Co-localization of Conditional eQTL and GWAS Signatures in Schizophrenia

Amanda Dobbyn[1,2], Laura M. Huckins[1,2], James Boocock[3], Laura G. Sloofman[1], Benjamin S. Glicksberg[2,4,5], Claudia Giambartolomei[3,6], Gabriel Hoffman[2,4], Thanneer Perumal[7], Kiran Girdhar[1,2], Yan Jiang[1,8], Douglas M. Ruderfer[9], Robin S. Kramer[10], Dalila Pinto[4,8,11,12], the CommonMind Consortium, Schahram Akbarian[1,8], Panos Roussos[1,4,8], Enrico Domenici[13], Bernie Devlin[14], Pamela Sklar[1,2,4,8], Eli A. Stahl[1,4,*,‡] and Solveig K. Sieberts[7,**,‡]

1) Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
3) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90024, USA
4) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
5) Institute for Next Generation Healthcare, Mount Sinai Health System, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
6) Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA
7) Systems Biology, Sage Bionetworks, Seattle, WA 98109, USA
8) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
9) Division of Genetic Medicine, Department of Medicine, Psychiatry and Biomedical Informatics, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37235, USA
10) Human Brain Collection Core, National Institute of Mental Health, Bethesda, MD 20892, USA
11) Department of Psychiatry and Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
12) Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
13) Laboratory of Neurogenomic Biomarkers, Centre for Integrative Biology (CIBIO), University of Trento, Trento, Italy
14) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213, USA

‡ These authors contributed equally to this work

* eli.stahl@mssm.edu
** solly.sieberts@sagebase.org

ABSTRACT

Causal genes and variants within genome-wide association study (GWAS) loci can be identified by integrating GWAS statistics with expression quantitative trait loci (eQTL) and determining which SNPs underlie both GWAS and eQTL signals. Most analyses, however, consider only the marginal eQTL signal, rather than dissecting this signal into multiple independent eQTL for each gene. Here we show that analyzing conditional eQTL signatures, which could be important under specific cellular or temporal contexts, leads to improved fine mapping of GWAS associations. Using genotypes and gene expression levels from post-mortem human brain samples (N=467) reported by the CommonMind Consortium (CMC), we find that conditional eQTL are widespread; 63% of genes with primary eQTL also have conditional eQTL. In addition, genomic features associated with conditional eQTL are consistent with context specific (i.e. tissue, cell type, or developmental time point specific) regulation of gene expression. Integrating the Psychiatric Genomics Consortium schizophrenia (SCZ) GWAS and CMC conditional eQTL data reveals forty loci with strong evidence for co-localization (posterior probability >0.8), including six loci with co-localization of conditional eQTL. Our co-localization analyses support previously reported genes and identify novel genes for schizophrenia risk, and provide specific hypotheses for their functional follow-up.

64 INTRODUCTION

65

66 Significant advances in understanding the genetic architecture of schizophrenia
67 have occurred over the last ten years. However, for common variants identified in
68 genome-wide association studies (GWAS), the success in locus identification is
69 not yet matched by an understanding of their underlying basic mechanism or
70 effect on pathophysiology. Expression quantitative trait loci (eQTL), which are
71 responsible for a significant proportion of variation in gene expression, could
72 serve as a link between the numerous non-coding genetic associations that have
73 been identified in GWAS and susceptibility to common diseases directly through
74 their association with gene expression regulation.[1-4] Indeed, results from eQTL
75 mapping studies have been successfully utilized to identify genes and causal
76 variants from GWAS for various complex phenotypes, including asthma, body
77 mass index, celiac disease, and Crohn's disease.[5-8]

78

79 Studies integrating eQTL and GWAS data have almost exclusively used marginal
80 association statistics which typically represent the primary, or most significant,
81 eQTL signal when assessing co-localization with GWAS, ignoring other SNPs
82 that affect expression independently of the primary eQTL for a given gene.
83 However, recent findings indicating that conditionally independent eQTL are
84 widespread[9-11] motivate examination of the extent to which considering
85 conditional eQTL may provide additional power to identify likely causal genes in a
86 GWAS locus. Recent reports provide evidence that conditional eQTL are less

87 frequently shared across tissues than primary eQTL[9] and, like tissue and cell
88 type specific eQTL, are often found more distally to the genes they regulate.[9; 12;
89 13] These lines of evidence suggest that conditionally independent eQTL may
90 contribute to tissue- or other context-specific gene regulation (e.g. specific to a
91 particular cell type, developmental stage, or stimulation condition).
92
93 Here, we leveraged genotype and dorsolateral prefrontal cortex (DLPFC)
94 expression data provided by the CommonMind Consortium (CMC) to elucidate
95 the role of conditional eQTL in the etiology of schizophrenia (SCZ). Currently
96 comprising the largest existing postmortem brain genomic resource at nearly 600
97 samples, the CMC is generating and making publicly available an unprecedented
98 array of functional genomic data, including gene expression (RNA-sequencing),
99 histone modification (chromatin immunoprecipitation, ChIP-seq), and SNP
100 genotypes, from individuals with psychiatric disorders as well as unaffected
101 controls.[14] We utilized SNP dosage and RNA-sequencing (RNA-seq) data from
102 the CMC to identify primary and conditionally independent eQTL. We then
103 characterized the resulting eQTL on various genomic attributes including
104 distance to transcription start site, and their genes' specificity across tissues, cell-
105 types, and developmental periods. In addition, we quantified enrichment of
106 primary and conditional eQTL in promoter and enhancer functional genomic
107 elements inferred from epigenomic data. Finally, we isolated each independent
108 eQTL signal by conducting a series of "all-but-one" conditional analyses for

genes with multiple independent eQTL, and assessed the overlap between all eQTL association signals and the SCZ GWAS signals.

MATERIAL AND METHODS

CommonMind Consortium Data

We used pre-QC'ed genotype and expression data made available from the CommonMind Consortium, and detailed information on quality control, data adjustment and normalization procedures can be found in Fromer et. al.[14] Briefly, samples were genotyped at 958,178 markers using the Illumina Infinium HumanOmniExpressExome array, and markers were removed on the basis of having no alternate alleles, having a genotyping call rate ≤ 0.98, or a Hardy-Weinberg P-value $< 5x10^{-5}$. After phasing and imputation using the 1000 Genomes Phase 1 integrated reference then filtering out variants with INFO < 0.8 or MAF < 0.05, the total number of markers included in the analysis increased to approximately 6.4 million. Gene expression was assayed via RNA-seq using 100 base pair paired end reads, and mapped to human Ensembl gene reference (v70) using TopHat version 2.0.9 and Bowtie version 2.1.0. After discarding genes with less than 1 CPM (counts per million) in at least 50% of the samples, RNA-seq data for a total of 16,423 Ensembl genes were considered for analysis. The expression data was voom-adjusted for both known covariates (RIN, library batch, institution, diagnosis, post-mortem interval, and sex) and surrogate

132 variable analysis (SVA) identified surrogate variables. After the removal of
133 individuals that did not pass RNA sample QC (including but not limited to: having
134 RIN < 5.5, having less than 50 million total reads or more than 5% of reads
135 aligning to rRNA, having any discordance between genotyping and RNA-seq
136 data, and having RNA outlier status or evidence for contamination), and retaining
137 only genetically-identified European-ancestry individuals, a total of 467 samples
138 were used for downstream analyses. These 467 individuals comprised 209 SCZ
139 cases, 52 AFF (Bipolar, Major depressive disorder, or Mood disorder,
140 unspecified) cases, and 206 controls.
141
142 **eQTL Identification**
143
144 To identify primary and conditional cis-eQTL, we a conducted forward stepwise
145 conditional analysis implemented in MatrixEQTL[15] using genotype data at 6.4
146 million markers and RNA-seq data for 16,423 genes. For each gene with at least
147 one cis-eQTL (gene ± 1 Mb) association at a 5% false discovery rate (FDR), the
148 most significant SNP was added as a covariate in order to identify additional
149 independent associations. This procedure was repeated iteratively until no further
150 FDR significant eQTL were identified. We used a linear regression model,
151 adjusting for diagnosis and five ancestry covariates inferred by GemTools.
152 Following eQTL identification, only autosomal eQTL were retained for
153 downstream analyses.
154

155 **Replication in Independent Datasets**

156

157 Replication was performed in HBCC microarray cohort (dbGaP ID phs000979,
158 see Web Resources) and in the ROSMAP[16] RNA-seq cohort by fitting the
159 stepwise regression models identified in the CMC data. For cases in which a
160 marker was unavailable in the replication cohort, all models including that marker
161 (i.e. for that eQTL and higher-order eQTL conditional on it, for a given gene)
162 were omitted from replication.

163

164 Data from the HBCC cohort was QC'ed and normalized as described in Fromer
165 et al.[14] DLPFC tissue was profiled on the Illumina HumanHT-12_V4 Beadchips
166 and normalized in an analogous manner to the CMC data. Genotypes were
167 obtained using the HumanHap650Yv3 or Human1MDuov3 chips and imputed to
168 1000 Genomes Phase 1. Replication of the eQTL models was performed on 279
169 genetically inferred Caucasian samples (76 controls, 72 SCZ, 43 BP, 88 MDD)
170 adjusting for diagnosis and five ancestry components.

