent picture.

352 Here, using replication tests for Individual eQTL and testing for aggregate effect of eQTL
353 independently identified from CommonMind Consortium, we provided strong evidence
354 supporting prevalent translational attenuation of eQTL effects in the human brain. Although
355 operating at the molecular level, our study remained observational. Datatype-specific features
356 could potentially confound the results. For example, while mass spectrometry data aggregate
357 peptide count information by protein groups, sequencing data aggregates read count information
358 by genes. Mismatch between protein group and gene definition could result in mismatch in
359 quantification. Future work replicating this observation and elucidating molecular mechanisms of

360 translational attenuation of eQTL effects are needed to provide a clear understanding of the
361 phenomenon.

362 Following this observation, our current work focuses on the important question of whether the
363 attenuated eQTLs are still functionally (biologically) relevant. We attempted to address this
364 question by exploring the relevance of attenuated eQTL in SCZ, a neuropsychiatric disorder that
365 is highly heritable. We took two approaches to identify risk genes that have either datatype-
366 specific QTL signals or attenuated eQTL. An attenuated eQTL by definition is a genetic variant
367 that has significant association between its genotype and transcript level of the linked gene, but no
368 significant association in the downstream datatypes, either the ribosome occupancy or the protein
369 level (or both). In essence, attempts to identify attenuated eQTL (or datatype-specific QTL) are
370 dealing with the challenge of separating true negatives from false negatives. While we applied
371 very conservative criteria to identify candidate genes and performed subsequent evaluations, such
372 caveat is important to keep in mind when interpreting our results. Overall, from both approaches
373 we found clear examples of SCZ risk genes that appear to have attenuated eQTL. However, in
374 both cases, we were not able to evaluate whether datatype-specific QTLs were more likely to
375 explain SCZ risk than shared QTLs.

376 In the first approach, we used PrediXcan to aggregate variant impact at the gene level in order to
377 identify datatype-specific risk genes. Our method is distinct from published work on datatype
378 specific QTLs that took a SNP-based approach (*27*). Using a gene-based approach, we aimed to
379 increase the interpretability of the results and decrease the challenge of the needle in the haystack
380 problem. Indeed, our approach reduced our search space to 633 genes, and among these genes we
381 identified 213 datatype-specific QTL genes. These datatype-specific QTL genes show no
382 significant change after regressing out the other datatypes; they remain an R^2 higher than 0.9
383 between imputed expression levels generated from models with and without conditioning on the

384 other datatypes. Follow up mechanistic studies on these datatype-specific QTL genes could
385 provide insights on the nature of datatype-specific genetic regulation. From the same analysis we
386 also identified 164 shared QTL genes. Note that the number of shared and datatype-specific QTL
387 genes found here cannot be compared directly. By requiring the models built based on regressed
388 data to pass certain quality metric, our method filtered out genes with strong sharing between
389 QTL types from the analysis. The number of shared QTL genes identified here is therefore
390 underrepresenting the actual count. In the end, this approach produced five schizophrenia risk
391 genes based on signal colocalization. The limited number of genes identified prevented us from
392 performing the subsequent comparison between datatype-specific QTL genes and shared QTL
393 genes.

394 In the second approach, we first identified risk genes using a TWAS approach, S-PrediXcan
395 method. We then replicated the TWAS risk genes using two-sample Mendelian Randomization
396 with Egger intercept tests. While the majority of the TWAS risk genes were replicated in two-
397 sample Mendelian Randomization tests, more than 80% of the failed replication tests failed
398 because of the Egger intercept test. Our results, therefore, clearly confirmed the need of better
399 controlling for potential horizontal pleiotropy in TWAS studies (*40*). About 40% of the risk genes
400 that we identified here have not been previously reported as SCZ risk genes. These novel SCZ
401 risk genes tend to have weak GWAS signals and are therefore challenging to identify without the
402 help of functional genomics data (see Fig. 3D for an example). Finally, for each replicated risk
403 genes we further tested causality between QTL types using One-Sample MR to identify causal
404 regulatory mechanisms. We found ten risk genes that are likely contributing to SCZ risk through
405 transcriptionally regulated protein level. On the other hand, we also identify nine genes that show
406 no significant MR tests and 20 genes that only have significant MR test results from one of the
407 two pathways tested. The interpretation of the relative numbers here, however, is challenging.
408 While these failed MR tests appear to have significantly smaller causal effect, many of these tests

409 also used weaker instrument variables, which could underestimate the causal effect and resulted
410 in false negatives. With the current dataset, we unfortunately do not have enough data points to
411 estimate the proportion of false negative in our results. On the other hand, we do see interesting
412 examples supporting the possibility of SCZ risk driven by variants impacting transcriptional gene
413 regulation that does not propagate to the ribosome occupancy level or the protein level (Fig. 4C).
414 We hope by presenting the current results, our findings can inspire future studies on this topic.

415 An alternative, and perhaps more practical, way of looking at the relevance of attenuated eQTL is
416 to ask whether it is still relevant to use transcriptome profiling for risk gene identification. Since
417 many of the eQTL effects do not appear to affect protein level, should we collectively switch to
418 profile downstream protein level directly to maximize our ability in risk gene identification. Our
419 results clearly reject this idea. Likely because of the differences in the number of genes covered
420 across datatypes, our RNA-Seq data, which covers more genes than the other two datatypes,
421 explained more than twice the percentage of SCZ heritability explained by our ribo-seq or protein
422 data. When looking at the number of risk genes identified by using S-PrediXcan, we again found
423 the most risk genes from analyzing RNA-Seq data. Importantly, when comparing the risk genes
424 identified across the three datatypes, we found limited overlaps. In other words, at current sample
425 size, each datatype appears to be more powerful in finding risk genes at different loci, and by
426 collecting all three, we benefited from the complementary results. Case in point, because
427 proteomics data were not used in the functional genomics studies of SCZ risk genes that we
428 compared our results to, five out of the eight risk genes identified using protein data in this study
429 were not reported in the previous studies (i.e. on the "no match" list). Whether such inter-datatype
430 differences in risk gene identification would persist with increasing sample size remains an open
431 question. However, at the current scale (i.e. a sample size of ~200) there are clear benefits from
432 profiling all three data types for risk gene identification.

433 **Materials and Methods**

434 **Data sources**

435 The SNP genotypes (*24*), RNA-Seq (*29*), and quantitative mass spectrometry (*24*) data generated
436 for prefrontal cortex tissue samples of the BrainGVEX cohort were downloaded from the
437 PsychENCODE Synapse portal (https://www.synapse.org/#!Synapse:syn4590909) (See Table S5
438 for a summary on the number of samples in each dataset and their respective overlap with the
439 samples in the genotype data). The BrainGVEX cohort includes 420 Caucasians, 2 Hispanics, 1
440 African American, 3 Asian Americans, and 14 unknown-origin individuals. The DRAMS
441 software was used to detect and correct mixed-up samples (*47*). We also used PGC3 SCZ GWAS
442 summary statistics data obtained from the Psychiatric Genomics Consortium (*17*).

443 **Ribosome profiling**

444 Ribosome Profiling experiments were performed using Illumina's TrueSeq Ribo Profile
445 (Mammalian) Kit. TrueSeq Ribo Profile (Mammalian) Kit was developed for cell lines. We
446 adapted the protocol to frozen tissue samples with MP Biomedicals™ FastPrep-24™ Classic
447 Bead Beating Grinder and Lysis System. Specifically, 60-80 mg of frozen brain tissue was
448 homogenized in Lysing Matrix D tubes containing 800 µl polysome lysis buffer. The Lysing
449 tubes were placed on the FastPrep-24™ homogenizer set at 4.0 m/s with 20 s increments until no
450 visible chunks of tissue remained. Tissue lysate was incubated on ice for 10 min followed by
451 centrifugation at 20,000g at 4 °C for 10 minutes to pellet cell debris. Clarified lysate was
452 subsequently used for ribo-seq library preparation following TrueSeq Ribo Profile (Mammalian)
453 Kit instructions. Indexed libraries were pooled and then sequenced on an Illumina HiSeq 4000.
454 Note that the experimental protocol for TrueSeq Ribo Profile (Mammalian) Kit that we followed
455 is a modified version of the previous ribo-seq protocol published by Ingolia and colleagues (*26*),

456 and it has the following key modifications: Monosome isolation was performed using Sephacryl
457 S400 spin columns (GE; 27–5140-01) on a tabletop centrifuge instead of ultra-high speed
458 centrifugation in sucrose cushion. Ribosomal RNA depletion was carried out by using Ribo-Zero
459 Magnetic Kits and this step is moved up to right after the monosomal RNA isolation step and
460 before the Ribosome Protected Fragment gel isolation step.

461 **Data processing, gene expression quantification, and normalization**

462 For RNA-Seq data, we obtained the FASTQ raw data from the PsychENCODE BrainGVEX
463 project. Then we used cutadapt (v1.12) to trim adapter for fastq data with code "cutadapt –
464 minimum-length=5 –quality-base=33 -q 30 -a
465 AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A
466 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT". Then we mapped trimmed fastq onto
467 GENCODE Release 19 (GRCh37.p13) genome with same version's GTF by STAR (v2.4.2a).
468 Then we used RSEM software to quantify the read counts for each gene (*48*). The cpm() function
469 in the R package "limma" was used to calculate the log-transformed CPM (Counts per million).
470 We filtered out the genes with CPM < 1 in more than 75% samples and the samples with network
471 connectivity (*49*) z score < -5 (Fig. S7), which resulted in a total of 17,207 genes from 423
472 samples in the quantification table. We then used normalize.quantiles() function in the R package
473 "preprocessCore" (*50*) to normalize expression level for each sample.

