Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS

Amanda Dobbyn,1,2 Laura M. Huckins,1,2 James Boocock,3 Laura G. Sloofman,1 Benjamin S. Glicksberg,7,4,5 Claudia Giambartolomei,6 Gabriel E. Hoffman,1,2,4 Thanneer M. Perumal,7 Kiran Girdhar,1,2 Yan Jiang,1,8 Towfique Raj,2,4,9 Douglas M. Ruderfer,10 Robin S. Kramer,11 Dalila Pinto,5,4,8,12,13 the CommonMind Consortium, Schahram Akbarian,1,8 Panos Roussos,1,4,8 Enrico Domenici,14,15 Bernie Devlin,16 Pamela Sklar,1,2,4,8,17 Eli A. Stahl,1,4,18,* and Solveig K. Sieberts7,18,*

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 variants underlie both GWAS and eQTL signals. Most analyses, however, consider only the marginal eQTL signal, rather than dissect this signal into multiple conditionally independent signals 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 (e.g., tissue-, cell type-, or developmental time point-specific) regulation of gene expression. Integrating the 2014 Psychiatric Genomics Consortium schizophrenia (SCZ) GWAS and CMC primary and conditional eQTL data reveals 40 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, identify novel genes associated with schizophrenia risk, and provide specific hypotheses for their functional follow-up.

Introduction

Significant advances in understanding the genetic architecture of schizophrenia (MIM: 181500) have occurred within the last 10 years. However, for common variants identified in genome-wide association studies (GWASs), the success in locus identification is not yet matched by an understanding of their underlying basic mechanism or effect on pathophysiology. Expression quantitative trait loci (eQTL), which are responsible for a significant proportion of variation in gene expression, could serve as a link between the numerous non-coding genetic associations that have been identified in GWASs and susceptibility to common diseases directly through their association with gene expression regulation.1–4 Accordingly, results from eQTL mapping studies have been successfully utilized to identify genes and causal variants from GWASs for various complex phenotypes, including asthma (MIM: 600807), body mass index (MIM: 601665), celiac disease (MIM: 212750), and Crohn disease (MIM: 266600).5–8 Studies integrating eQTL and GWAS data have almost exclusively used marginal association statistics which typically represent the primary, or most significant, eQTL signal when assessing co-localization with GWASs, ignoring other SNPs that affect expression independently of the primary eQTL for a given gene. However, recent findings indicating that conditionally independent eQTL are widespread9–12 motivate examination of the extent to which considering conditional eQTL may provide additional power to identify likely causal genes in a GWAS locus. Recent reports provide evidence that conditional eQTL are less frequently shared across tissues than primary

1Division of Psychiatric Genomics, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 2Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 3Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90024, USA; 4Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 5Department for Next Generation Healthcare, Mount Sinai Health System, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 6Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; 7Systems Biology, Sage Bionetworks, Seattle, WA 98109, USA; 8Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 9Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 10Division of Genetic Medicine, Department of Medicine, Psychiatry and Biomedical Informatics, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37235, USA; 11Human Brain Collection Core, National Institute of Mental Health, Bethesda, MD 20892, USA; 12Department of Psychiatry and Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 13Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 14Laboratory of Neurogenomic Biomarkers, Centre for Integrative Biology (CIBIO), University of Trento, Trento, Italy; 15The Microsoft Research - University of Trento Centre for Computational and Systems Biology (COSBI), Rovereto (TN), Italy; 16Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA
17Deceased November 20, 2017
18These authors contributed equally to this work
*Correspondence: eli.stahl@mssm.edu (E.A.S.), solvey.sieberts@sagebase.org (S.K.S.)
© 2018 The Authors. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
eQTL\textsuperscript{10} and, like tissue- and cell-type-specific eQTL, are often found more distally to the genes they regulate.\textsuperscript{10,13,14} These lines of evidence suggest that conditionally independent eQTL may contribute to tissue-specific or other context-specific gene regulation (e.g., specific to a particular cell type, developmental stage, or stimulation condition). One mechanism by which disease risk could potentially be mediated by a conditional eQTL is the disruption of a tissue-specific enhancer by a given variant, leading to the dysregulation of the relevant eGene in only the tissue for which the enhancer is specific. For example, an eQTL affecting Parkinson disease risk through expression of SNCA was recently shown to act through the disruption of an enhancer;\textsuperscript{15} if this enhancer is specific to a disease-relevant cell type, such as nerve cells of the substantia nigra, then it could manifest as a conditional eQTL since it would be only partially represented in brain homogenate.