171

172 ROSMAP data were obtained from the AMP-AD Knowledge Portal (see Web
173 Resources). Quantile normalized FPKM expression values were adjusted for Age
174 of Death, RIN, PMI, and hidden confounders from SVA, conditional on diagnosis.
175 Only genes with FPKM > 0 in > 50 samples were considered in analyses. QC'ed
176 genotypes were also obtained from the AMP-AD Knowledge Portal and imputed
177 to the Haplotype Reference Consortium (v1.1)[17] reference panel via the Michigan

178 Imputation Server.[18] Only markers with imputation quality score $R^2 \geq 0.7$ were
179 considered in the replication analysis. GemTools was used to infer ancestry
180 components as for CMC above. After QC, 494 samples were used for eQTL
181 replication in a linear regression model that also adjusted for diagnosis
182 (Alzheimer's disease, mild cognitive impairment, no cognitive impairment and
183 other) and four ancestry components.
184
185 **Modeling Number of eQTL per Gene on Genomic Features**
186
187 We considered three genomic features (gene length, number of LD blocks in the
188 cis-region, and genic constraint score) for our modeling analyses. Gene lengths
189 were calculated using Ensembl gene locations. We obtained LD blocks from the
190 LDetect Bitbucket site to tally the number of LD blocks overlapping each gene's
191 cis-region (gene $\pm$ 1Mb). We obtained Loss-of-Function-based genic constraint
192 scores from the Exome Aggregation Consortium (ExAC). A negative-binomial
193 generalized linear regression model was used to model the number of eQTL per
194 gene based on the above variables; results were qualitatively the same using
195 linear regression of Box-Cox transformed eQTL numbers. Backward-forward
196 stepwise regression using the full model with interaction terms for these three
197 variables was used to determine the relationship between genomic conditions
198 and eQTL number. These analyses were implemented in R.
199

200 Cis-heritability of gene expression was estimated using the same CMC data used
201 for eQTL detection, using all markers in the cis-region using GCTA[19], and SNP-
202 heritability estimates were included in the modeling described above.
203
204 Tissue, cell type, and developmental time point specificity were measured using
205 the expression specificity metric Tau.[20; 21] Tissue specificity for each gene was
206 calculated using publicly available expression data for 53 tissues from the GTEx
207 project[22] (release V6p). Expression for each tissue was summarized as the log2
208 of the median expression plus one, and then used to calculate tissue specificity
209 Tau. Cell type specificity for each gene was computed using publicly available
210 single-cell RNA-sequencing expression data[23] generated from human cortex and
211 hippocampus tissues. Raw expression counts for 285 cells comprising six major
212 cell types of the brain were obtained from GEO (GSE67835) and counts data
213 were library normalized to CPM. Expression for each cell type was then
214 summarized as the log2 of the mean expression plus one, and then used to
215 compute cell type specificity Tau. Developmental time point specificity for each
216 gene was calculated using publicly available DLPFC expression data for 27 time
217 points, clustered into eight biologically relevant groups, from the BrainSpan atlas
218 (see Web resources). Eight developmental periods[24] were defined as follows:
219 early prenatal (8-12 pcw), early mid-prenatal (13-17 pcw), late mid-prenatal (19-
220 24 pcw), late prenatal (25-37 pcw), infancy (4 mos - 1 yr), childhood (2 - 11 yr),
221 adolescence (13 - 19 yr), and adulthood (21 yr +). Expression for each time point
222 was summarized as the log2 of the median expression plus one, and then used

223 to calculate developmental period specificity (Tau). Each Tau was added to the
224 above modeling of eQTL number in turn, as well as all together.
225
226 **Enrichment Analyses**
227
228 We divided eQTL into separate subgroups by stepwise conditional order (first,
229 second, third, and greater than third), and created sets of matched SNPs drawn
230 from the SNPsnap database for each subgroup, matching on minor allele
231 frequency, gene density (number of genes within 1Mb of the SNP), distance from
232 SNP to TSS of the nearest gene, and LD (number of LD-partners within $r^2 \geq 0.8$).
233 For each subgroup of eQTL, we performed a logistic regression of status as
234 eQTL or matched SNP on overlap with functional annotation, including the four
235 SNP matching parameters as covariates. Enrichment was taken as the
236 regression coefficient estimate, interpretable as the log-odds ratio for being an
237 eQTL given a functional annotation. Functional annotations tested included:
238 DLPFC promoters and enhancers (TssA and Enh+EnhG, respectively, from the
239 NIH Roadmap Epigenomics Project[25] ChromHMM[26] core 15-state model), Brain
240 promoters and enhancers (union of all brain region TssA and Enh+EnhG,
241 respectively, from the NIH Roadmap Epigenomics Project ChromHMM core 15-
242 state model), and pre-frontal cortex (PFC) neuronal (NeuN+) and non-neuronal
243 (NeuN-) nuclei H3K4me3 and H3K27ac ChIP-seq marks from the CMC. For each
244 data source, Roadmap DLPFC, brain, CMC NeuN+, NeuN-, active promoter and

245 enhancer (or H3K4me3 and H3K27ac) annotation were tested for enrichment
246 jointly.
247
248 **Conditional eQTL Analyses**
249
250 In order to isolate each conditionally independent cis-eQTL association, we
251 carried out a series of “all-but-one” conditional analyses, implemented within
252 MatrixEQTL[15], for each gene possessing more than one independent eQTL. As
253 these conditional eQTL signals were to be used to test for co-localization with the
254 SCZ GWAS signals, we limited these analyses to those genes (346 in total) with
255 eQTL overlapping GWAS loci. For each of these genes, we conducted an ‘”all-
256 but-one” analysis for each independent eQTL by regressing the given gene’s
257 expression data on the dosage data, including all of the other independent eQTL
258 for that gene as covariates in addition to diagnosis and five ancestry
259 components. For example, three conditional analyses would be conducted for a
260 gene with three independent eQTL: one analysis conditioning on the secondary
261 and tertiary eQTL, one analysis conditioning on the primary and tertiary, and one
262 analysis conditioning on the primary and secondary. In this manner we generated
263 summary statistics for each independent eQTL in isolation, conditional on all of
264 the other independent eQTL for that gene.
265
266
267

268 **Co-localization Analyses**

269

270 For our co-localization analyses, we used summary statistics and genomic
271 intervals from the 2014 PGC SCZ GWAS.[27] We included 217 loci at a P-value
272 threshold of $1x10^{-6}$ (omitting the complex MHC locus), defined these loci by their
273 LD $r^2 \geq 0.6$ with the lead SNP, and then merged overlapping loci. GWAS and
274 eQTL signatures were qualitatively compared using P-P plots, rendered in R, and
275 LocusZoom[28] plots.

276

277 We tested for co-localization using an updated version of COLOC[29] R functions,
278 which we name COLOC2 (see Web Resources) which incorporates several
279 improvements to the method. First, COLOC2 preprocesses data by aligning
280 eQTL and GWAS summary statistics for each eQTL cis-region. Second, the
281 COLOC2 model optionally incorporates changes implemented in gwas-pw[30].
282 Briefly, we implemented learning mixture proportions of five hypotheses (H_0, no
283 association; H_1, GWAS association only; H_2, eQTL association only; H_3, both but
284 not co-localized; and H_4, both and co-localized) from the data. COLOC2 uses
285 these proportions as priors (or optionally, COLOC default or user specified priors)
286 in the empirical Bayesian calculation of the posterior probability of co-localization
287 for each locus (eQTL cis-region). COLOC2 averages per-SNP Wakefield
288 asymptotic Bayes factors (WABF)[31] across three different values for the WABF
289 prior variance term, 0.01, 0.1, and 0.5, and provides options for specifying

290 phenotypic variance, estimating it from case-control proportions, or estimating it
291 from the data.
292
293 RESULTS
294
295 **Identification of eQTL**
296
297 Primary and conditional eQTL were identified using genotype and RNA-seq data
298 from the CommonMind Consortium post-mortem DLPFC samples (467
299 European-ancestry cases and controls).[14] We identified 16,273 conditional eQTL
300 in addition to the 13,137 primary eQTL we previously reported[14] for a total of
301 29,410 independent cis-eQTL for 15,817 autosomal genes. Of the genes tested,
302 81% (12,813 genes) had at least one eQTL and 63% of these (51% of all genes)
303 also had at least one conditional eQTL, with an average of 1.83 independent
304 eQTL per gene (2.26 among those with at least one eQTL), and a maximum of
305 16 eQTL (Figure 1). Conversely, when examining the distributions for the number
306 of genes whose expression was affected by each eQTL (**Table S1**), the majority
307 of eQTL were specific for a single gene, and only a small fraction of eQTL,
308 1.47%, affected more than one gene, with a maximum of six genes affected by a
309 single eQTL.
310
311 We tested conditional eQTL for replication in two independent data sets, the
312 National Institute of Mental Health's Human Brain Collection Core (HBCC,

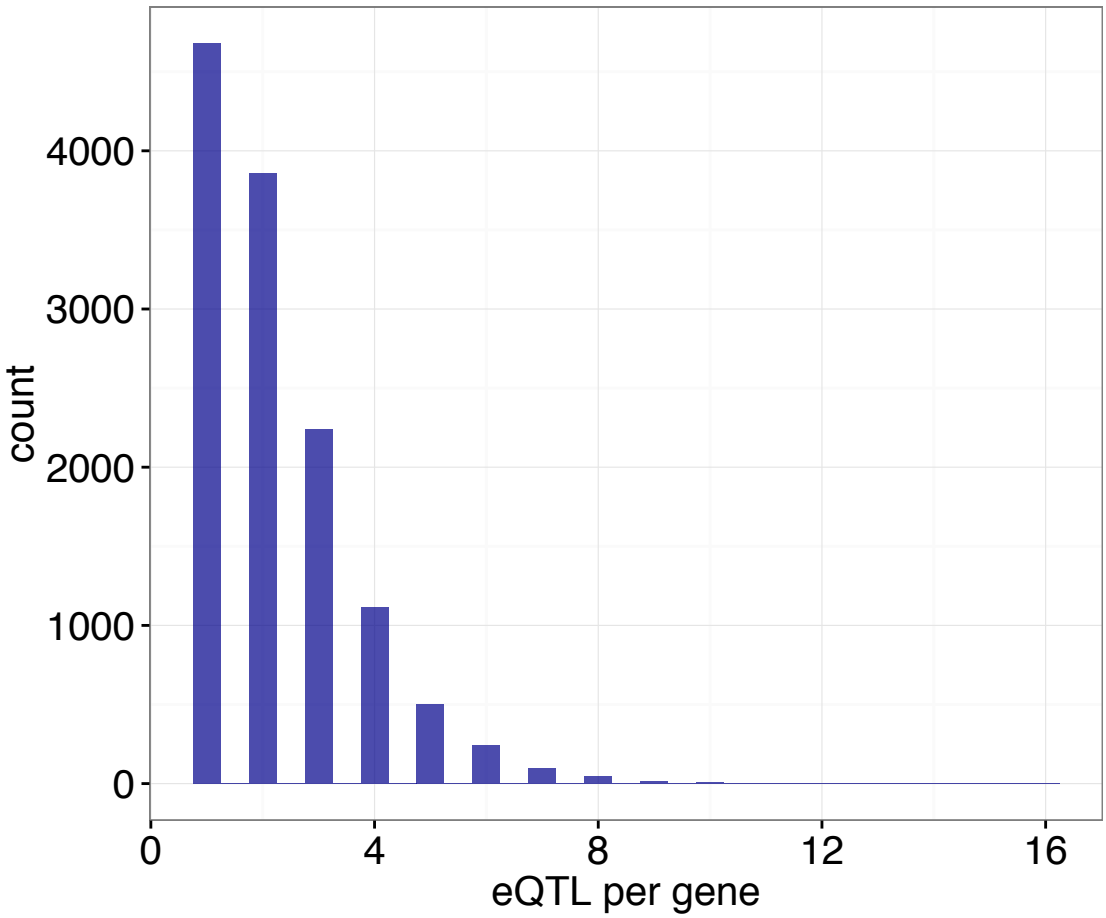

Figure 1. Distribution of the Number of Independent eQTL per Gene

Counts of the numbers of genes (y-axis) regulated by N ($1 \leq N \leq 16$) independent eQTL (x-axis). Plotted are 28,895 cis-eQTL with FDR ≤ 5%, for 12,813 autosomal genes. For genes with eQTL, there are an average of 2.26 eQTL per gene and a maximum of 16 eQTL per gene.