474 For ribo-seq data we used cutadapt (v1.12) to trim adapter for fastq data with code "cutadapt -a
475 AGATCGGAAGAGCACACGTCT –quality-base=33 –minimum-length=25 –discard-
476 untrimmed". The trimmed reads were then mapped against a FASTA file of rRNA, tRNA, and
477 snoRNA sequence using bowtie2 (*51*) to filter out uninformative reads. The filtered reads were
478 then mapped to GENCODE Release 19 (GRCh37.p13) genome with corresponding transcript

479 model GTF file using STAR (v2.4.2a). The uniquely mapped reads, as defined by the NH:i:1 flag
480 of the alignment files, were kept for subsequent analysis. We used the featureCounts() function in
481 the R package "subreads" (*52*) to calculate gene level read counts for ribosome occupancy. The
482 cpm() function in the R package "limma" was used to calculate log-transformed CPM value. We
483 filtered out genes with CPM < 0.1 in more than 90% samples and the samples with network
484 connectivity (*49*) z score < -3.5 (Fig. S8), which resulted in a total of 26,552 genes quantitated
485 from 197 samples in the quantification table. We then used normalize.quantiles() function in the
486 R package "preprocessCore" (*50*) to normalize ribosome occupancy level for each sample.

487 For quantitative mass spectrometry data, we obtained protein quantification table from the
488 PsychENCODE BrainGVEX Synapse portal (https://www.synapse.org/#!Synapse:syn26017684).
489 This table includes abundance quantification for 11,433 proteins from 266 samples. The data
490 processing steps for producing the mass spectrometry quantification table is detailed in Luo et al.
491 (*24*). We further log-transformed and quantile normalized protein level quantification for each
492 sample. We matched the protein ID to gene ID according to the UCSC database, which resulted in
493 8,223 genes.

494 **QTL mapping**

495 *Estimating and adjusting for unwanted factors*

496 We used the R package "PEER" (*53*) to estimate hidden factors for RNA-Seq, ribo-seq, and
497 quantitative mass spectrometry data separately. The principle of selecting unwanted hidden
498 factors was to maximize the variance explained with the least number of factors. We identified 30,
499 20, and 6 hidden factors to remove from RNA-Seq, ribo-seq, and mass spectrometry data,
500 respectively (Fig. S9). For each gene from each datatype, we adjusted the expression level by
501 fitting the selected hidden factors as predictors in a linear model and taking the residuals as the

502 adjusted expression level. The adjusted expression levels were then further centered by mean and
503 scaled by standard deviation.

504 *Genotype association tests*

505 We identified *cis*-region expression QTL (*cis*-eQTL), ribo-seq QTL (*cis*-rQTL), and protein QTL
506 (*cis*-pQTL) separately using QTLTools (1.3.1) (*54*). For each gene, we defined cis-region as the
507 region ranging from 1Mb upstream of the 5' end of the gene to 1Mb downstream of the 3' end of
508 the gene (i.e. the length of the gene body plus 2 Mb in size). We tested all common SNPs (MAF >
509 0.05) within the cis-region using 10,000 permutations to create empirical null distributions and
510 used the beta-approximation approach implemented in QTLTools to estimate the empirical p
511 values. For each gene, we selected the SNP with the most significant empirical p value from the
512 genotype-expression level association tests to represent the QTL signal. To calculate a genome-
513 wide FDR, we used qvalue() function of the qvalue R package for multiple testing correction and
514 set a qvalue < 0.1 (i.e. 10% FDR) cutoff to identify significant QTLs.

515 **Estimating mediated SCZ heritability of each omics**

516 To estimate the proportion of SCZ heritability mediated by different omics, we used MESC (*33*).
517 In the first step, we calculated the overall expression score using the unwanted-factor-adjusted-
518 expression data (see QTL mapping section) as the individual-level gene expression data with the
519 corresponding BrainGVEX genotype data and the 1000 genome phase 3 genotype data as the
520 ancestry matched genotype data for GWAS. In the second step, we used the overall expression
521 score calculated in the previous step and PGC3 SCZ GWAS summary statistics to estimate SCZ
522 heritability mediated by each omics.

523 **Identifying shared and datatype-specific QTL genes**

524 *Building PrediXcan gene expression prediction models*

525 We used the PrediXcan software (*13*) to separately build gene expression prediction models for
526 RNA-Seq, ribo-seq, and quantitative mass spectrometry. For each gene, an Elastic Net algorithm
527 was used for feature selection from SNPs located within the cis-region defined for each gene (i.e.
528 gene body +/- 1Mb flanking regions) based on results from a ten-fold cross-validation. After that,
529 weights were produced for every selected SNP of each gene, which were used in the prediction
530 models. For each gene, we calculated the Pearson correlation between predicted and observed
531 gene expression (Cross-validation R, R_{cv}), which was considered as a metric for prediction
532 accuracy. Only the genes with $R_{cv} > 0.1$ and $P < 0.05$ were retained. We produced prediction
533 models for the unified set of 5,080 genes and 175 samples shared across omics.

534 *Building conditionally independent prediction models*

535 To identify shared and datatype-specific QTL genes among different omics, we built
536 conditionally independent prediction models for each omics. For each gene, we regressed from
537 the data the genetic regulation signals of all other omics types (i.e. the imputed quantification
538 level of all other omics types). The models were built based on the regressed expression (i.e. the
539 residual from the regression). Note that because in many cases multiple protein isoform
540 quantifications match to one gene, when building model for gene-based quantification datatypes,
541 such as mRNA and ribosome occupancy, we aggregate data across all protein isoforms for each
542 gene. In contrast, when building model for protein data, quantifications between isoforms were
543 kept separate.

544 More specifically, we took the following ten-step procedure to identify shared-QTL genes and to
545 define datatype-specific QTL genes:

546 Step 1: For each gene, we built original prediction models using equation 1.

547 $$E_k = \sum_{i=0}^{n} w_i S_i + \varepsilon \quad (1)$$

548 Where E_k denotes the observed expression of omics type k; w_i denotes the weight of the i-th SNP
549 S_i; n denotes the genotype dosage of selected SNPs by the elastic net.

550 Step 2: For each datatype, we then applied the respective model to the genotype data and
551 calculated the imputed expression $\widehat{E_k} = \sum_{i=0}^{n} w_i S_i$.

552 Step 3: For each gene in datatype k, we regressed the imputed expression of the other datatypes
553 from the observed expression using equation 2 and kept the residual.

554 $$E_k = \beta_1 \widehat{E_1} + \beta_2 \widehat{E_2} + \cdots + \beta_{l-1} \widehat{E_{l-1}} + e_k \quad (2)$$

555 Where β_l denotes slope of the *l*-th other datatypes, $\widehat{E_l}$ denotes the imputed expression of the *l*-th
556 other datatypes, and e_k denotes the residual for datatype k.

557 Step 4: For each gene in the datatype k, the expression value conditioning on the genetic
558 regulation of all the other datatypes was calculated as follow.

559 $$R_k = \overline{E_k} + e_k \quad (3)$$

560 Step 5: After that, we can build the conditional prediction models (i.e. estimate the weight
561 parameters) using equation 4.

562 $$R_k = \sum_{i=0}^{n} m_i S_i + \varepsilon \quad (4)$$

563 Where m_i is the weight of i-th SNP S_i for the conditional prediction model.

564 Step 6: We calculated the square of correlation of imputed gene expression between the original
565 model and conditional model and named it as observed R_c^2.

566 Step 7: To determine whether a gene share genetic regulation signals with other datatypes, we
567 used permutation to create a null distribution of R_c^2. We permuted the sample label for imputed
568 expression of the other datatypes, and repeated step 3 and step 4 to build the conditional
569 prediction model using permuted data, which we named conditional permutation model. We then
570 calculated permutation R_c^2 of imputed gene expression between the original model and
571 conditional permutation model.

572 Step 8: We performed 50 permutations and calculated empirical p values of the observed R_c^2
573 based on its rank among the permutation R_c^2 (i.e. ordered from the smallest to the greatest).

574 Step 9: We used Benjamini-Hochberg control procedure to adjust the empirical p values to
575 calculate FDR for shared-QTL genes.

576 Step 10: We defined datatype-specific QTL genes as genes with FDR > 0.1 and R_c^2 >0.9, and the
577 shared-QTL genes as genes with FDR <0.1 and R_c^2 <0.5.

578 **Colocalization**

579 We used *coloc* (*11*) to detect signal colocalization between SCZ GWAS and each QTL type at the
580 cis-region of each QTL gene. For each QTL gene, for all common SNPs (MAF > 0.05) within the
581 cis-region, we use QTL effect and GWAS summary statistics as input for coloc analysis. For QTL
582 effect, we calculated slope and square of slope' standard error from a linear model fit using SNP
583 genotype as the predictor of gene quantifications. We then used coloc.abf() function in the R

584 package "*coloc*" (*11*) to calculate the posterior possibility of each hypothesis using the default
585 prior. We use posterior probability of 70% for the colocalization hypothesis (i.e. PPH4) as the
586 cutoff for reporting our colocalization findings.

587 **TWAS identification of SCZ risk genes**

588 Gene-level association tests were performed for SCZ using S-PrediXcan (*34*) based on the
589 prediction models built using our omics data, which is described in the "building prediction
590 models" section, and the SCZ GWAS summary statistics data from PGC3 (*17*). The association
591 tests were done separately for each datatype. For protein data, we performed omnibus test to
592 incorporate p values of all protein isoforms together to calculate a single p value for the
593 corresponding gene. The family wise error rates for SCZ risk genes were calculated using
594 Bonferroni correction of nominal p values.

595 **Two-sample Mendelian randomization (MR)**

596 To identify causal relationships between each datatype and SCZ we used MR analysis. Here we
597 used fine mapped QTL SNPs as instruments, gene expression quantification at each datatype as
598 exposure, and SCZ GWAS signal as outcome.