Here, we leveraged genotype and dorsolateral prefrontal cortex (DLPFC) expression data provided by the CommonMind Consortium (CMC) to elucidate the role of conditional eQTL in the etiology of schizophrenia (SCZ). Currently comprising the largest existing postmortem brain genomic resource at nearly 600 samples, the CMC is generating and making publicly available an unprecedented array of functional genomic data, including gene expression (RNA sequencing), histone modification (chromatin immunoprecipitation [ChIP-seq]), and SNP genotypes, from individuals with psychiatric disorders as well as unaffected controls.\textsuperscript{16} We utilized SNP dosage and RNA-sequencing (RNA-seq) data from the CMC to identify primary and conditionally independent eQTL. We then characterized the resulting eQTL on various genomic attributes including distance to transcription start site and their genes’ specificities across tissues, cell types, and developmental periods. In addition, we quantified enrichment of primary and conditional eQTL in promoter and enhancer functional genomic elements inferred from epigenomic data. Finally, we isolated each independent eQTL signal by conducting a series of “all-but-one” conditional analyses for genes with multiple independent eQTL and then assessed the overlap between all eQTL association signals and the schizophrenia GWAS signals.

### Material and Methods

**CommonMind Consortium Data**

We used pre-QC’ed genotype and expression data from the CommonMind Consortium, and detailed information on quality control, data adjustment, and normalization procedures can be found in Fromer et al.\textsuperscript{16} Briefly, samples were genotyped at 958,178 markers using the Illumina Infinium HumanOmniplex Exome array and markers were removed on the basis of having no alternate alleles, having a genotyping call rate \( \leq 0.98\), or having a Hardy-Weinberg p value \( < 5 \times 10^{-5} \). After QC, 668 individuals genotyped at 767,368 markers were used for imputation. Phasing was performed on each chromosome using ShapeIt v2.1.790,\textsuperscript{17} and variants were imputed in 5 Mb segments with Impute v2.3.118 using the 1000 Genomes Phase 1 integrated reference panel,\textsuperscript{19} excluding singleton variants. After phasing and imputation, then filtering out variants with INFO \( < 0.8 \) or MAF \( < 0.05 \), the number of markers included in the analysis totaled approximately 6.4 million. Gene expression was assayed via RNA-seq using 100 base pair paired-end reads and was mapped to human Ensembl gene reference (v.70) using TopHat v.2.0.9 and Bowtie v.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 was considered for analysis. The expression data was voom-adjusted for both known covariates (RIN, library batch, institution, diagnosis, post-mortem interval, and sex) and 20 surrogate variables identified via surrogate variable analysis (SVA).\textsuperscript{20} After the removal of samples that did not pass RNA sample QC (including but not limited to: having RIN \( < 5.5 \), having less than 50 million total reads or more than 5% of reads aligning to rRNA, having any discordance between genotyping and RNA-seq data, and having RNA outlier status or evidence for contamination) and retaining only genetically identified European-ancestry individuals, a total of 467 samples was used for downstream analyses. These 467 individuals comprised 209 SCZ-affected case subjects, 52 AFF (bipolar, major depressive disorder, or mood disorder, unspecified)-affected case subjects, and 206 control subjects.