313 N=279, microarray expression data) and the Religious Orders Study / Memory
314 and Aging Project[16] (ROSMAP, N=494, RNA-seq expression). For each gene the
315 same models were evaluated that were identified in forward-stepwise conditional
316 analysis in the CMC data. We observed strong evidence of replication for both
317 primary and conditional eQTL in the HBCC and ROSMAP post-mortem brain
318 cohorts (**Table S2**). The estimated proportion of true associations (π_1) in
319 ROSMAP was 0.57 and 0.26 for primary and conditional eQTL, respectively; in
320 HBCC π_1 was 0.46 and 0.20 for primary and conditional eQTL. Thus replication
321 was stronger for primary than for conditional eQTL, as expected given their
322 stronger effect sizes. Replication rates were somewhat higher in the RNA-seq
323 ROSMAP data than in HBCC.
324
325 **Genomic Characterization of Primary and Conditional eQTL**
326
327 According to prior results, eQTL that are shared across tissues and cell types
328 tend to be located closer to transcription start sites (TSS) than context specific
329 eQTL.[9; 12; 13] We therefore examined the relationship between primary or
330 conditional eQTL status and distance to its gene's transcription start site. Primary
331 eQTL fall closer to the TSS than conditional eQTL (Figure 2): primary eQTL
332 occur at a median distance of 70.4 Kb from the TSS versus a median distance of
333 302 Kb for conditional eQTL. This difference holds true even more proximally to
334 TSS (**Figure S1**); 8.1 and 2.5 percent of primary and conditional eQTL,
335 respectively, fall within three Kb of the TSS.

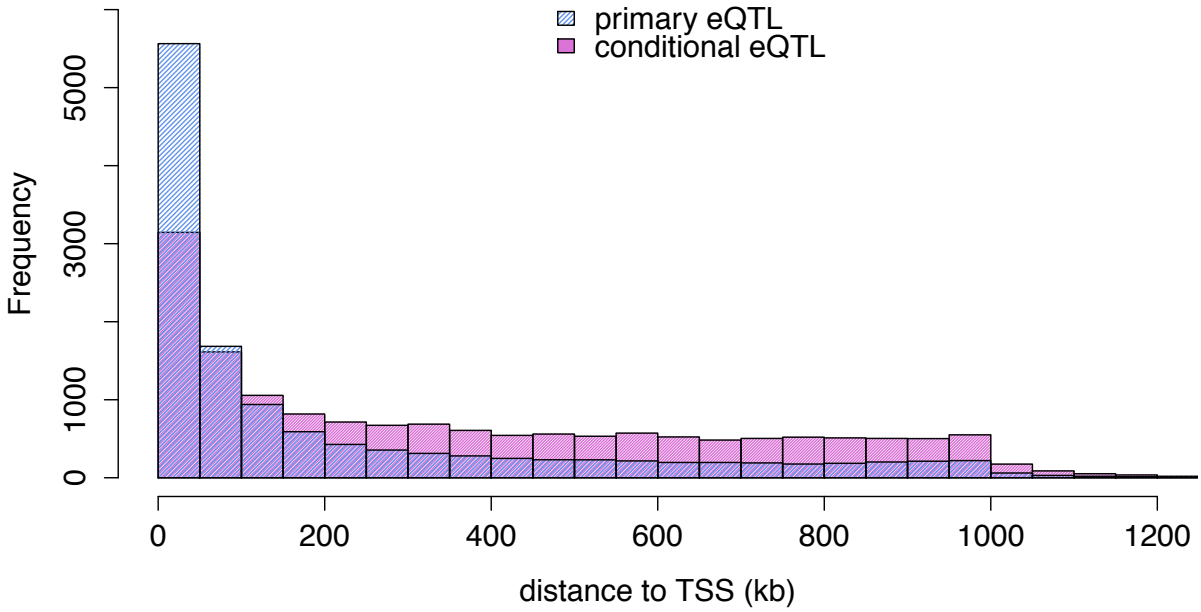

Figure 2. Distance from eQTL to transcription start site (TSS)

Overlapping histograms showing the numbers of eQTL (y-axis) occurring at increasing distances to TSS (x-axis), for primary eQTL (blue) and conditional eQTL (pink).

336

337 We next characterized the relationship between the number of independent
338 eQTL per gene and three different genomic features: gene length, number of LD
339 blocks[32] in the gene's cis-region (±1 Mb) and Exome Aggregation Consortium
340 (ExAC) genic constraint score,[33] including possible interactions. The best
341 multivariate model for eQTL number included gene length, number of LD blocks
342 and genic constraint as predictors, as well as a gene length-LD blocks interaction
343 (Table 1). The number of independent eQTL was positively correlated with gene
344 length and number of LD blocks, and negatively correlated with genic constraint
345 score (**Figure S2**).

346

347 We next examined the variance of gene expression explained by cis-region
348 SNPs, or cis-SNP-heritability, estimated by linear mixed model variance
349 component analysis[19] (**Figure S3**). We found a strong effect of estimated cis-
350 heritability on number of independent eQTL (Table 1, **Figure S4**). In a joint model
351 with cis-SNP-heritability, the main effects of gene length, number of LD blocks
352 and genic constraint on eQTL number remained at least nominally significant.

353

354 Finally we addressed whether genes with conditional eQTL exhibit greater
355 context specificity as measured by the robust expression specificity metric Tau.[20;
356 21] We calculated Tau across 53 tissues from the Genotype-Tissue Expression
357 (GTEx) project, across six DLPFC cell types (astrocytes, endothelial cells,
358 microglia, neurons, oligodendrocytes, and oligodendrocyte progenitor cells) from

Table 1. Number of Independent eQTL Modeled on Genomic Features

Predictor	Model 1 Estimate	Model 1 Robust SE	Model 1 Pr(>|z|)	Model 2 Estimate	Model 2 Robust SE	Model 2 Pr(>|z|)	Model 3 Estimate	Model 3 Robust SE	Model 3 Pr(>|z|)
log(Gene length)	0.27	0.04	5.16E-12	0.16	0.03	2.20E-06	0.17	0.03	9.87E-07
LD blocks	0.59	0.17	6.47E-04	0.33	0.15	2.92E-02	0.37	0.15	1.55E-02
log(Gene length) : LD blocks	-0.03	0.02	7.77E-02	-0.01	0.01	5.65E-01	-0.01	0.01	4.11E-01
Constraint	-0.61	0.03	5.93E-85	-0.20	0.03	2.93E-13	-0.15	0.03	5.41E-08
Cis-heritability	-	-	-	7.03	0.18	0.00	7.02	0.18	0.00
Tau (tissue)	-	-	-	-	-	-	0.08	0.08	2.76E-01
Tau (DLPFC cell type)	-	-	-	-	-	-	0.20	0.09	3.69E-02
Tau (developmental time point)	-	-	-	-	-	-	0.17	0.09	5.99E-02

359 single cell RNA-seq[23], and across eight developmental periods[24] (early prenatal,
360 early mid-prenatal, late mid-prenatal, late prenatal, infant, child, adolescent, and
361 adult) from the BrainSpan atlas DLPFC RNA-seq data**.** We confirmed that higher
362 values of Tau reflect expression specificity, by comparing the distributions of all
363 three Tau measures for all genes with the distributions for a subset of
364 housekeeping genes[34] (**Figure S5**). We found positive correlations between
365 eQTL number and tissue, cell type, and developmental specificities (Table 1,
366 **Table S3, Figure S6**). The strongest correlation was with DLPFC cell type Tau,
367 which is consistent with previous data demonstrating tissue specific, cell type
368 dependent expression in blood;[35] however, we note that all three Tau sets were
369 inter-correlated (**Table S3**).
370
371 **Epigenetic Enrichment Analyses**
372
373 One way in which eQTL may affect gene expression is through alteration of cis-
374 regulatory elements such as promoters and enhancers. Putative causal eQTL
375 variants have been shown to be enriched in genomic regions containing
376 functional annotations such as DNase hypersensitive sites, transcription factor
377 binding sites, promoters, and enhancers.[36-39] Our observation that conditional
378 eQTL fall farther from transcription start sites than primary eQTL led us to
379 hypothesize that primary eQTL may affect transcription levels by altering
380 functional sites in promoters whereas conditional eQTL may do so by altering
381 more distal regulatory elements such as enhancers. We therefore assessed

382 enrichment of primary and conditional eQTL in DLPFC and brain active promoter
383 (TssA) and enhancer (merged Enh and EnhG) states derived from the NIH
384 Roadmap Epigenomics Project,[25; 26] and in H3K4me3 and H3K27ac ChIP-seq
385 peaks from a subset of the CMC post-mortem DLPFC samples. We performed
386 logistic regression of SNP status (eQTL versus random matched SNP) on
387 overlap with functional annotations, separately for each eQTL order (primary,
388 secondary, tertiary, and greater than tertiary).
389
390 Both primary and conditional eQTL were significantly enriched in both promoter
391 and enhancer chromatin states (Figure 3A-B, **Table S4**). Chromatin states from
392 the DLPFC showed stronger eQTL enrichment than did the Brain annotation
393 formed by merging all individual brain region chromatin states. We found that
394 enrichments in both the DLPFC and Brain annotations generally decreased with
395 higher conditional order of eQTL, particularly for the active promoter state. A
396 similar pattern was observed when examining enrichment in neuronal nuclei
397 (NeuN+) ChIP-seq peaks (Figure 3C), using the overlap of H3K4me3 and
398 H3K27ac ChIP-seq peaks as a proxy for active promoters and H3K27ac peaks
399 that do not overlap H3K4me3 peaks as a (relatively non-specific) proxy for
400 enhancers.[26] These analyses showed decreasing enrichment in both promoters
401 and enhancers with increasing eQTL order, with a more marked decrease in the
402 promoters. Though there was also significant enrichment of eQTL in non-
403 neuronal nuclei (NeuN-) ChIP-seq peaks, decreasing with higher eQTL order,
404 this trend of a more marked decrease in active promoters was not observed in