599 More specifically, we took the following steps to test for causal relationship between gene
600 regulation at each omics level and SCZ:

601 Step 1: Here we used two methods of selecting instrument SNPs. For each gene of each QTL type,
602 we performed LD clumping by PLINK with "–clump-kb 1000 –clump-r2 0.5" parameters to
603 select instrument SNPs with p value < 0.05. Another method is fine-mapping by susie_rss()
604 function with default parameters in R package "SuSiE" (*43*).

605 Step 2: For each gene, we used harmonise_data() function in R package "TwoSampleMR" (*55*) to
606 harmonize QTL of each datatype and SCZ GWAS SNPs to be in the same direction (i.e. effect
607 relative to the same allele).

608 Step 3: We then performed Univariable Mendelian randomization (UVMR) for each gene in each
609 datatype separately. UVMR analysis was done using mr() function, which includes IVW and
610 Egger methods, in the R package "TwoSampleMR" (*55*).

611 Step 4: We used the intercept test (i.e. the Egger method) to test for horizontal pleiotropy, and
612 used predictor coefficient and its corresponding p value from IVW to determine the effect size
613 and significance of causal effects for each datatype on SCZ.

614 Step 5: We used Benjamini-Hochberg to adjust for multiple testing.

615 Step 6: We define a gene by datatype combination as causal for SCZ if the causal effect test
616 passed the FDR < 0.1 cutoff and the Egger intercept test passed the intercept p value > 0.05 cutoff.

617 **One-sample Mendelian randomization**

618 For one-sample MR we used two-stage least squares (2SLS) approach to find causal relationships
619 between datatypes. We performed the following analysis in two iterations, both following the
620 direction of genetic information flow. In the first iteration, we tested causal relationships between
621 transcript level and ribosome occupancy level (i.e. RNA -> ribo). In the second iteration, we
622 tested causal relationships between ribosome occupancy level and protein level (i.e. ribo ->
623 protein).

624 Step 1: We used the same two methods for instrument SNP selection as described in the previous
625 section. Here we tested two pathways (mRNA -> ribo and ribo -> protein). For mRNA -> ribo, we

626 used the SNPs of eQTL with p value < 0.05 and LD clumping by PLINK with "–clump-kb 1000 –
627 clump-r2 0.5" parameters. For ribo -> protein, we used the SNPs of rQTL with p value < 0.05 and
628 LD clumping by PLINK with "–clump-kb 1000 –clump-r2 0.5".

629 Step 2: We then combined genotype and quantification data of the relevant datatypes into a
630 dataframe: dat_{mr} for RNA -> ribo pathway, dat_{rp} for ribo -> protein pathway.

631 Step 3: For each gene, we then set formula in R package "ivreg": ivreg("ribo ~ mRNA |
632 SNP_1+SNP_2+SNP_3+…+SNP_n", data = dat_{mr}) and ivreg("protein ~ ribo |
633 SNP_1+SNP_2+SNP_3+…+SNP_m", data = dat_{rp}).

634 Step 4: We used the summary() function to get slope, p value of slope, intercept, p value of
635 intercept, F-statistic and p value of F-statistic of each ivreg object. P value of Intercept was used
636 to test horizontal pleiotropy, F-statistic was used to check instrument strength.

637 Step 5: We used Benjamini-Hochberg control procedure to adjust for multiple testing.

638 Step 6: We defined a causal relationship for each gene by pathway combination as the causal
639 effect test passed the FDR < 0.1 cutoff and the Egger intercept test passed the intercept p value >
640 0.05 cutoff.

641 **References**

642 1. A. Fialkowski, T. M. Beasley, H. K. Tiwari, "[11 – Multifactorial Inheritance and Complex Diseases]" in
643 *Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics (Seventh Edition)*, R. E.
644 Pyeritz, B. R. Korf, W. W. Grody, Eds. (Academic Press, 2019), pp. 323–358.

645 2. A. S. Brown, The environment and susceptibility to schizophrenia. *Prog. Neurobiol*. 93, 23–58 (2011).

646 3. N. Solovieff, C. Cotsapas, P. H. Lee, S. M. Purcell, J. W. Smoller, Pleiotropy in complex traits: challenges and
647 strategies. *Nat. Rev. Genet.* 14, 483–495 (2013).

648 4. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic
649 associations with FUMA. *Nat. Commun.* 8, 1826 (2017).

650 5. M. G. Heckman, C. Labbé, A. L. Kolicheski, A. I. Soto-Beasley, R. L. Walton, R. R. Valentino, E. R. Brennan,
651 P. W. Johnson, S. Baheti, V. Sarangi, Y. Ren, R. J. Uitti, Z. K. Wszolek, O. A. Ross, Fine-mapping of the non-
652 coding variation driving the Caucasian LRRK2 GWAS signal in Parkinson‘s disease. *Parkinsonism Relat.*
653 *Disord*. 83, 22–30 (2021).

654 6. D. J. Schaid, W. Chen, N. B. Larson, From genome-wide associations to candidate causal variants by statistical
655 fine-mapping. *Nat. Rev. Genet.* 19, 491–504 (2018).

656 7. A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. J. H. Penninx, R. Jansen, E. J. C. de Geus, D. I.
657 Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimäki, E.
658 Raitoharju, M. Kähönen, I. Seppälä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, B.
659 Pasaniuc, Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48, 245–
660 252 (2016).

661 8. M. Fromer, P. Roussos, S. K. Sieberts, J. S. Johnson, D. H. Kavanagh, T. M. Perumal, D. M. Ruderfer, E. C.
662 Oh, A. Topol, H. R. Shah, L. L. Klei, R. Kramer, D. Pinto, Z. H. Gümüş, A. E. Cicek, K. K. Dang, A. Browne,
663 C. Lu, L. Xie, B. Readhead, E. A. Stahl, J. Xiao, M. Parvizi, T. Hamamsy, J. F. Fullard, Y.-C. Wang, M. C.
664 Mahajan, J. M. J. Derry, J. T. Dudley, S. E. Hemby, B. A. Logsdon, K. Talbot, T. Raj, D. A. Bennett, P. L. De
665 Jager, J. Zhu, B. Zhang, P. F. Sullivan, A. Chess, S. M. Purcell, L. A. Shinobu, L. M. Mangravite, H.
666 Toyoshiba, R. E. Gur, C.-G. Hahn, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum, E.
667 E. Schadt, K. Hirai, K. Roeder, K. J. Brennand, N. Katsanis, E. Domenici, B. Devlin, P. Sklar, Gene expression
668 elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci*. 19, 1442–1453 (2016).

669 9. ENCODE Project Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T.
670 Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He,

671 M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J.
672 Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri,
673 J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J.
674 Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham,
675 J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E.
676 Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded
677 encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 583, 699–710 (2020).

678 10. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*.
679 369, 1318–1330 (2020).

680 11. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian
681 test for colocalisation between pairs of genetic association studies using summary statistics. *PloS Genet*. 10,
682 e1004383 (2014).

683 12. Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R.
684 Wray, P. M. Visscher, J. Yang, Integration of summary data from GWAS and eQTL studies predicts complex
685 trait gene targets. *Nat. Genet*. 48, 481–487 (2016).

686 13. E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J.
687 C. Denny, D. L. Nicolae, N. J. Cox, H. K. Im, A gene-based association method for mapping traits using
688 reference transcriptome data. *Nat. Genet*. 47, 1091–1098 (2015).

689 14. R. S. Kahn, I. E. Sommer, R. M. Murray, A. Meyer-Lindenberg, D. R. Weinberger, T. D. Cannon, M. O
690 'Donovan, C. U. Correll, J. M. Kane, J. van Os, T. R. Insel, Schizophrenia. *Nat. Rev. Dis. Primers*. 1, 15067
691 (2015).

692 15. P. Lichtenstein, B. H. Yip, C. Björk, Y. Pawitan, T. D. Cannon, P. F. Sullivan, C. M. Hultman, Common
693 genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study.
694 *Lancet*. 373, 234–239 (2009).

16. P. F. Sullivan, K. S. Kendler, M. C. Neale, Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 60, 1187–1192 (2003).