**eQTL Identification**

An overview of our workflow can be found in Figure S1. First, to identify primary and conditional cis-eQTL, we conducted forward stepwise conditional analysis implemented in MatrixEQTL\textsuperscript{21} using genotype data at 6.4 million markers and RNA-seq data for 16,423 genes. FDR was initially assessed using the Benjamini-Hochberg algorithm across all cis-eQTL tests within each chromosome. FDR was not re-assessed at each conditional step; instead, a fixed p value threshold was used as the inclusion criteria in the stepwise model selection. For each gene with at least one cis-eQTL (gene \( \pm 1 \) Mb) association at a 5% false discovery rate (FDR), the most significant SNP was added as a covariate in order to identify additional independent associations (considered significant if the p value achieved was less than that corresponding to the initial 5% FDR for primary eQTL). This procedure was repeated iteratively until no further eQTL met the p value threshold criteria. We used a linear regression model, adjusting for diagnosis and five ancestry covariates inferred by GemTools. Following eQTL identification, only autosomal eQTL were retained for downstream analyses.

**Replication in Independent Datasets**

Replication was performed in the HBBC microarray cohort (dbGaP: phs000979, see Web Resources) and in the ROSMAP\textsuperscript{22} RNA-seq cohort by fitting the stepwise regression models identified in the CMC data. For cases in which a marker was unavailable in the replication cohort, all models including that marker (i.e., for that eQTL and higher-order eQTL conditional on it, for a given gene) were omitted from replication.

Data from the HBBC cohort was QC’ed and normalized as described in Fromer et al.\textsuperscript{16} DLPFC tissue was profiled on the Illumina HumanHT-12_V4 BeadChip and normalized in an analogous manner to the CMC data. Genotypes were obtained using the HumanHap650V2 or Human1MDuov3 chips and imputed using the 1000 Genomes Phase 1 reference panel. Replication of the eQTL models was performed on 279 genetically inferred European-ancestry samples (76 control subjects, 72 SCZ-affected
ROSMAP data were obtained from the AMP-AD Knowledge Portal (see Web Resources). Quantile normalized FPKM expression values were adjusted for age of death, RIN, PMI, and 31 hidden confounders from SVA, conditional on diagnosis. Only genes with FPKM > 0 in more than 50 samples were retained. QC’ed genotypes were also obtained from the AMP-AD Knowledge Portal and imputed to the HapMap Reference Consortium (v.1.1) reference panel via the Michigan Imputation Server. Only markers with imputation quality score R² ≥ 0.7 were considered in the replication analysis. GenTools was used to infer ancestry components as was done for the CMC data above. After QC, 494 samples were used for eQTL replication in a linear regression model that also adjusted for diagnosis (Alzheimer disease, mild cognitive impairment, no cognitive impairment, and other) and four ancestry components.

### Modeling Number of eQTL per Gene on Genomic Features

We considered three genomic features (gene length, number of LD blocks in the cis-region, and genic constraint score) for our modeling analyses. Gene lengths were calculated using Ensembl gene locations. We obtained LD blocks from the LDetect Bitbucket site to tally the number of blocks overlapping each gene’s cis-region (gene ± 1 Mb). We obtained loss-of-function-based genic constraint scores from the Exome Aggregation Consortium (ExAC). A negative binomial generalized linear regression model was used to model the number of eQTL per gene based on the above variables; results were qualitatively the same using linear regression of Box-Cox transformed eQTL numbers. Backward-forward stepwise regression using the full model with interaction terms for these three variables was used to determine the relationship between genomic attributes and eQTL number. These analyses were implemented in R. cis-heritability of gene expression was estimated using the same CMC data that were used for eQTL detection, including all markers in the cis-region (gene ± 1 Mb). SNP-heritability estimates were then added to the modeling procedure described above.

Tissue, cell type, and developmental time point specificity were measured using the expression specificity metric Tau. Tissue specificity for each gene was calculated using publicly available expression data for 53 tissues from the GTEx project (release V6p). Expression for each tissue was summarized as the log2 of the median expression plus one, and then used to calculate tissue specificity Tau. Cell type specificity for each gene was computed using publicly available single-cell RNA-sequencing expression data generated from human cortex and hippocampus tissues. Raw expression counts for 285 cells comprising six major cell types of the brain were obtained from GEO (GSE67835) and counts data were library normalized to CPM. Expression for each cell type was summarized as the log2 of the mean expression plus one, and then used to compute cell type specificity Tau. Developmental time point specificity for each gene was calculated using publicly available DLPFC expression data for 27 time points, clustered into eight biologically relevant groups, from the BrainSpan atlas (see Web Resources). Eight developmental periods were defined as follows: early prenatal (8–12 pcw), early mid-prenatal (13–17 pcw), late mid-prenatal (19–24 pcw), late prenatal (25–37 pcw), infancy (4 months–1 year), childhood (2–11 years), adolescence (13–19 years), and adulthood (21+ years). Expression for each time point was summarized as the log2 of the median expression plus one and then used to calculate developmental period specificity Tau. Each Tau was added to the above model for eQTL number individually, as well as all together.