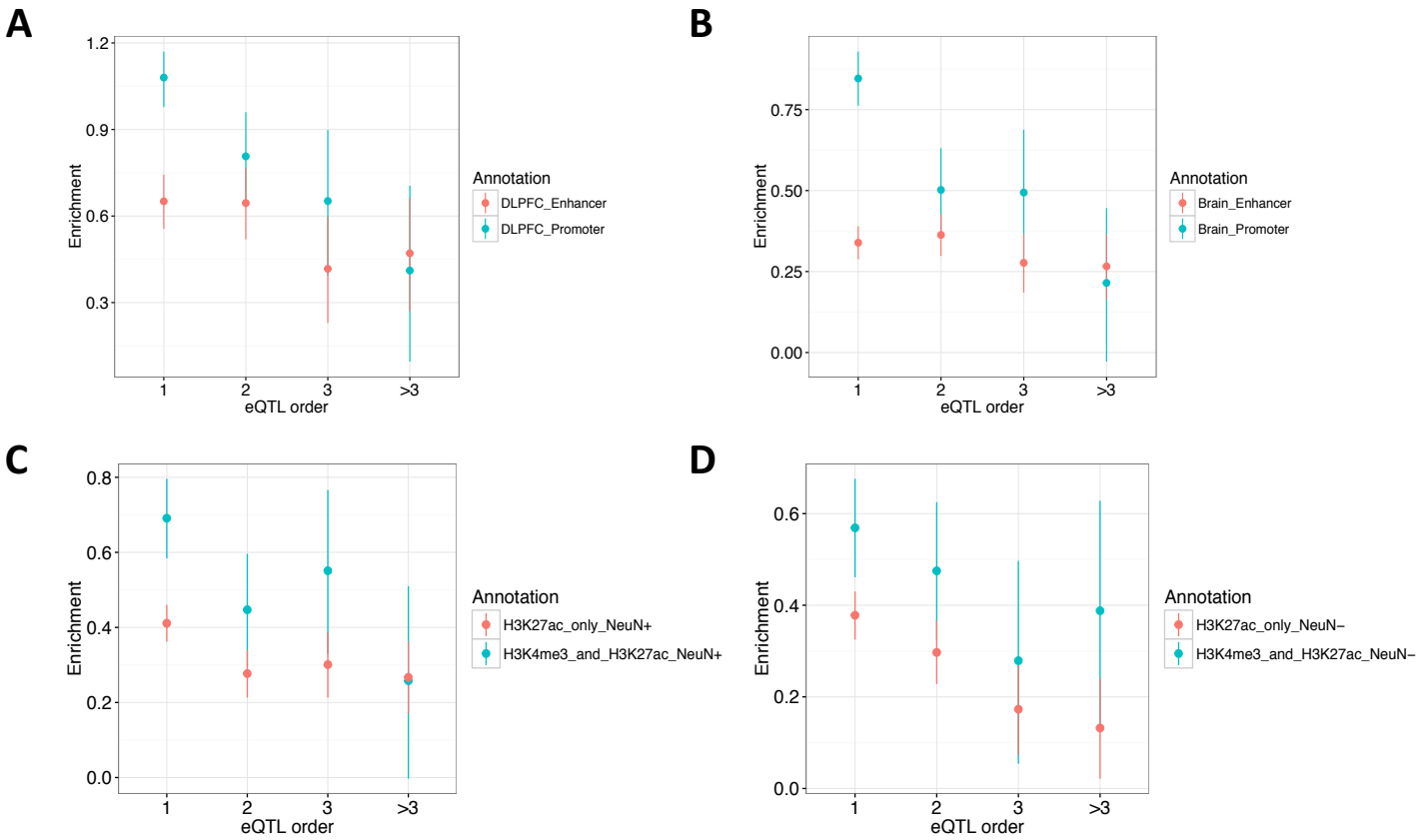

Figure 3. Enrichments of Primary and Conditional eQTL in Active Regulatory Annotations

Plotted are enrichments (estimate ± 95% CI from logistic regression, y-axes) of primary (x-axis eQTL order = 1) and conditional (eQTL order = 2, 3, >3) eQTL in functional annotations. Panels (A) and (B) show enrichment in DLPFC and Brain (union of all individual Brain regions) active promoter (turquoise) and enhancer (orange) ChromHMM states from the NIH Roadmap Epigenomics Project. Panel (C) shows enrichment in neuronal nuclei (NeuN+), for the intersection of H3K4me3 and H3K27ac ChIP-seq peaks (turquoise) and for H3K27 peaks that do not overlap H3K4me3 peaks (orange). Panel (D) shows enrichments in the same annotations, but for non-neuronal nuclei (NeuN-).

405 non-neuronal DLPFC nuclei (Figure 3D). Enrichment results for H3K4me3 and
406 H3K27ac ChIP-seq peaks are shown in Figure S7.
407
408 **eQTL Co-localization with SCZ GWAS**
409
410 We performed co-localization analyses in order to evaluate the extent of overlap
411 between eQTL and GWAS signatures in schizophrenia, and to identify putative
412 causal genes from GWAS associations. Considering 217 loci (**Table S5)** with
413 lead SNPs reaching a significance threshold of $P < 1x10^{-6}$ from the recent
414 Psychiatric Genomics Consortium schizophrenia GWAS,[27] we tabulated the
415 number of eQTL (FDR $\leq$ 5%) falling within GWAS loci. A total of 114 out of 217
416 loci contained primary and/or conditional eQTL for 346 genes; 110 of these
417 genes had one eQTL only, and 236 genes had more than one independent
418 eQTL.
419
420 To quantitatively compare the SCZ GWAS and eQTL association signatures, we
421 modified the R package COLOC[29] for Bayesian inference of co-localization
422 between the two sets of summary statistics across each gene's cis-region.
423 COLOC2, our modified implementation of COLOC, analyzes the hierarchical
424 model of pw-GWAS,[30] with likelihood-based estimation of dataset-wide
425 probabilities of five hypotheses (H_0, no association; H_1, GWAS association only;
426 H_2, eQTL association only; H_3, both but not co-localized; and H_4, both and co-
427 localized). We then used these probabilities as priors to calculate empirical

428 Bayesian posterior probabilities for the five hypotheses for each locus, in
429 particular PP_{H4} for co-localization.
430
431 For genes with conditional eQTL overlapping SCZ GWAS loci, summary
432 statistics from “all-but-one” conditional eQTL analyses were assessed for co-
433 localization with the GWAS signature (Figure 4). To illustrate this analytical
434 strategy, we show eQTL results for the iron responsive element binding protein 2
435 gene *IREB2* (*chr15:78729773-78793798*) as an example. Forward stepwise
436 selection analysis identified two independent cis-eQTL for *IREB2*. In order to
437 generate summary statistics for each eQTL in isolation, we conducted two “all-
438 but-one” conditional analyses, in each analysis conditioning on all but a focal
439 independent eQTL (for *IREB2* this entailed conditioning on only one eQTL per
440 conditional analysis, but involved conditioning on up to six eQTL across genes in
441 the SCZ co-localization analysis). We then tested for co-localization between the
442 GWAS and all of the conditional summary statistics using COLOC2. In the case
443 of *IREB2*, the conditional eQTL (rs7171869) was implicated as co-localized with
444 the GWAS signal at this locus with a posterior probability for co-localization PP_{H4}
445 = 0.94. A qualitative examination of the *IREB2* locus supported the COLOC2
446 results: the correlation between the GWAS P-values and conditional eQTL P-
447 values was higher than that between the GWAS and primary eQTL P-values
448 (Figure 5A). In addition, the GWAS signature for the locus more closely
449 resembled the conditional eQTL signature than either the non-conditional eQTL
450 signature or the primary eQTL signature (Figure 5B).

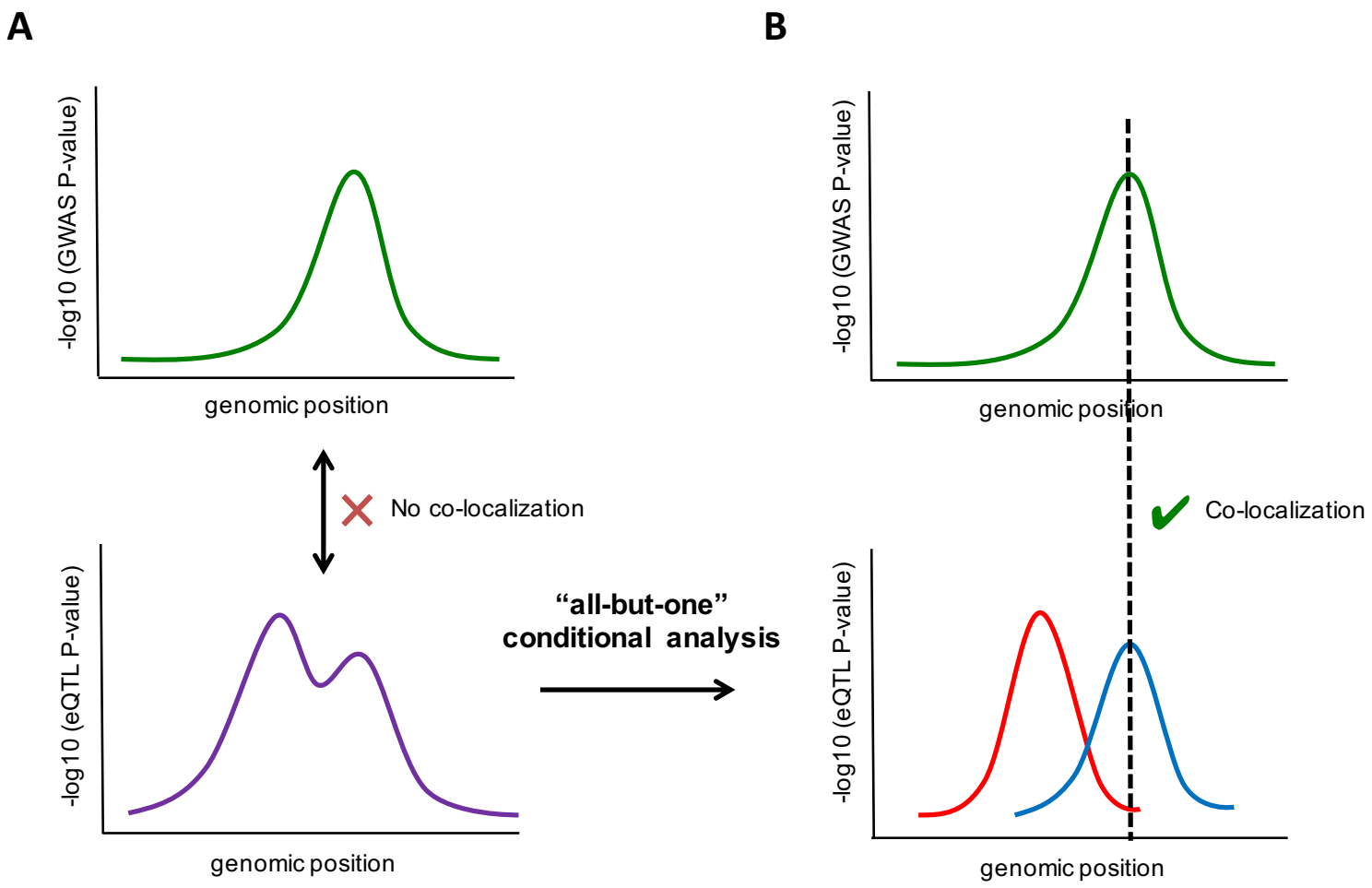

Figure 4. Conditional "All-but-One" Analysis to Isolate Independent eQTL Signatures

Panel (A) shows a hypothetical GWAS signature (top, green) at a given locus, and an overlapping hypothetical eQTL signature (bottom, purple), which comprises two independent eQTL. Panel (B) shows the same hypothetical GWAS and eQTL signatures after the "all-but-one" conditional eQTL analysis isolating the primary (red) and secondary (blue) eQTL signatures. Before conditional analysis there is a lack of co-localization between the GWAS signature and eQTL signature. After all-but-one conditional analysis, there is evidence for co-localization between the secondary eQTL and GWAS signatures.