17. V. Trubetskoy, A. F. Pardiñas, T. Qi, G. Panagiotaropoulou, S. Awasthi, T. B. Bigdeli, J. Bryois, C.-Y. Chen, C. A. Dennison, L. S. Hall, M. Lam, K. Watanabe, O. Frei, T. Ge, J. C. Harwood, F. Koopmans, S. Magnusson, A. L. Richards, J. Sidorenko, Y. Wu, J. Zeng, J. Grove, M. Kim, Z. Li, G. Voloudakis, W. Zhang, M. Adams, I. Agartz, E. G. Atkinson, E. Agerbo, M. Al Eissa, M. Albus, M. Alexander, B. Z. Alizadeh, K. Alptekin, T. D. Als, F. Amin, V. Arolt, M. Arrojo, L. Athanasiu, M. H. Azevedo, S. A. Bacanu, N. J. Bass, M. Begemann, R. A. Belliveau, J. Bene, B. Benyamin, S. E. Bergen, G. Blasi, J. Bobes, S. Bonassi, A. Braun, R. A. Bressan, E. J. Bromet, R. Bruggeman, P. F. Buckley, R. L. Buckner, J. Bybjerg-Grauholm, W. Cahn, M. J. Cairns, M. E. Calkins, V. J. Carr, D. Castle, S. V. Catts, K. D. Chambert, R. C. K. Chan, B. Chaumette, W. Cheng, E. F. C. Cheung, S. A. Chong, D. Cohen, A. Consoli, Q. Cordeiro, J. Costas, C. Curtis, M. Davidson, K. L. Davis, L. de Haan, F. Degenhardt, L. E. DeLisi, D. Demontis, F. Dickerson, D. Dikeos, T. Dinan, S. Djurovic, J. Duan, G. Ducci, F. Dudbridge, J. G. Eriksson, L. Fañanás, S. V. Faraone, A. Fiorentino, A. Forstner, J. Frank, N. B. Freimer, M. Fromer, A. Frustaci, A. Gadelha, G. Genovese, E. S. Gershon, M. Giannitelli, I. Giegling, P. Giusti-Rodríguez, S. Godard, J. I. Goldstein, J. González Peñas, A. González-Pinto, S. Gopal, J. Gratten, M. F. Green, T. A. Greenwood, O. Guillin, S. Gülöksüz, R. E. Gur, R. C. Gur, B. Gutiérrez, E. Hahn, H. Hakonarson, V. Haroutunian, A. M. Hartmann, C. Harvey, C. Hayward, F. A. Henskens, S. Herms, P. Hoffmann, D. P. Howrigan, M. Ikeda, C. Iyegbe, I. Joa, A. Julià, A. K. Kähler, T. Kam-Thong, Y. Kamatani, S. Karachanak-Yankova, O. Kebir, M. C. Keller, B. J. Kelly, A. Khrunin, S.-W. Kim, J. Klovins, N. Kondratiev, B. Konte, J. Kraft, M. Kubo, V. Kučinskas, Z. A. Kučinskiene, A. Kusumawardhani, H. Kuzelova-Ptackova, S. Landi, L. C. Lazzeroni, P. H. Lee, S. E. Legge, D. S. Lehrer, R. Lencer, B. Lerer, M. Li, J. Lieberman, G. A. Light, S. Limborska, C.-M. Liu, J. Lönnqvist, C. M. Loughland, J. Lubinski, J. J. Luykx, A. Lynham, M. Macek, A. Mackinnon, P. K. E. Magnusson, B. S. Maher, W. Maier, D. Malaspina, J. Mallet, S. R. Marder, S. Marsal, A. R. Martin, L. Martorell, M. Mattheisen, R. W. McCarley, C. McDonald, J. J. McGrath, H. Medeiros, S. Meier, B. Melegh, I. Melle, R. I. Mesholam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, M. Mitjans, E. Molden, E. Molina, M. D. Molto, V. Mondelli, C. Moreno, C. P. Morley, G. Muntané, K. C. Murphy, I. Myin-Germeys, I. Nenadić, G. Nestadt, L. Nikitina-Zake, C. Noto, K. H. Nuechterlein, N. L. O'Brien, F. A. O'Neill, S.-Y. Oh, A. Olincy, V. K. Ota, C. Pantelis, G. N. Papadimitriou, M. Parellada, T. Paunio, R. Pellegrino, S. Periyasamy, D. O. Perkins, B. Pfuhlmann, O. Pietiläinen, J. Pimm, D. Porteous, J. Powell, D. Quattrone, D.

724 Quested, A. D. Radant, A. Rampino, M. H. Rapaport, A. Rautanen, A. Reichenberg, C. Roe, J. L. Roffman, J.
725 Roth, M. Rothermundt, B. P. F. Rutten, S. Saker-Delye, V. Salomaa, J. Sanjuan, M. L. Santoro, A. Savitz, U.
726 Schall, R. J. Scott, L. J. Seidman, S. I. Sharp, J. Shi, L. J. Siever, E. Sigurdsson, K. Sim, N. Skarabis, P.
727 Slominsky, H.-C. So, J. L. Sobell, E. Söderman, H. J. Stain, N. E. Steen, A. A. Steixner-Kumar, E. Stögmann,
728 W. S. Stone, R. E. Straub, F. Streit, E. Strengman, T. S. Stroup, M. Subramaniam, C. A. Sugar, J. Suvisaari, D.
729 M. Svrakic, N. R. Swerdlow, J. P. Szatkiewicz, T. M. T. Ta, A. Takahashi, C. Terao, F. Thibaut, D. Toncheva,
730 P. A. Tooney, S. Torretta, S. Tosato, G. B. Tura, B. I. Turetsky, A. Üçok, A. Vaaler, T. van Amelsvoort, R. van
731 Winkel, J. Veijola, J. Waddington, H. Walter, A. Waterreus, B. T. Webb, M. Weiser, N. M. Williams, S. H.
732 Witt, B. K. Wormley, J. Q. Wu, Z. Xu, R. Yolken, C. C. Zai, W. Zhou, F. Zhu, F. Zimprich, E. C. Atbaşoğlu,
733 M. Ayub, C. Benner, A. Bertolino, D. W. Black, N. J. Bray, G. Breen, N. G. Buccola, W. F. Byerley, W. J.
734 Chen, C. R. Cloninger, B. Crespo-Facorro, G. Donohoe, R. Freedman, C. Galletly, M. J. Gandal, M. Gennarelli,
735 D. M. Hougaard, H.-G. Hwu, A. V. Jablensky, S. A. McCarroll, J. L. Moran, O. Mors, P. B. Mortensen, B.
736 Müller-Myhsok, A. L. Neil, M. Nordentoft, M. T. Pato, T. L. Petryshen, M. Pirinen, A. E. Pulver, T. G.
737 Schulze, J. M. Silverman, J. W. Smoller, E. A. Stahl, D. W. Tsuang, E. Vilella, S.-H. Wang, S. Xu, R.
738 Adolfsson, C. Arango, B. T. Baune, S. I. Belangero, A. D. Børglum, D. Braff, E. Bramon, J. D. Buxbaum, D.
739 Campion, J. A. Cervilla, S. Cichon, D. A. Collier, A. Corvin, D. Curtis, M. D. Forti, E. Domenici, H.
740 Ehrenreich, V. Escott-Price, T. Esko, A. H. Fanous, A. Gareeva, M. Gawlik, P. V. Gejman, M. Gill, S. J. Glatt,
741 V. Golimbet, K. S. Hong, C. M. Hultman, S. E. Hyman, N. Iwata, E. G. Jönsson, R. S. Kahn, J. L. Kennedy, E.
742 Khusnutdinova, G. Kirov, J. A. Knowles, M.-O. Krebs, C. Laurent-Levinson, J. Lee, T. Lencz, D. F. Levinson,
743 Q. S. Li, J. Liu, A. K. Malhotra, D. Malhotra, A. McIntosh, A. McQuillin, P. R. Menezes, V. A. Morgan, D. W.
744 Morris, B. J. Mowry, R. M. Murray, V. Nimgaonkar, M. M. Nöthen, R. A. Ophoff, S. A. Paciga, A. Palotie, C.
745 N. Pato, S. Qin, M. Rietschel, B. P. Riley, M. Rivera, D. Rujescu, M. C. Saka, A. R. Sanders, S. G. Schwab, A.
746 Serretti, P. C. Sham, Y. Shi, D. St Clair, H. Stefánsson, K. Stefansson, M. T. Tsuang, J. van Os, M. P. Vawter,
747 D. R. Weinberger, T. Werge, D. B. Wildenauer, X. Yu, W. Yue, P. A. Holmans, A. J. Pocklington, P. Roussos,
748 E. Vassos, M. Verhage, P. M. Visscher, J. Yang, D. Posthuma, O. A. Andreassen, K. S. Kendler, M. J. Owen,
749 N. R. Wray, M. J. Daly, H. Huang, B. M. Neale, P. F. Sullivan, S. Ripke, J. T. R. Walters, M. C. O'Donovan,
750 Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 604, 502–508 (2022).

751 18. J. B. W. Wolf, Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial.
752 *Mol Ecol Resour*. 13, 559–572 (2013).

753 19. L. Shang, J. A. Smith, W. Zhao, M. Kho, S. T. Turner, T. H. Mosley, S. L. R. Kardia, X. Zhou, Genetic
754 Architecture of Gene Expression in European and African Americans: An eQTL Mapping Study in GENOA.
755 *Am. J. Hum. Genet*. 106, 496–512 (2020).

756 20. U. Võsa, A. Claringbould, H.-J. Westra, M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber,
757 S. Yazar, H. Brugge, R. Oelen, D. H. de Vries, M. G. P. van der Wijst, S. Kasela, N. Pervjakova, I. Alves, M.-J.
758 Favé, M. Agbessi, M. W. Christiansen, R. Jansen, I. Seppälä, L. Tong, A. Teumer, K. Schramm, G. Hemani, J.
759 Verlouw, H. Yaghootkar, R. Sönmez Flitman, A. Brown, V. Kukushkina, A. Kalnapenkis, S. Rüeger, E. Porcu,
760 J. Kronberg, J. Kettunen, B. Lee, F. Zhang, T. Qi, J. A. Hernandez, W. Arindrarto, F. Beutner, BIOS
761 Consortium, i2QTL Consortium, J. Dmitrieva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, A. W.
762 Hewitt, M. Kähönen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei,
763 Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M. G. Nivard, B. W. J. H. Penninx, J. K. Pritchard, O. T.
764 Raitakari, O. Rotzschke, E. P. Slagboom, C. D. A. Stehouwer, M. Stumvoll, P. Sullivan, P. A. C. 't Hoen, J.
765 Thiery, A. Tönjes, J. van Dongen, M. van Iterson, J. H. Veldink, U. Völker, R. Warmerdam, C. Wijmenga, M.
766 Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, E. Dermitzakis, S. Bergmann, T.
767 Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. L. Pierce, T. Lehtimäki, D. I. Boomsma, B. M. Psaty, S. A.
768 Gharib, P. Awadalla, L. Milani, W. H. Ouwehand, K. Downes, O. Stegle, A. Battle, P. M. Visscher, J. Yang, M.
769 Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, Large-scale cis- and trans-eQTL analyses identify thousands
770 of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet*. 53, 1300–1310 (2021).