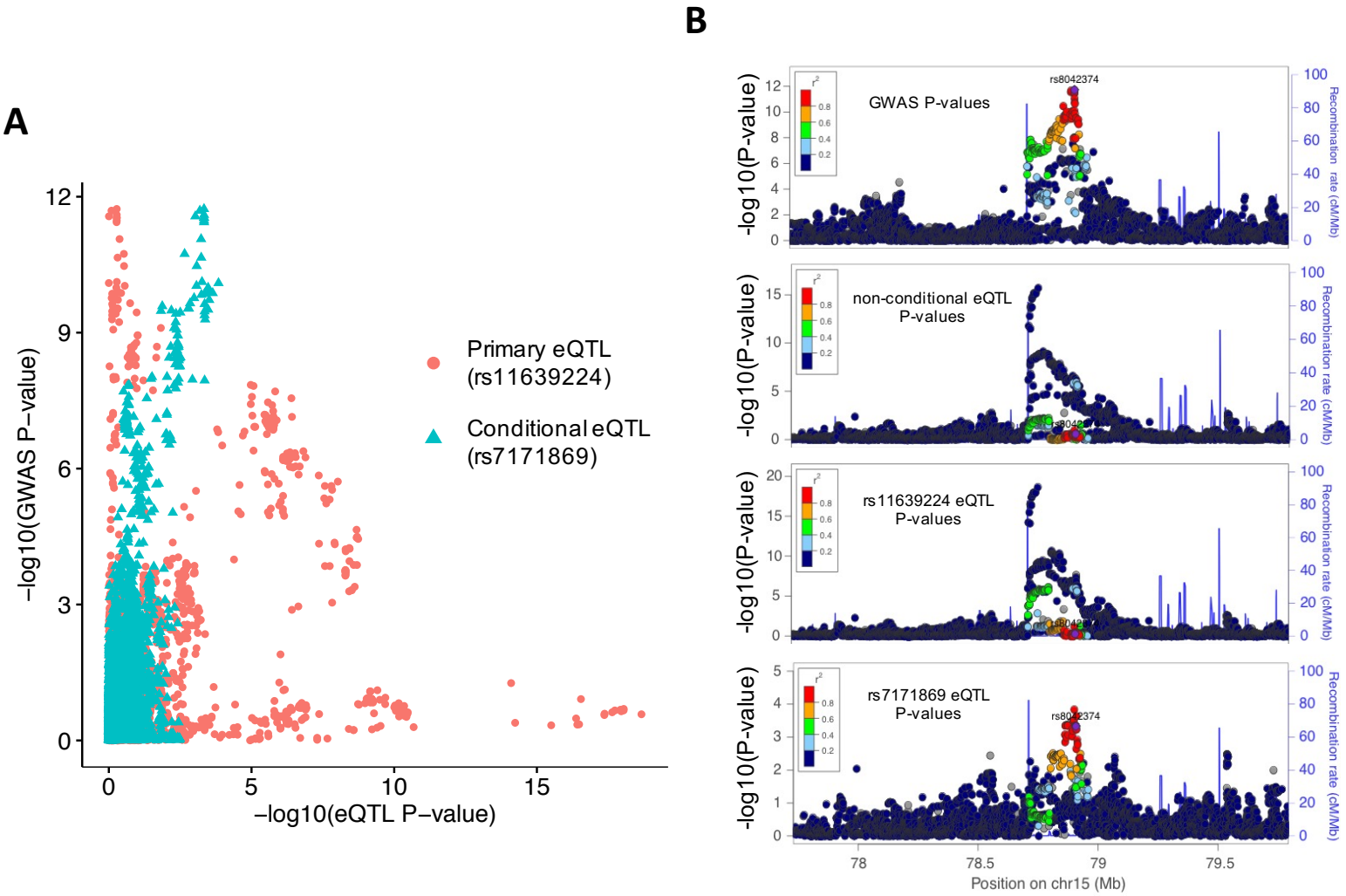

Figure 5. GWAS Signature for *IREB2* Co-localizes with the Conditional eQTL Signature

Panel (A) shows a P-P plot comparing -$\log_{10}$ P-values from GWAS (y-axis) and "all-but-one" conditional eQTL analysis (x-axis), which shows the highest correlation between the GWAS and the conditional eQTL (rs7171869, turquoise triangles). Panel (B) shows LocusZoom plots for the *IREB2* locus, where the GWAS signal (top) more closely resembles the conditional eQTL signal (rs7171869, bottom) than the primary eQTL signal (rs11639224, third from top) or non-conditional eQTL signal (second from top). For all LocusZoom plots LD is colored with respect to the GWAS lead SNP (rs8042374, labelled).

451

452 We found that 40 loci contained genes with strong evidence of co-localization
453 between eQTL and GWAS signatures, with posterior probability of H_4 (PP_{H4}) ≥
454 0.8 (Table 2, **Table S6**). When restricting to genome-wide significance for the
455 GWAS, we found co-localization in 24 of the 108 loci. Given the correlations
456 between number of independent eQTL and expression specificity scores (Tau)
457 across tissues, cell types and development, we tabulated the reported genes'
458 Tau percentiles and expression levels, to highlight contexts in which the genes
459 are specifically expressed (Table 2, **Table S9**). We acknowledge that while
460 posterior probability $PP_{H4} \geq 0.8$ demonstrates strong Bayesian evidence for co-
461 localization, it is an arbitrary threshold for characterizing loci as SCZ-eQTL co-
462 localized; we find that many loci with $PP_{H4} \geq 0.5$ appear qualitatively consistent
463 with co-localization.

464

465 Importantly, for six of the 40 co-localizing loci, a conditional rather than primary
466 eQTL co-localized with the GWAS with compelling qualitative support (Table 2,
467 Figure 5, **Figures S8-S12**). The genes showing strong evidence for conditional
468 eQTL co-localization include *SLC35E2*, *PROX1-AS1*, *PPM1M*, *SDAD1P1*,
469 *STAT6*, and *IREB2*. Also notable are the occurrences of complex patterns of co-
470 localization for some loci; for example, three loci showed evidence for co-
471 localization with a primary eQTL for one gene and a conditional eQTL for
472 another.

473

Table 2. GWAS-eQTL Co-localized Loci

Chr	Range.left	Range.right	GWAS SNP	GWAS P-value	eSNP	eSNP P-value	primary/ conditional	PP_{H4}	Gene	Relevant tissue / cell type / develop. period (See Table S9)
1	2372401	2402501	rs4648845	4.033E-09	rs12037821	6.4E-04	conditional	0.87	SLC35E2	- / - / Early mid-prenatal
1	8355697	8638984	rs301797	2.7E-09	rs138050288	3.8E-05	primary	0.95	RERE	- / - / -
1	30412551	30443951	rs1498232	1.3E-09	rs2015244	9.7E-12	primary	0.99	PTPRU	- / Neurons / Early mid-prenatal
1	163582923	163766623	rs7521492	5.6E-07	rs10799961	2.4E-11	primary	0.91	PBX1	- / - / Early prenatal
1	205015255	205189455	rs16937	8.7E-07	rs12724651	5.3E-07	primary	0.89	TMEM81	- / Neurons / -
					rs12031350	8.2E-06	conditional	0.87	RBBP5	- / - / -
1	214137889	214163689	rs7529073	9.7E-07	rs1431983	1.7E-04	conditional	0.93	PROX1-AS1	Cerebellar Hemisphere / Neurons / Adult
2	73194203	73900439	rs56145559	8.4E-08	rs11679809	2.0E-37	primary	0.86	ALMS1P	Testis / - / -
2	110262036	110398236	rs9330316	7.7E-08	rs892464	1.3E-28	primary	0.92	SEPT10	- / - / Late prenatal
2	198148577	198835577	rs6434928	1.5E-11	rs12621129	2.2E-12	primary	0.94	SF3B1	- / - / -
2	200715237	201247789	rs281768	3.5E-14	rs35220450	1.6E-10	primary	0.95	FTCDNL1, AC073043.2	- / - / Adult
					rs186546506	8.8E-04	conditional	0.83	LINC01792, AC007163.3	Putamen (basal ganglia) / - / Adult
2	208371631	208531731	rs2709410	4.1E-06	rs34171849	2.2E-16	primary	0.88	METTL21A	- / - / -
					rs2551656	1.6E-09	primary	0.86	CREB1	- / - / Early prenatal
2	220033801	220071601	rs6707588	9.5E-07	rs11236	1.1E-09	primary	0.92	CNPPD1	- / - / -
3	36843183	36945783	rs75968099	3.4E-12	rs9834970	1.9E-05	primary	0.94	DCLK3	Nerve - Tibial / Neurons / Infant
3	52281078	53539269	rs2535627	3.956E-11	rs6801235	2.8E-08	conditional	0.86	PPM1M	- / Neurons / Late prenatal
3	63792650	64004050	rs832187	2.6E-08	rs113386200	3.0E-12	primary	0.98	THOC7	- / - / -
3	135807405	136615405	rs7432375	5.3E-11	rs10935184	7.7E-25	primary	0.93	PCCB	- / - / -
4	170357552	170646052	rs10520163	1.0E-08	rs7438	1.5E-10	primary	0.97	CLCN3	- / - / -
5	45291475	46404116	rs1501357	1.2E-08	rs9292918	4.5E-05	primary	0.94	BRCAT54, RP11-53O19.1	- / - / Adult
6	83779798	84407274	rs3798869	1.2E-09	rs2016358	8.7E-13	primary	0.90	SNAP91	Cerebellar Hemisphere / - / -
6	108875527	109019327	rs9398171	3.4E-08	rs111727905	3.8E-06	primary	0.97	ZNF259P1	- / - / Early mid-prenatal

7	21485312	21545712	rs73060317	6.6E-07	rs12672629	3.6E-05	primary	0.92	SP4	- / - / Early prenatal
8	8088038	10056127	rs2945232	2.0E-08	rs2980441	1.9E-36	primary	0.82	FAM86B3P	- / - / Adolescence
8	26181524	26279124	rs1042992	2.9E-07	rs17055186	5.9E-25	conditional	0.91	SDAD1P1	Testis / - / Adult
8	38020424	38310924	rs57709857	2.3E-07	rs17175814	2.1E-07	primary	0.88	WHSC1L1	- / - / Early prenatal
8	144822546	144871746	rs11784536	1.5E-05	rs12541792	2.7E-37	primary	0.90	FAM83H	Esophagus - Mucosa / Oligodendrocytes / Adolescence
9	26839508	26909408	rs10967586	4.7E-07	rs12345197	1.3E-06	primary	0.80	IFT74	- / - / -
11	46340213	46751213	rs7951870	8.3E-11	rs10160701	5.1E-05	primary	0.88	MDK	- / - / Early mid-prenatal
12	57428314	57497814	rs324017	2.1E-07	rs4559	4.2E-08	conditional	0.91	STAT6	- / Microglia / Adolescence
14	35421614	35847614	rs77477310	1.8E-07	rs1028449	8.1E-04	primary	0.84	RP11-85K15.2	- / - / -
15	78803032	78926732	rs8042374	1.865E-12	rs7171869	1.4E-04	conditional	0.94	IREB2	- / - /Early prenatal
15	84661161	85153461	rs12902973	8.4E-11	rs35677834	1.6E-21	primary	0.80	LOC101929479, RP11-561C5.3	Ovary / - / Early mid-prenatal
15	91416560	91436560	rs4702	2.3E-12	rs4702	9.3E-10	primary	1.00	FURIN	- / Endothelial cells / Adolescence
16	4447751	4596451	rs6500602	2.8E-07	rs3747580	2.3E-16	primary	0.90	CORO7	- / - / -
					rs8046295	4.8E-15	primary	0.89	NMRAL1	- / - / -
16	29924377	30144877	rs12691307	1.3E-10	rs4788203	2.0E-05	primary	0.88	TMEM219	- / - / -
					rs3935873	7.5E-14	primary	0.87	INO80E	- / Neurons / -
					rs4787491	3.5E-05	conditional	0.82	DOC2A	Brain - Cortex / Neurons / Adolescence
16	58669293	58691393	rs12325245	1.1E-08	rs11647976	4.8E-04	primary	0.94	CNOT1	- / - / -
17	17722402	18030202	rs8082590	6.8E-09	rs4072739	3.6E-15	primary	0.92	DRG2	- / - / -
19	11839736	11859736	rs3095917	1.6E-06	rs72986630	2.4E-07	primary	1.00	ZNF823	- / Endothelial cells / Early prenatal
19	19374022	19658022	rs2905426	6.9E-09	rs2965199	9.2E-36	primary	0.87	GATAD2A	- / - / -
19	50067499	50135399	rs56873913	2.2E-07	rs5023763	5.5E-05	primary	0.93	SNRNP70	- / - / -
22	41408556	42689414	rs9607782	6.8E-12	rs9607782	2.0E-04	primary	0.96	RANGAP1	- / - / -

474 **Comparison with Previous Co-localization Analyses**

475

476 In our prior analyses[14] we reported a co-localization analysis of the 108 genome-
477 wide significant schizophrenia GWAS loci and non-conditional eQTL using
478 Sherlock.[40] Those results and our current findings are highly concordant (**Table**
479 **S7**). Eleven loci were reported as co-localized in both analyses. Thirteen loci
480 were co-localized ($PP_{H4} \geq 0.8$) in our analysis but not previously, twelve of which
481 showed suggestive significance in Sherlock ($P<2x10^{-4}$), or in one case involved a
482 conditional eQTL (*SLC35E2*) in our analysis. Six loci were co-localized in the
483 previous study but not in the current analysis; three of these resulted from
484 differences in study design such as GWAS locus definition and eQTL overlap
485 criteria, and two were suggestive in the current analysis ($0.65< PP_{H4} <0.8$). The
486 one remaining discrepant locus (*chr8:143302933-143403527*) was found to co-
487 localize with *TSNARE1* eQTL previously (Sherlock $P=8.24x10^{-7}$), but not here
488 (COLOC2 primary eQTL $PP_{H4}=0.074$, $PP_{H3}=0.93$), and indeed a qualitative
489 comparison of the eQTL and GWAS data did not appear to support co-
490 localization (**Figure S13**).

491

492 In the present analysis we considered not only primary but also conditional eQTL
493 association signals for co-localization with the GWAS, allowing us to detect loci
494 where co-localization may be obscured by multiple association signals in non-
495 conditional eQTL analysis. We also compared our conditional co-localization
496 results with results using non-conditional eQTL analysis, using the same

497 COLOC2 method and SCZ GWAS loci (**Table S8**). Conditional and non-
498 conditional COLOC2 results were highly concordant, with slightly higher PP_{H4}s
499 resulting from the same WABFs because of a higher prior probability of co-
500 localization estimated in the non-conditional COLOC2 analysis. Thirty-five loci
501 were co-localized in both analyses, and five loci that were co-localized in the
502 non-conditional analysis only were all highly suggestive in the conditional
503 analysis ($0.65 < PP_{H4} < 0.8$). The five loci that were co-localized only in the
504 conditional COLOC2 analysis involved conditional and not primary eQTL.
505
506 DISCUSSION
507
508 We utilized genotype and expression data from 467 human post-mortem brain
509 samples from the DLPFC to conduct eQTL mapping analyses, to characterize
510 both primary and conditionally independent eQTL. We then identified co-
511 localization between SCZ GWAS and our eQTL association signals, including
512 conditional eQTL. Our principal findings include four major observations. First,
513 we detect that conditional eQTL are widespread in the brain tissue samples we
514 investigated. In 63% of genes with at least one eQTL, we found multiple
515 statistically independent eQTL (8,136 genes). This demonstrates that genetic
516 variation affecting RNA abundance is incompletely characterized by focusing
517 only on one primary eQTL per gene, which is the case currently for most eQTL
518 studies. We suggest that these conditional eQTL may represent regulatory

519 variation specific to biological contexts not necessarily well represented in the
520 transcriptomic data at hand.
521
522 Second, we find the genomics of conditional eQTL and their genes are consistent
523 with complex, context-specific regulation of gene expression. Conditional eQTL
524 occur farther from transcription start sites than primary eQTL, consistent with
525 effects on distal regulatory elements. Genes with more independent eQTL tend
526 to be larger and span multiple recombination hotspot intervals, and tend to be
527 less constrained at the protein level. While these associations may reflect in part
528 greater power to detect independent eQTL that are not in linkage disequilibrium
529 and that have greater phenotypic variance, they are also consistent with more
530 complex regulation and greater potential for regulatory genetic variation. The
531 strong association of eQTL number with gene expression cis-SNP-heritability
532 shows that conditional eQTL contribute to regulatory genetic variation.
533 Importantly, associations with specificity of expression across tissues,
534 developmental periods, and cell types determined from single-cell RNA
535 sequencing data, suggest that context specificity plays a role in the occurrence of
536 multiple statistically independent eQTL. Cell type specificity is particularly
537 strongly correlated with eQTL number, consistent with those cell types being
538 present in the current tissue-homogenate data.
539
540 Both primary and conditional eQTL are enriched in both active promoter and
541 enhancer regions, and their enrichment in active promoters diminishes with

increasing conditional eQTL order. In other words, conditional eQTL show greater enrichment in enhancers relative to promoters than do primary eQTL. We note that these enrichment analyses are less well powered for conditional eQTL than for primary eQTL, both because of smaller effect sizes of conditional eQTL, and because of statistical error introduced by forward stepwise conditional analyses.[41-43]

Third, we highlight the importance of examining conditional eQTL for co-localization with GWAS. In at least six out of 40 loci showing GWAS-eQTL co-localization, a conditional eQTL signal co-localizes with SCZ risk. If we had considered only primary eQTL in the analyses, these instances of co-localization would have been missed. Conditional eQTL that co-localize with disease risk may reflect regulatory mechanisms that are important in a key developmental period or individual cell type, and may be missed when focusing on primary eQTL discovered in adult whole tissue. As further efforts are made to generate data across ranges of tissues or individual cell-types, we may have a better ability to directly identify regulatory variants specific to these contexts. However if a variant is primarily active in a very specific time point or stimulus condition, capturing data reflecting this condition will remain challenging. Conditional co-localization analysis in well-powered eQTL cohorts may best identify the genes driving these trait associations, though further validation work will be required to understand the mechanism by which the gene contributes to disease risk.

565 Fourth, we have identified a number of candidate genes for which genetic
566 variation for expression co-localizes with genetic variation for schizophrenia risk
567 (Table 2), including cases of co-localization with conditional eQTL. Genetic co-
568 localization is expected if gene expression causally mediates disease risk,
569 although we recognize that co-localization could also result from pleiotropy or
570 linkage, particularly in regions of extensive linkage disequilibrium and haplotype
571 structure.[44; 45] Our analyses prioritize 24 genes among the 111 genes within 23
572 genome-wide significant SCZ loci (GWAS $P < 5x10^{-8}$), and 22 genes in 17
573 suggestive ($P < 1x10^{-6}$) loci. Among the candidates are genes that have
574 previously been implicated in SCZ etiology, such as *FURIN*,[14] as well as
575 alternative candidates in well-known SCZ loci – *DCLK3* in the *TRANK1* locus,[46]
576 *PPM1M* in the *ITIH1* locus,[47] *IREB2* in the *CHRNA3* locus,[27] and *GATAD2A* in
577 the *NCAN* locus.[27] Our candidates include several genes not previously
578 considered as candidates,[27] in some cases - *SLC35E2*, *PTPRU*, *LINC01792*,
579 *DCLK3*, *PPM1M*, *LOC101929479* - because the genes themselves do not
580 overlap the GWAS locus regions but their eQTL do. We also find several non-
581 coding RNA genes - *PROX1-AS1*, *FTCDNL1*, *LINC01792*, *BRCAT54*.
582
583 In an effort to highlight specific developmental periods or cell types for follow-up,
584 we have tabulated expression specificity in GTEx tissues, brain sample cell types
585 from single-cell RNA-seq,[23] and in BrainSpan DLPFC developmental periods, for
586 all identified genes (Table 2, **Table S9**). Their expression contexts show a
587 diversity of patterns, and can provide clues to generate specific hypotheses for

588 functional follow-up of their potential roles in SCZ. Among DLPFC cell types, we
589 find several genes that are specific to neurons and examples of genes specific to
590 oligodendrocytes or endothelial cells (Table 2). Among DLPFC developmental
591 periods, we see diverse expression patterns ranging from early prenatal and
592 broader prenatal, to perinatal (late prenatal and infancy periods), to
593 adolescent/adult expression. We note no clear pattern of correlation between cell
594 type and developmental expression patterns, for example neuronal cell type
595 expressed genes include genes with prenatal, perinatal and postnatal
596 expression. Interestingly, however, all genes broadly expressed across cell types
597 show prenatal expression (**Table S9**).
598
599 *IREB2* (iron regulatory element binding protein 2), highlighted in our analyses as
600 a conditional eQTL hit, is a key regulator of iron homeostasis[48; 49] and has been
601 implicated in neurodegenerative disorders.[50; 51] Mouse *IREB2* homolog Irp2
602 knockouts exhibit impairments in coordination and balance, exploration, and
603 nociception.[49] The *IREB2* locus includes the *CHRNA3-CHRNA5-CHRNB4*
604 nicotinic receptor cluster associated with schizophrenia[27] as well as nicotine
605 dependence and smoking behavior,[52] lung cancer,[53; 54] and COPD.[55] We note
606 that the *IREB2* conditional eQTL is associated with *CHRNA3* and *CHRNA5*
607 expression in cerebellum, caudate and some non-brain tissues in GTEx, but both
608 genes are too lowly expressed for eQTL analysis in the CMC DLPFC samples.
609 Therefore, we cannot rule out the possibility that other genes may be causal of
610 SCZ risk at this locus, perhaps in other brain regions.