771 21. E. Ferkingstad, P. Sulem, B. A. Atlason, G. Sveinbjornsson, M. I. Magnusson, E. L. Styrmisdottir, K.
772 Gunnarsdottir, A. Helgason, A. Oddsson, B. V. Halldorsson, B. O. Jensson, F. Zink, G. H. Halldorsson, G.
773 Masson, G. A. Arnadottir, H. Katrinardottir, K. Juliusson, M. K. Magnusson, O. T. Magnusson, R.
774 Fridriksdottir, S. Saevarsdottir, S. A. Gudjonsson, S. N. Stacey, S. Rognvaldsson, T. Eiriksdottir, T. A.
775 Olafsdottir, V. Steinthorsdottir, V. Tragante, M. O. Ulfarsson, H. Stefansson, I. Jonsdottir, H. Holm, T. Rafnar,
776 P. Melsted, J. Saemundsdottir, G. L. Norddahl, S. H. Lund, D. F. Gudbjartsson, U. Thorsteinsdottir, K.
777 Stefansson, Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet*. 53, 1712–
778 1721 (2021).

779 22. C. Giambartolomei, J. Zhenli Liu, W. Zhang, M. Hauberg, H. Shi, J. Boocock, J. Pickrell, A. E. Jaffe,
780 CommonMind Consortium, B. Pasaniuc, P. Roussos, A Bayesian framework for multiple trait colocalization
781 from summary association statistics. *Bioinformatics*. 34, 2538–2545 (2018).

782 23. C. N. Foley, J. R. Staley, P. G. Breen, B. B. Sun, P. D. W. Kirk, S. Burgess, J. M. M. Howson, A fast and
783 efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat.*
784 *Commun*. 12, 764 (2021).

785 24. J. Luo, M. Niu, L. Li, D. Kong, Y. Jiang, A. Shieh, L. Cheng, G. Giase, K. Grennan, K. White, C. Chen, S.
786 Wang, D. Pinto, Y. Wang, C. Liu, J. Peng, X. Wang, https://doi.org/10.21203/rs.3.rs-1633422/v1 (2022).

787 25. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, Genome-wide analysis in vivo of
788 translation with nucleotide resolution using ribosome profiling. *Science*. 324, 218–223 (2009).

789 26. N. T. Ingolia, G. A. Brar, S. Rouskin, A. M. McGeachy, J. S. Weissman, The ribosome profiling strategy for
790 monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc*. 7,
791 1534–1550 (2012).

792 27. A. Battle, Z. Khan, S. H. Wang, A. Mitrano, M. J. Ford, J. K. Pritchard, Y. Gilad, Impact of regulatory
793 variation from RNA to protein. *Science*. 347, 664–667 (2015).

794 28. S. Akbarian, C. Liu, J. A. Knowles, F. M. Vaccarino, P. J. Farnham, G. E. Crawford, A. E. Jaffe, D. Pinto, S.
795 Dracheva, D. H. Geschwind, J. Mill, A. C. Nairn, A. Abyzov, S. Pochareddy, S. Prabhakar, S. Weissman, P. F.
796 Sullivan, M. W. State, Z. Weng, M. A. Peters, K. P. White, M. B. Gerstein, A. Amiri, C. Armoskus, A. E.
797 Ashley-Koch, T. Bae, A. Beckel-Mitchener, B. P. Berman, G. A. Coetzee, G. Coppola, N. Francoeur, M.
798 Fromer, R. Gao, K. Grennan, J. Herstein, D. H. Kavanagh, N. A. Ivanov, Y. Jiang, R. R. Kitchen, A.
799 Kozlenkov, M. Kundakovic, M. Li, Z. Li, S. Liu, L. M. Mangravite, E. Mattei, E. Markenscoff-Papadimitriou,
800 F. C. P. Navarro, N. North, L. Omberg, D. Panchision, N. Parikshak, J. Poschmann, A. J. Price, M. Purcaro, T.
801 E. Reddy, P. Roussos, S. Schreiner, S. Scuderi, R. Sebra, M. Shibata, A. W. Shieh, M. Skarica, W. Sun, V.
802 Swarup, A. Thomas, J. Tsuji, H. van Bakel, D. Wang, Y. Wang, K. Wang, D. M. Werling, A. J. Willsey, H.

803 Witt, H. Won, C. C. Y. Wong, G. A. Wray, E. Y. Wu, X. Xu, L. Yao, G. Senthil, T. Lehner, P. Sklar, N. Sestan,
804 The PsychENCODE project. *Nat. Neurosci*. 18, 1707–1712 (2015).

805 29. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese,
806 Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y.
807 Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof, J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe,
808 PsychENCODE Consortium, M. A. Peters, M. Gerstein, C. Liu, L. M. Iakoucheva, D. Pinto, D. H. Geschwind,
809 Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 362,
810 eaat8127 (2018).

811 30. G. A. Brar, J. S. Weissman, Ribosome profiling reveals the what, when, where and how of protein synthesis.
812 *Nat. Rev. Mol. Cell Biol*. 16, 651–664 (2015).

813 31. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* 100,
814 9440–9445 (2003).

815 32. G. E. Hoffman, J. Bendl, G. Voloudakis, K. S. Montgomery, L. Sloofman, Y.-C. Wang, H. R. Shah, M. E.
816 Hauberg, J. S. Johnson, K. Girdhar, L. Song, J. F. Fullard, R. Kramer, C.-G. Hahn, R. Gur, S. Marenco, B. K.
817 Lipska, D. A. Lewis, V. Haroutunian, S. Hemby, P. Sullivan, S. Akbarian, A. Chess, J. D. Buxbaum, G. E.
818 Crawford, E. Domenici, B. Devlin, S. K. Sieberts, M. A. Peters, P. Roussos, CommonMind Consortium
819 provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci. Data*. 6, 180 (2019).

820 33. D. W. Yao, L. J. O'Connor, A. L. Price, A. Gusev, Quantifying genetic effects on disease mediated by assayed
821 gene expression levels. *Nat. Genet*. 52, 626–633 (2020).

822 34. A. N. Barbeira, S. P. Dickinson, R. Bonazzola, J. Zheng, H. E. Wheeler, J. M. Torres, E. S. Torstenson, K. P.
823 Shah, T. Garcia, T. L. Edwards, E. A. Stahl, L. M. Huckins, D. L. Nicolae, N. J. Cox, H. K. Im, Exploring the
824 phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics.
825 *Nat. Commun*. 9, 1825 (2018).

35. S. Ripke, B. M. Neale, A. Corvin, J. T. R. Walters, K.-H. Farh, P. A. Holmans, P. Lee, B. Bulik-Sullivan, D. A. Collier, H. Huang, T. H. Pers, I. Agartz, E. Agerbo, M. Albus, M. Alexander, F. Amin, S. A. Bacanu, M. Begemann, R. A. Belliveau Jr, J. Bene, S. E. Bergen, E. Bevilacqua, T. B. Bigdeli, D. W. Black, R. Bruggeman, N. G. Buccola, R. L. Buckner, W. Byerley, W. Cahn, G. Cai, D. Campion, R. M. Cantor, V. J. Carr, N. Carrera, S. V. Catts, K. D. Chambert, R. C. K. Chan, R. Y. L. Chen, E. Y. H. Chen, W. Cheng, E. F. C. Cheung, S. Ann Chong, C. Robert Cloninger, D. Cohen, N. Cohen, P. Cormican, N. Craddock, J. J. Crowley, D. Curtis, M. Davidson, K. L. Davis, F. Degenhardt, J. Del Favero, D. Demontis, D. Dikeos, T. Dinan, S. Djurovic, G. Donohoe, E. Drapeau, J. Duan, F. Dudbridge, N. Durmishi, P. Eichhammer, J. Eriksson, V. Escott-Price, L. Essioux, A. H. Fanous, M. S. Farrell, J. Frank, L. Franke, R. Freedman, N. B. Freimer, M. Friedl, J. I. Friedman, M. Fromer, G. Genovese, L. Georgieva, I. Giegling, P. Giusti-Rodríguez, S. Godard, J. I. Goldstein, V. Golimbet, S. Gopal, J. Gratten, L. de Haan, C. Hammer, M. L. Hamshere, M. Hansen, T. Hansen, V. Haroutunian, A. M. Hartmann, F. A. Henskens, S. Herms, J. N. Hirschhorn, P. Hoffmann, A. Hofman, M. V. Hollegaard, D. M. Hougaard, M. Ikeda, I. Joa, A. Julià, R. S. Kahn, L. Kalaydjieva, S. Karachanak-Yankova, J. Karjalainen, D. Kavanagh, M. C. Keller, J. L. Kennedy, A. Khrunin, Y. Kim, J. Klovins, J. A. Knowles, B. Konte, V. Kucinskas, Z. Ausrele Kucinskiene, H. Kuzelova-Ptackova, A. K. Kähler, C. Laurent, J. Lee Chee Keong, S. Hong Lee, S. E. Legge, B. Lerer, M. Li, T. Li, K.-Y. Liang, J. Lieberman, S. Limborska, C. M. Loughland, J. Lubinski, J. Lönnqvist, M. Macek Jr, P. K. E. Magnusson, B. S. Maher, W. Maier, J. Mallet, S. Marsal, M. Mattheisen, M. Mattingsdal, R. W. McCarley, C. McDonald, A. M. McIntosh, S. Meier, C. J. Meijer, B. Melegh, I. Melle, R. I. Mesholam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, Y. Mokrab, D. W. Morris, O. Mors, K. C. Murphy, R. M. Murray, I. Myin-Germeys, B. Müller-Myhsok, M. Nelis, I. Nenadic, D. A. Nertney, G. Nestadt, K. K. Nicodemus, L. Nikitina-Zake, L. Nisenbaum, A. Nordin, E. O ‘Callaghan, C. O‘Dushlaine, F. A. O‘Neill, S.-Y. Oh, A. Olincy, L. Olsen, J. Van Os, C. Pantelis, G. N. Papadimitriou, S. Papiol, E. Parkhomenko, M. T. Pato, T. Paunio, M. Pejovic-Milovancevic, D. O. Perkins, O. Pietiläinen, J. Pimm, A. J. Pocklington, J. Powell, A. Price, A. E. Pulver, S. M. Purcell, D. Quested, H. B. Rasmussen, A. Reichenberg, M. A. Reimers, A. L. Richards, J. L. Roffman, P. Roussos, D. M. Ruderfer, V. Salomaa, A. R. Sanders, U. Schall, C. R. Schubert, T. G. Schulze, S. G. Schwab, E. M. Scolnick, R. J. Scott, L. J. Seidman, J. Shi, E. Sigurdsson, T. Silagadze, J. M. Silverman, K. Sim, P. Slominsky, J. W. Smoller, H.-C. So, ChrisC. A. Spencer, E. A. Stahl, H. Stefansson, S. Steinberg, E. Stogmann, R. E. Straub, E. Strengman, J. Strohmaier, T. Scott Stroup, M. Subramaniam, J. Suvisaari, D. M. Svrakic, J. P. Szatkiewicz, E. Söderman, S. Thirumalai, D. Toncheva, S. Tosato, J. Veijola, J. Waddington, D. Walsh, D. Wang, Q. Wang, B. T. Webb, M.