611

612 A conditional eQTL for *STAT6* co-localizes with a suggestive SCZ GWAS signal
613 ($P=2x10^{-7}$).[27] The immune related transcription factor STAT6 induces interleukin
614 4 (IL4)-mediated anti-apoptotic activity of T helper cells, and the locus is
615 associated with migraine[56; 57] and brain glioma,[58] as well as several
616 immune/inflammatory diseases.[59-61] STAT6 also activates neuronal
617 progenitor/stem cells and neurogenesis,[62] making it intriguing as an immune-
618 related SCZ candidate given recent observations about the role of complement
619 factor 4 (*C4*) gene as a SCZ risk gene,[63] and prior work potentially implicating
620 microglia.[64] Consistent with a role in immune-mediated synaptic pruning, *STAT6*
621 expression is broadly postnatal and specific to microglia and neurons (**Table S9**).
622

623 Finally, a conditional eQTL for PROX1 Antisense RNA 1 (*PROX1-AS1*; *chr1,*
624 *214Mb*) co-localizes with a suggestive SCZ locus ($P=9.7x10^{-7}$). The Prospero
625 Homeobox 1 (PROX1) transcription factor, involved in development and cell
626 differentiation in several tissues, including oligodendrocytes[65] and GABAnergic
627 interneurons[66] in the brain. This lncRNA has been implicated as aberrantly
628 expressed in several cancers, is upregulated in the cell cycle S-phase, and
629 promotes G1/S transition in cell culture.[67] Like *STAT6*, *PROX1-AS1* expression is
630 specific to neurons and mature oligodendrocytes, and is expressed postnatally
631 (**Table S9**).
632

633 In conclusion, we find that conditional eQTL are wide spread, and are consistent
634 with complex and context specific regulation. Accounting for conditional eQTL
635 leads to new findings of GWAS-eQTL co-localization, and generates specific
636 hypotheses for gene expression possibly mediating disease risk.
637
638 ACKNOWLEDGMENTS
639
640 Data were generated as part of the CommonMind Consortium supported by
641 funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche
642 Ltd and NIH grants R01MH085542, R01MH093725, P50MH066392,
643 P50MH080405, R01MH097276, RO1-MH-075916, P50M096891,
644 P50MH084053S1, R37MH057881 and R37MH057881S1,
645 HHSN271201300031C, AG02219, AG05138 and MH06692. Brain tissue for the
646 study was obtained from the following brain bank collections: the Mount Sinai
647 NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's
648 Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and
649 Tissue Repositories and the NIMH Human Brain Collection Core. CMC
650 Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount
651 Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-
652 Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba
653 (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux
654 (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks),
655 Thomas Lehner, Barbara Lipska (NIMH). ROSMAP study data were provided by

656 the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago.
657 Data collection was supported through funding by NIA grants P30AG10161,
658 R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984,
659 U01AG46152, the Illinois Department of Public Health, and the Translational
660 Genomics Research Institute. The Genotype-Tissue Expression (GTEx) Project
661 was supported by the Common Fund of the Office of the Director of the National
662 Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The
663 data used for the analyses described in this manuscript were obtained from
664 the GTEx Portal on 09/05/16. BrainSpan: Atlas of the Developing Human Brain
665 [Internet]. Funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01,
666 and 1RC2MH089929-01.
667

668 WEB RESOURCES

669

670 AMP-AD Knowledge Portal, (https://www.synapse.org/ampad)
671 BrainSpan atlas, http://www.brainspan.org/
672 CommonMind Consortium data, http://www.synapse.org/CMC
673 CommonMind Consortium ChIP-seq data,
674 https://www.synapse.org/#!Synapse:syn8040458
675 COLOC2, https://github.com/Stahl-Lab-MSSM/coloc2
676 Darmanis et. al. single-cell RNA-seq data, https://www.ncbi.nlm.nih.gov/geo/,
677 accession number GSE67835
678 ExAC Functional Gene Constraint, http://exac.broadinstitute.org/downloads

679 GCTA, http://cnsgenomics.com/software/gcta/

680 GemTools, http://www.wpic.pitt.edu/wpiccompgen/GemTools/GemTools.htm

681 GTEx Portal, https://gtexportal.org/home/

682 HBCC microarray cohort, dbGaP (ID:

683 phs000979.v1.p1), https://www.ncbi.nlm.nih.gov/gap

684 LDetect LD blocks, https://bitbucket.org/nygcresearch/ldetect-data/overview

685 NIH Roadmap Epigenomics Project chromatin state learning,

686 http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state

687 R language and environment, https://www.r-project.org/

688 SNPsnap, https://data.broadinstitute.org/mpg/snpsnap/

689

690 REFERENCES

691

692 1. Gilad, Y., Rifkin, S.A., and Pritchard, J.K. (2008). Revealing the architecture of gene
693 regulation: the promise of eQTL studies. Trends in genetics : TIG 24, 408-415.
694 2. Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping
695 complex disease traits with global gene expression. Nature reviews Genetics
696 10, 184-194.
697 3. Montgomery, S.B., and Dermitzakis, E.T. (2011). From expression QTLs to
698 personalized transcriptomics. Nature reviews Genetics 12, 277-282.
699 4. Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex
700 traits and disease. Nature reviews Genetics 16, 197-212.
701 5. Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M.,
702 von Berg, A., Bufe, A., Rietschel, E., et al. (2007). Genetic variants regulating
703 ORMDL3 expression contribute to the risk of childhood asthma. Nature 448,
704 470-473.
705 6. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U.,
706 Lango Allen, H., Lindgren, C.M., Luan, J., Magi, R., et al. (2010). Association
707 analyses of 249,796 individuals reveal 18 new loci associated with body
708 mass index. Nature genetics 42, 937-948.
709 7. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A.,
710 Zhernakova, A., Heap, G.A., Adany, R., Aromaa, A., et al. (2010). Multiple
711 common variants for celiac disease influencing immune gene expression.
712 Nature genetics 42, 295-302.

713 8. Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., Vermeire,
714 S., Dewit, O., de Vos, M., Dixon, A., et al. (2007). Novel Crohn disease locus
715 identified by genome-wide association maps to a gene desert on 5p13.1 and
716 modulates expression of PTGER4. PLoS genetics 3, e58.
717 9. Aguet, F., Brown, A.A., Castel, S., Davis, J.R., Mohammadi, P., Segre, A.V., Zappala, A.,
718 Abell, N.S., Fresard, L., and Gamazon, E.R. (2016). Local genetic effects on
719 gene expression across 44 human tissues. In. (bioRxiv.
720 10. Jansen, R., Hottenga, J.J., Nivard, M.G., Abdellaoui, A., Laport, B., de Geus, E.J.,
721 Wright, F.A., Penninx, B.W., and Boomsma, D.I. (2017). Conditional eQTL
722 Analysis Reveals Allelic Heterogeneity of Gene Expression. Human molecular
723 genetics Epub ahead of print.
724 11. Zeng, B., Lloyd-Jones, L., Holloway, A., Marigorta, U.M., Metspalu, A., Montgomery,
725 G.W., Esko, T., Brigham, K.L., Quyyumi, A.A., Idaghdour, Y., et al. (2016).
726 Constraints on eQTL fine mapping in the presence of multi-site local
727 regulation of gene expression. In. (bioRxiv.
728 12. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen,
729 H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M., et al. (2009).
730 Common regulatory variation impacts gene expression in a cell type-
731 dependent manner. Science 325, 1246-1250.
732 13. Liu, X., Finucane, H.K., Gusev, A., Bhatia, G., Gazal, S., O'Connor, L., Bulik-Sullivan,
733 B., Wright, F.A., Sullivan, P.F., Neale, B.M., et al. (2017). Functional
734 Architectures of Local and Distal Regulation of Gene Expression in Multiple
735 Human Tissues. American journal of human genetics Epub ahead of print.
736 14. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal,
737 T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene
738 expression elucidates functional impact of polygenic risk for schizophrenia.
739 Nature neuroscience 19, 1442-1453.
740 15. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix
741 operations. Bioinformatics 28, 1353-1358.
742 16. De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton,
743 M.L., Keenan, B.T., Ernst, J., McCabe, C., et al. (2014). Alzheimer's disease:
744 early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other
745 loci. Nature neuroscience 17, 1156-1163.
746 17. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang,
747 H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of
748 64,976 haplotypes for genotype imputation. Nature genetics 48, 1279-1283.
749 18. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I.,
750 Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype
751 imputation service and methods. Nature genetics 48, 1284-1287.
752 19. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for
753 genome-wide complex trait analysis. American journal of human genetics 88,
754 76-82.
755 20. Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-
756 Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide
757 midrange transcription profiles reveal expression level relationships in
758 human tissue specification. Bioinformatics 21, 650-659.