856 Weiser, D. B. Wildenauer, N. M. Williams, S. Williams, S. H. Witt, A. R. Wolen, E. H. M. Wong, B. K.
857 Wormley, H. Simon Xi, C. C. Zai, X. Zheng, F. Zimprich, N. R. Wray, K. Stefansson, P. M. Visscher, W.
858 Trust Case-Control Consortium, R. Adolfsson, O. A. Andreassen, D. H. R. Blackwood, E. Bramon, J. D.
859 Buxbaum, A. D. Børglum, S. Cichon, A. Darvasi, E. Domenici, H. Ehrenreich, T. Esko, P. V. Gejman, M. Gill,
860 H. Gurling, C. M. Hultman, N. Iwata, A. V. Jablensky, E. G. Jönsson, K. S. Kendler, G. Kirov, J. Knight, T.
861 Lencz, D. F. Levinson, Q. S. Li, J. Liu, A. K. Malhotra, S. A. McCarroll, A. McQuillin, J. L. Moran, P. B.
862 Mortensen, B. J. Mowry, M. M. Nöthen, R. A. Ophoff, M. J. Owen, A. Palotie, C. N. Pato, T. L. Petryshen, D.
863 Posthuma, M. Rietschel, B. P. Riley, D. Rujescu, P. C. Sham, P. Sklar, D. St Clair, D. R. Weinberger, J. R.
864 Wendland, T. Werge, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Psychosis
865 Endophenotypes International Consortium, Biological insights from 108 schizophrenia-associated genetic loci.
866 *Nature*. 511, 421–427 (2014).

867 36. A. F. Pardiñas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D.
868 Cameron, M. L. Hamshere, J. Han, L. Hubbard, A. Lynham, K. Mantripragada, E. Rees, J. H. MacCabe, S. A.
869 McCarroll, B. T. Baune, G. Breen, E. M. Byrne, U. Dannlowski, T. C. Eley, C. Hayward, N. G. Martin, A. M.
870 McIntosh, R. Plomin, D. J. Porteous, N. R. Wray, A. Caballero, D. H. Geschwind, L. M. Huckins, D. M.
871 Ruderfer, E. Santiago, P. Sklar, E. A. Stahl, H. Won, E. Agerbo, T. D. Als, O. A. Andreassen, M. Bækvad-
872 Hansen, P. B. Mortensen, C. B. Pedersen, A. D. Børglum, J. Bybjerg-Grauholm, S. Djurovic, N. Durmishi, M.
873 G. Pedersen, V. Golimbet, J. Grove, D. M. Hougaard, M. Mattheisen, E. Molden, O. Mors, M. Nordentoft, M.
874 Pejovic-Milovancevic, E. Sigurdsson, T. Silagadze, C. S. Hansen, K. Stefansson, H. Stefansson, S. Steinberg, S.
875 Tosato, T. Werge, GERAD1 Consortium, CRESTAR Consortium, D. A. Collier, D. Rujescu, G. Kirov, M. J.
876 Owen, M. C. O'Donovan, J. T. R. Walters, Common schizophrenia alleles are enriched in mutation-intolerant
877 genes and in regions under strong background selection. *Nat. Genet*. 50, 381–389 (2018).

878 37. M. Lam, C.-Y. Chen, Z. Li, A. R. Martin, J. Bryois, X. Ma, H. Gaspar, M. Ikeda, B. Benyamin, B. C. Brown, R.
879 Liu, W. Zhou, L. Guan, Y. Kamatani, S.-W. Kim, M. Kubo, A. A. A. A. Kusumawardhani, C.-M. Liu, H. Ma,
880 S. Periyasamy, A. Takahashi, Z. Xu, H. Yu, F. Zhu, Schizophrenia Working Group of the Psychiatric
881 Genomics Consortium, Indonesia Schizophrenia Consortium, Genetic Research on schizophreniA neTwork-
882 China and the Netherlands (GREAT-CN), W. J. Chen, S. Faraone, S. J. Glatt, L. He, S. E. Hyman, H.-G. Hwu,
883 S. A. McCarroll, B. M. Neale, P. Sklar, D. B. Wildenauer, X. Yu, D. Zhang, B. J. Mowry, J. Lee, P. Holmans,

884 S. Xu, P. F. Sullivan, S. Ripke, M. C. O'Donovan, M. J. Daly, S. Qin, P. Sham, N. Iwata, K. S. Hong, S. G.
885 Schwab, W. Yue, M. Tsuang, J. Liu, X. Ma, R. S. Kahn, Y. Shi, H. Huang, Comparative genetic architectures
886 of schizophrenia in East Asian and European populations. *Nat. Genet*. 51, 1670–1678 (2019).

887 38. Y. Wu, X. Li, J. Liu, X.-J. Luo, Y.-G. Yao, SZDB2.0: an updated comprehensive resource for schizophrenia
888 research. *Hum. Genet*. 139, 1285–1297 (2020).

889 39. D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, E. R. Gamazon, A unified framework for joint-tissue
890 transcriptome-wide association and Mendelian randomization analysis. *Nat. Genet*. 52, 1239–1246 (2020).

891 40. S. Burgess, S. G. Thompson, Interpreting findings from Mendelian randomization using the MR-Egger method.
892 *Eur. J. Epidemiol*. 32, 377–389 (2017).

893 41. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de
894 Bakker, M. J. Daly, P. C. Sham, PLINK: A Tool Set for Whole-Genome Association and Population-Based
895 Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575 (2007).

896 42. A. Gkatzionis, S. Burgess, P. J. Newcombe, Statistical methods for cis-Mendelian randomization with two-
897 sample summary-level data. *Genet. Epidemiol*. 47, 3–25 (2023).

898 43. G. Wang, A. Sarkar, P. Carbonetto, M. Stephens, A simple new approach to variable selection in regression,
899 with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol*. 82 (2020).

900 44. C. G. Artieri, H. B. Fraser, Evolution at two levels of gene expression in yeast. *Genome Res.* 24 (2014).

901 45. C. J. McManus, G. E. May, P. Spealman, A. Shteyman, Ribosome profiling reveals post-transcriptional
902 buffering of divergent gene expression in yeast. *Genome Res*. 24, 422–430 (2014).

903 46. F. W. Albert, D. Muzzey, J. S. Weissman, L. Kruglyak, Genetic influences on translation in yeast. *PloS Genet*.
904 10, e1004692 (2014).

905 47. Y. Jiang, G. Giase, K. Grennan, A. W. Shieh, Y. Xia, L. Han, Q. Wang, Q. Wei, R. Chen, S. Liu, K. P. White,
906 C. Chen, B. Li, C. Liu, DRAMS: A tool to detect and re-align mixed-up samples for integrative studies of
907 multi-omics data. *PloS Comput. Biol.* 16, e1007522 (2020).

908 48. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference
909 genome. *BMC Bioinformatics*. 12, 323 (2011).

910 49. M. C. Oldham, P. Langfelder, S. Horvath, Network methods for describing sample relationships in genomic
911 datasets: application to Huntington's disease. *BMC Syst. Biol.* 6, 63 (2012).

912 50. B. M. Bolstad, R. A. Irizarry, M. Astrand, T. P. Speed, A comparison of normalization methods for high
913 density oligonucleotide array data based on variance and bias. *Bioinformatics*. 19, 185–193 (2003).

914 51. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. 9, 357–359 (2012).

915 52. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence
916 reads to genomic features. *Bioinformatics*. 30, 923–930 (2014).

917 53. O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals
918 (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc*. 7, 500–507
919 (2012).

920 54. O. Delaneau, H. Ongen, A. A. Brown, A. Fort, N. I. Panousis, E. T. Dermitzakis, A complete tool set for
921 molecular QTL discovery and analysis. *Nat. Commun*. 8, 15452 (2017).

922 55. B. L. Pierce, S. Burgess, Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample
923 Instrumental Variable Estimators. *Am. J. Epidemiol.* 178, 1177–1184 (2013).

924 **Acknowledgments**

925 We thank Majd Alsayed, Miguel Brown, Dominic Fitzgerald, Amber Thomas, and Mimi
926 Brown for their assistance with RNA-Seq data production and processing. We thank Richard

927 Kopp at SUNY Upstate Medical University for his help with wordsmithing. We thank all of the
928 donors and their relatives participated in the brain collections at Stanley Medical Research
929 Institute and Banner Sun Health Research Institute. We thank Drs. Maree J. Webster and Thomas
930 Beach for their supports to make the tissue available. We thank the High Performance Computing
931 Center of Central South University for partial support of calculation.