759 21. Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2016). A benchmark of
760 gene expression tissue-specificity metrics. Briefings in bioinformatics 18,
761 205-214.
762 22. Consortium, G. (2013). The Genotype-Tissue Expression (GTEx) project. Nature
763 genetics 45, 580-585.
764 23. Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden
765 Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain
766 transcriptome diversity at the single cell level. Proc Natl Acad Sci U S A 112,
767 7285-7290.
768 24. Lin, G.N., Corominas, R., Lemmens, I., Yang, X., Tavernier, J., Hill, D.E., Vidal, M.,
769 Sebat, J., and Iakoucheva, L.M. (2015). Spatiotemporal 16p11.2 protein
770 network implicates cortical late mid-fetal brain development and KCTD13-
771 Cul3-RhoA pathway in psychiatric diseases. Neuron 85, 742-754.
772 25. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen,
773 A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015).
774 Integrative analysis of 111 reference human epigenomes. Nature 518, 317-
775 330.
776 26. Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state
777 discovery and characterization. Nat Methods 9, 215-216.
778 27. Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological
779 insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-
780 427.
781 28. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P.,
782 Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional
783 visualization of genome-wide association scan results. Bioinformatics 26,
784 2336-2337.
785 29. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace,
786 C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of
787 genetic association studies using summary statistics. PLoS genetics 10,
788 e1004383.
789 30. Pickrell, J.K., Berisa, T., Liu, J.Z., Segurel, L., Tung, J.Y., and Hinds, D.A. (2016).
790 Detection and interpretation of shared genetic influences on 42 human traits.
791 Nature genetics 48, 709-717.
792 31. Wakefield, J. (2009). Bayes factors for genome-wide association studies:
793 comparison with P-values. Genet Epidemiol 33, 79-86.
794 32. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage
795 disequilibrium blocks in human populations. Bioinformatics 32, 283-285.
796 33. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T.,
797 O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016).
798 Analysis of protein-coding genetic variation in 60,706 humans. Nature 536,
799 285-291.
800 34. Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited.
801 Trends in genetics : TIG 29, 569-574.
802 35. Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M.,
803 Arindrarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.J., et al. (2017).

804 Identification of context-dependent expression quantitative trait loci in
805 whole blood. Nature genetics 49, 139-145.
806 36. Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh,
807 N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic
808 fine mapping of causal autoimmune disease variants. Nature 518, 337-343.
809 37. Tak, Y.G., and Farnham, P.J. (2015). Making sense of GWAS: using epigenomics
810 and genome engineering to understand the functional relevance of SNPs in
811 non-coding regions of the human genome. Epigenetics Chromatin 8, 57.
812 38. Brown, C.D., Mangravite, L.M., and Engelhardt, B.E. (2013). Integrative modeling
813 of eQTLs and cis-regulatory elements suggests mechanisms underlying cell
814 type specificity of eQTLs. PLoS genetics 9, e1003649.
815 39. Gaffney, D.J., Veyrieras, J.B., Degner, J.F., Pique-Regi, R., Pai, A.A., Crawford, G.E.,
816 Stephens, M., Gilad, Y., and Pritchard, J.K. (2012). Dissecting the regulatory
817 architecture of gene expression QTLs. Genome Biol 13, R7.
818 40. He, X., Fuller, C.K., Song, Y., Meng, Q., Zhang, B., Yang, X., and Li, H. (2013).
819 Sherlock: detecting gene-disease associations by matching patterns of
820 expression QTL and GWAS. American journal of human genetics 92, 667-680.
821 41. Servin, B., and Stephens, M. (2007). Imputation-based analysis of association
822 studies: candidate regions and quantitative traits. PLoS genetics 3, e114.
823 42. Spain, S.L., and Barrett, J.C. (2015). Strategies for fine-mapping complex traits.
824 Human molecular genetics 24, R111-119.
825 43. Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B., and Eskin, E. (2014).
826 Identifying causal variants at loci with multiple signals of association.
827 Genetics 198, 497-508.
828 44. Guo, H., Fortune, M.D., Burren, O.S., Schofield, E., Todd, J.A., and Wallace, C.
829 (2015). Integration of disease association and eQTL data using a Bayesian
830 colocalisation approach highlights six candidate causal genes in immune-
831 mediated diseases. Human molecular genetics 24, 3305-3313.
832 45. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery,
833 G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., et al. (2016). Integration of
834 summary data from GWAS and eQTL studies predicts complex trait gene
835 targets. Nature genetics 48, 481-487.
836 46. Schizophrenia Psychiatric Genome-Wide Association Study, C. (2011). Genome-
837 wide association study identifies five new schizophrenia loci. Nature genetics
838 43, 969-976.
839 47. Sleiman, P., Wang, D., Glessner, J., Hadley, D., Gur, R.E., Cohen, N., Li, Q.,
840 Hakonarson, H., and Janssen, C.N.G.W.G. (2013). GWAS meta analysis
841 identifies TSNARE1 as a novel Schizophrenia / Bipolar susceptibility locus.
842 Scientific reports 3, 3075.
843 48. Kim, W.J., and Lee, S.D. (2015). Candidate genes for COPD: current evidence and
844 research. International journal of chronic obstructive pulmonary disease 10,
845 2249-2255.
846 49. Zumbrennen-Bullough, K.B., Becker, L., Garrett, L., Holter, S.M., Calzada-Wack, J.,
847 Mossbrugger, I., Quintanilla-Fend, L., Racz, I., Rathkolb, B., Klopstock, T., et al.
848 (2014). Abnormal brain iron metabolism in Irp2 deficient mice is associated
849 with mild neurological and behavioral impairments. PloS one 9, e98072.

850 50. Dusek, P., Jankovic, J., and Le, W. (2012). Iron dysregulation in movement
851 disorders. Neurobiol Dis 46, 1-18.
852 51. Rouault, T.A. (2013). Iron metabolism in the CNS: implications for
853 neurodegenerative diseases. Nat Rev Neurosci 14, 551-564.
854 52. Thorgeirsson, T.E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K.P.,
855 Manolescu, A., Thorleifsson, G., Stefansson, H., Ingason, A., et al. (2008). A
856 variant associated with nicotine dependence, lung cancer and peripheral
857 arterial disease. Nature 452, 638-642.
858 53. Amos, C.I., Wu, X., Broderick, P., Gorlov, I.P., Gu, J., Eisen, T., Dong, Q., Zhang, Q.,
859 Gu, X., Vijayakrishnan, J., et al. (2008). Genome-wide association scan of tag
860 SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nature
861 genetics 40, 616-622.
862 54. Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D.,
863 Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., et al. (2008).
864 A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor
865 subunit genes on 15q25. Nature 452, 633-637.
866 55. DeMeo, D.L., Mariani, T., Bhattacharya, S., Srisuma, S., Lange, C., Litonjua, A.,
867 Bueno, R., Pillai, S.G., Lomas, D.A., Sparrow, D., et al. (2009). Integration of
868 genomic and genetic approaches implicates IREB2 as a COPD susceptibility
869 gene. American journal of human genetics 85, 493-502.
870 56. Anttila, V., Winsvold, B.S., Gormley, P., Kurth, T., Bettella, F., McMahon, G., Kallela,
871 M., Malik, R., de Vries, B., Terwindt, G., et al. (2013). Genome-wide meta-
872 analysis identifies new susceptibility loci for migraine. Nature genetics 45,
873 912-917.
874 57. Gormley, P., Anttila, V., Winsvold, B.S., Palta, P., Esko, T., Pers, T.H., Farh, K.H.,
875 Cuenca-Leon, E., Muona, M., Furlotte, N.A., et al. (2016). Meta-analysis of
876 375,000 individuals identifies 38 susceptibility loci for migraine. Nature
877 genetics 48, 856-866.
878 58. Ruan, Z., Zhao, Y., Yan, L., Chen, H., Fan, W., Chen, J., Wu, Q., Qian, J., Zhang, T.,
879 Zhou, K., et al. (2011). Single nucleotide polymorphisms in IL-4Ra, IL-13 and
880 STAT6 genes occurs in brain glioma. Front Biosci (Elite Ed) 3, 33-45.
881 59. Sleiman, P.M., Wang, M.L., Cianferoni, A., Aceves, S., Gonsalves, N., Nadeau, K.,
882 Bredenoord, A.J., Furuta, G.T., Spergel, J.M., and Hakonarson, H. (2014). GWAS
883 identifies four novel eosinophilic esophagitis loci. Nat Commun 5, 5593.
884 60. Granada, M., Wilk, J.B., Tuzova, M., Strachan, D.P., Weidinger, S., Albrecht, E.,
885 Gieger, C., Heinrich, J., Himes, B.E., Hunninghake, G.M., et al. (2012). A
886 genome-wide association study of plasma total IgE concentrations in the
887 Framingham Heart Study. J Allergy Clin Immunol 129, 840-845 e821.
888 61. Bonnelykke, K., Matheson, M.C., Pers, T.H., Granell, R., Strachan, D.P., Alves, A.C.,
889 Linneberg, A., Curtin, J.A., Warrington, N.M., Standl, M., et al. (2013). Meta-
890 analysis of genome-wide association studies identifies ten loci influencing
891 allergic sensitization. Nature genetics 45, 902-906.
892 62. Bhattarai, P., Thomas, A.K., Cosacak, M.I., Papadimitriou, C., Mashkaryan, V., Froc,
893 C., Reinhardt, S., Kurth, T., Dahl, A., Zhang, Y., et al. (2016). IL4/STAT6
894 Signaling Activates Neural Stem Cell Proliferation and Neurogenesis upon
895 Amyloid-beta42 Aggregation in Adult Zebrafish Brain. Cell Rep 17, 941-948.

896 63. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N.,
897 Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al. (2016). Schizophrenia
898 risk from complex variation of complement component 4. Nature 530, 177-
899 183.
900 64. Schafer, D.P., Lehrman, E.K., Kautzman, A.G., Koyama, R., Mardinly, A.R.,
901 Yamasaki, R., Ransohoff, R.M., Greenberg, M.E., Barres, B.A., and Stevens, B.
902 (2012). Microglia sculpt postnatal neural circuits in an activity and
903 complement-dependent manner. Neuron 74, 691-705.
904 65. Kato, K., Konno, D., Berry, M., Matsuzaki, F., Logan, A., and Hidalgo, A. (2015).
905 Prox1 Inhibits Proliferation and Is Required for Differentiation of the
906 Oligodendrocyte Cell Lineage in the Mouse. PloS one 10, e0145334.
907 66. Miyoshi, G., Young, A., Petros, T., Karayannis, T., McKenzie Chang, M., Lavado, A.,
908 Iwano, T., Nakajima, M., Taniguchi, H., Huang, Z.J., et al. (2015). Prox1
909 Regulates the Subtype-Specific Development of Caudal Ganglionic Eminence-
910 Derived GABAergic Cortical Interneurons. J Neurosci 35, 12869-12889.
911 67. Yang, F., Yi, F., Zheng, Z., Ling, Z., Ding, J., Guo, J., Mao, W., Wang, X., Wang, X., Ding,
912 X., et al. (2012). Characterization of a carcinogenesis-associated long non-
913 coding RNA. RNA Biol 9, 110-116.
914
915