932 **Funding:**

933 SUNY Empire Innovation Program
934 National Institutes of Health grants U01MH122591
935 National Institutes of Health grants U01MH116489
936 National Institutes of Health grants R01MH110920
937 National Institutes of Health grants R01MH126459
938 National Institutes of Health grants U01MH103340
939 National Institutes of Health grants R01MH109715
940 National Institutes of Health grants R01MH110555
941 National Institutes of Health grants R21MH129817
942 National Natural Science Foundation of China (Grants Nos. 82022024, 31970572)
943 Science and technology innovation Program of Hunan Province (2021RC4018,
944 2021RC5027)

945 **Author contributions:**

946 Conceptualization: CL, KPW, JP, SHW, BL, XW
947 Methodology: AWS, LC, DZ, FW, MX, MN, DP, YJW, RV, CZ, KG, GG
948 Investigation: SHW, QL, YJ, CC, CY, AWS
949 Visualization: QL
950 Supervision: SHW, CL, CC, BL, KPW

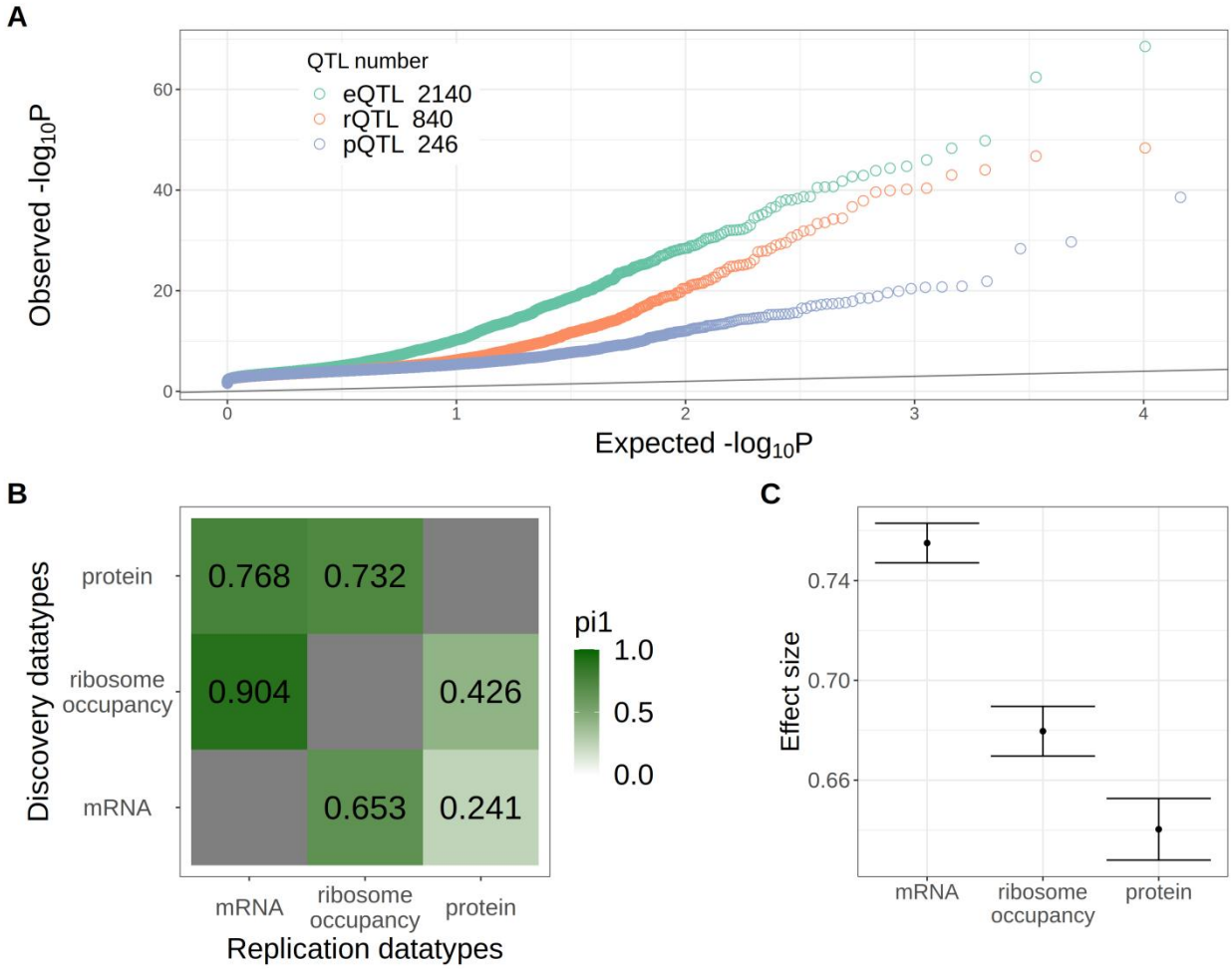
Writing—original draft: SHW, QL, YJ, CL

Writing—review & editing: SHW, CL, CC, QL

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data, code, and materials used in the analyses are available in https://www.synapse.org/#!Synapse:syn50698163. All data are available in the main text or the supplementary materials.

Figures and Tables



959 **Fig. 1. Genetic regulation of gene expression in the human brain. (A)** P-value quantile-
960 quantile plot between the observed (Y-axis) and the expected based on null distribution (X-axis).
961 The black line indicates the expected distribution of p values when there are no real QTL signals.
962 The number of cis-QTLs (i.e. the most significantly associated SNP for each gene) identified at
963 10% FDR is labeled in the top left inset. **(B)** Replication rate between QTL types. Proportions of
964 QTL replicated in the other two datatypes are listed in the 3X3 matrix. Each row is a discovery
965 datatype and each element of the row correspond to the proportion QTL signal replicated in the
966 datatype specified by the column label. For example, only 24.1% of the eQTL signal is replicated
967 in the protein data. **(C)** Effect size of CMC eQTL SNPs in BrainGVEX data. Mean and standard
968 error of absolute per allele effect across all CMC eQTL SNPs that were also analyzed in the
969 BrainGVEX union set is shown.

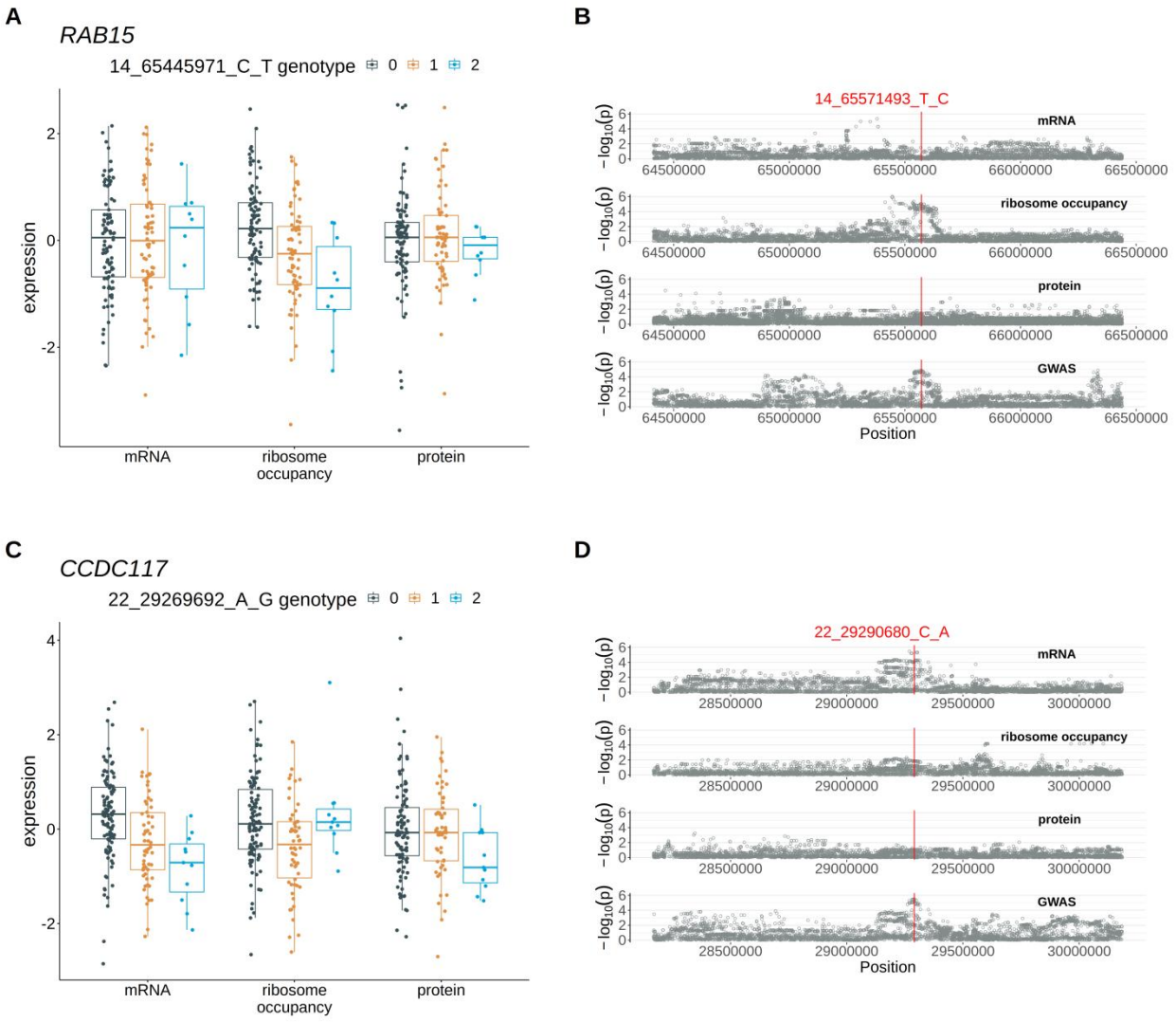


Fig. 2. Signal colocalization between schizophrenia GWAS and datatype-specific QTL. (A, C) Boxplots summarizing normalized gene expression level stratified by QTL genotype for *RAB15* rsQTL (A) and *CCDC117* esQTL (C). **(B, D)** Manhattan plots showing p value distribution for each QTL type and schizophrenia GWAS for the 1Mb QTL mapping window flanking *RAB15* (B) and *CCDC117* (D). The red line indicates the position of the lead colocalization SNP between datatype-specific QTL and schizophrenia GWAS.

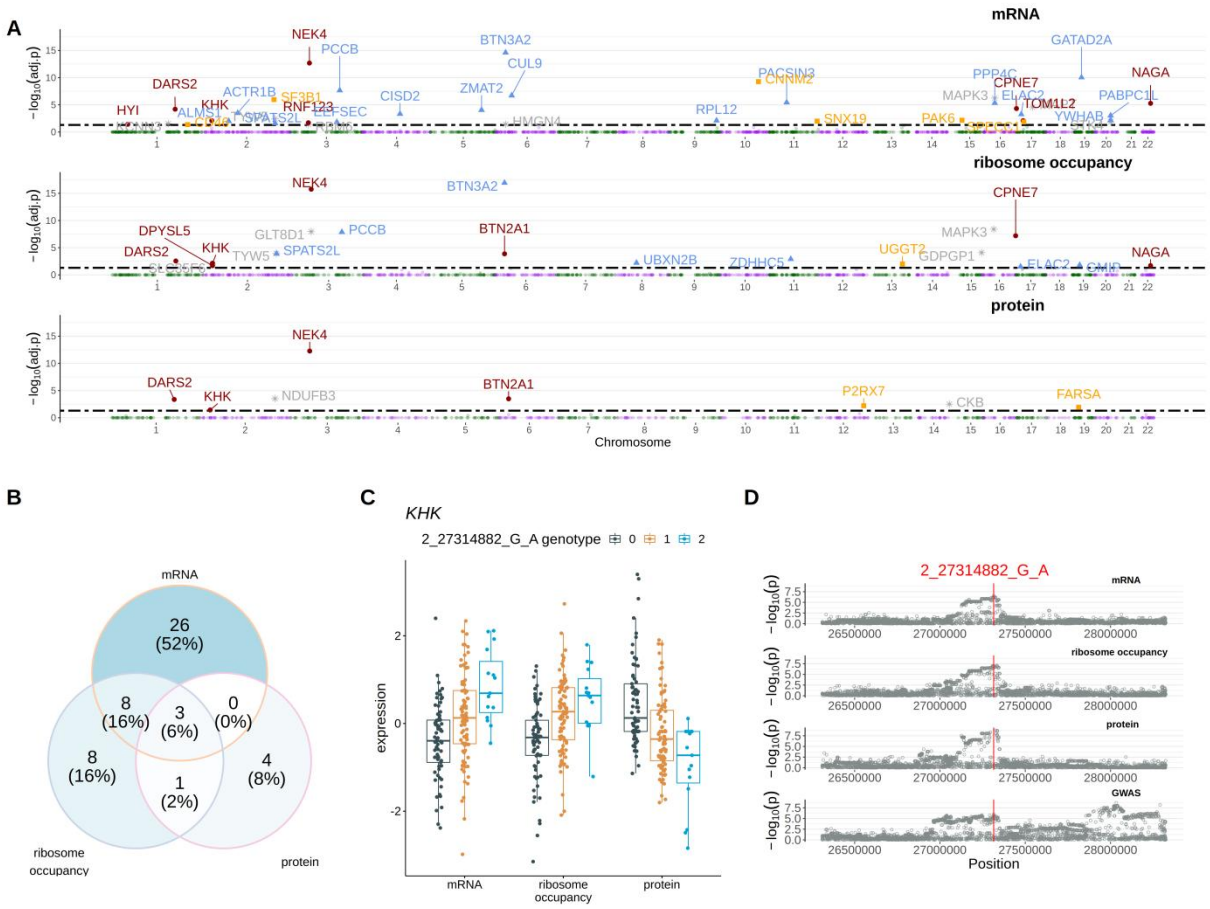


Fig. 3. Schizophrenia risk genes identified from each of the three datatypes RNA-Seq (mRNA), ribo-seq (ribosome occupancy), and proteomics (protein) using S-PrediXcan. (A) Manhattan plots showing significance level (i.e. -log10 FWER from S-PrediXcan) of gene-schizophrenia association across the genome for genes that pass the 5% FWER significance cutoff. The black horizontal dotted line indicated the significance cutoff. Risk genes are color-coded according to the MR test results. Grey asterisks mark the risk genes that failed the two-sample MR tests; red solid circle marks the risk genes that pass both one-sample MR tests (passing both one-sample MR tests suggest that transcriptionally regulated protein level differences between individuals drives the disease risk); blue triangle marks the risk genes that pass one of the two one-sample MR tests; orange rectangle marks the risk genes that failed both one-sample MR tests. **(B)** Venn diagram illustrating the number and corresponding percentage of overlapping risk genes

990 between datatypes. **(C)** Boxplots summarizing normalized gene expression level stratified by
991 eQTL genotype for *KHK*. **(D)** Manhattan plots showing p value distribution for each QTL type
992 and schizophrenia GWAS for the 1Mb QTL mapping window flanking *KHK*. The red line
993 indicates the position of the lead colocalization SNP between eQTL and schizophrenia GWAS.
994

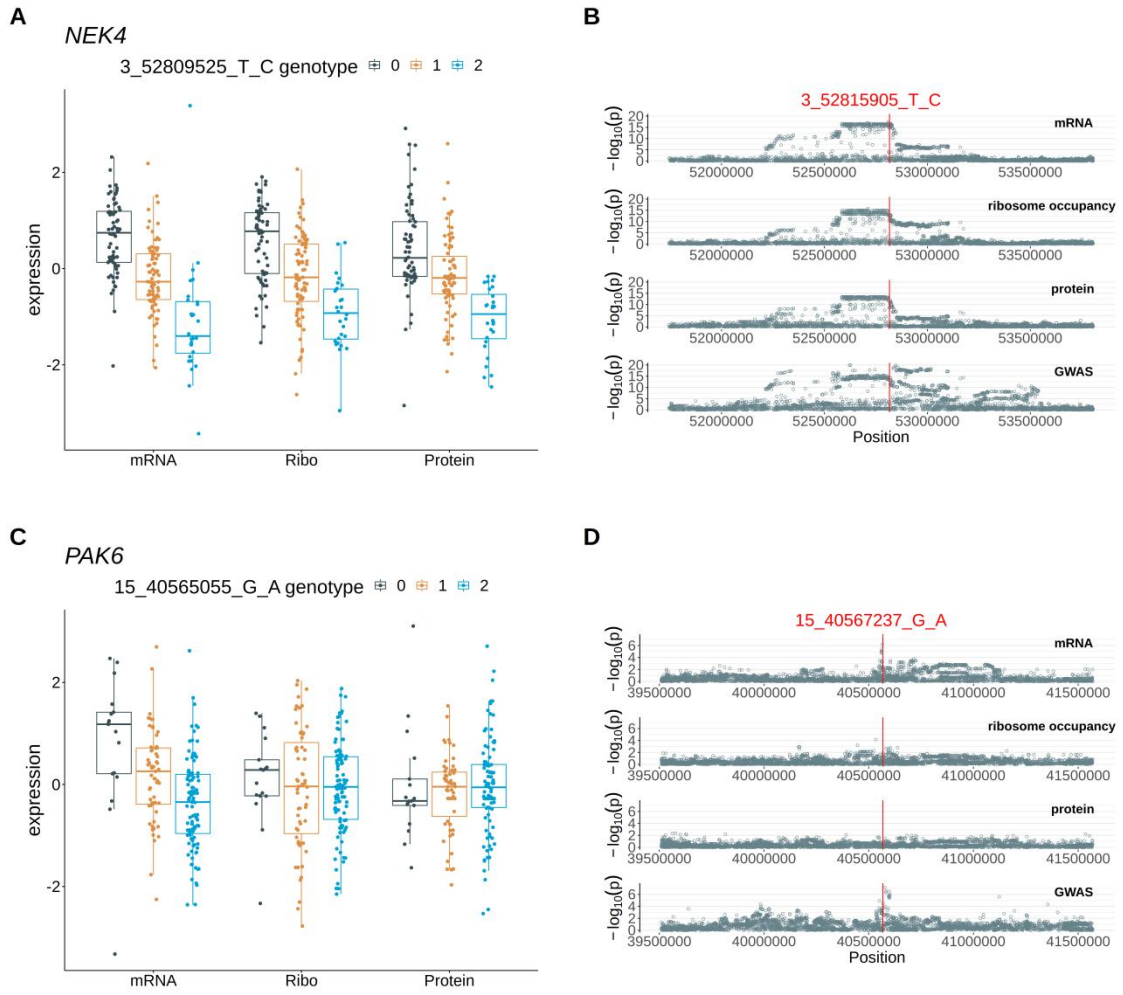


Fig. 4. Signal colocalization between schizophrenia GWAS and eQTL of example risk genes. **(A, C)** Boxplots summarizing normalized gene expression level stratified by QTL genotype for *NEK4* eQTL (A) and *PAK6* eQTL (C). **(B, D)** Manhattan plots showing p value distribution for each QTL type and schizophrenia GWAS for the 1Mb QTL mapping window flanking *NEK4* (B) and *PAK6* (D). The red line indicates the position of the colocalization lead SNP between eQTL and schizophrenia GWAS.