### Enrichment Analyses

We divided eQTL into separate subgroups by stepwise conditional order (first, second, and greater than second) and created sets of matched SNPs drawn from the SNPsnap database for each subgroup, matching on minor allele frequency, gene density (number of genes within 1 Mb of the SNP), distance from SNP to TSS of the nearest gene, and LD (number of LD-partners within R² ≥ 0.8). For each subgroup of eQTL, we performed a logistic regression of status as eQTL or matched SNP on overlap with functional annotation, including the four SNP matching parameters as covariates. Enrichment was taken as the regression coefficient estimate, interpretable as the log-odds ratio for being an eQTL given a functional annotation. Functional annotations tested included: brain promoters and enhancers (union of all brain region TssA and EnhþEnhG intervals, respectively, from the NIH Roadmap Epigenomics Project ChromHMM core 15-state model), brain-specific promoters and enhancers (the union of all brain region TssA and EnhþEnhG intervals, excluding those present in seven other non-brain tissues/cell types: primary T helper cells from peripheral blood, osteoblast primary cells, HUES64 cells, adipose nuclei, liver, NHLF lung fibroblast primary cells, and NHEK-epidermal keratinocyte primary cells), and pre-frontal cortex (PFC) neuronal (NeuN+) and non-neuronal (NeuN–) nucleus H3K4me3 and H3K27ac ChIP-seq marks from the CMC. For each data source, active promoter and enhancer (or H3K4me3 and H3K27ac) annotations were tested for enrichment jointly. This analysis was repeated but restricting to matched SNPs located within 1 Mb of any of the 16,423 genes that were tested for eQTL, in order to determine whether the enrichment estimates were inflated due to the proximity of our primary and conditional eQTL to brain-expressed genes, which may be more likely to occur near active regulatory regions in the brain. In addition, to ensure that any enrichment patterns observed were not due to varying effect size among primary and conditional eQTL, the enrichment analyses were also carried out taking into account the variance in expression explained by each eQTL. Variance explained (R²) was estimated using the variancePartition R package, and eQTL were stratified into three R² bins: bin 1, 1 x 10⁻² ≤ R² ≤ 1.75 x 10⁻²; bin 2, 1.75 x 10⁻² ≤ R² ≤ 2.25 x 10⁻²; and bin 3, 2.25 x 10⁻² ≤ R² ≤ 3 x 10⁻². Logistic regression of status as eQTL or matched SNP was then carried out separately for each R² bin, within each eQTL order.

### Conditional eQTL Analyses

In order to isolate each conditionally independent cis-eQTL association, we carried out a series of “all-but-one” conditional analyses, implemented within MatrixEQTL, for each gene possessing more than one independent eQTL. As these conditional eQTL signals were to be used to test for co-localization with the SCZ GWAS signals, we limited these analyses to those genes (346 in total) with eQTL overlapping GWAS loci. For each of these genes, we conducted an all-but-one analysis for each independent eQTL by regressing the given gene’s expression data on the dosage data, including all of the other independent eQTL for that gene as covariates in addition to diagnosis and five ancestry components. For example, three conditional analyses would be conducted for a
gene with three independent eQTL: one analysis conditioning on the secondary and tertiary eQTL, one analysis conditioning on the primary and tertiary, and one analysis conditioning on the primary and secondary. In this manner we generated summary statistics for each independent eQTL in isolation, conditional on all of the other independent eQTL for that gene.

Co-localization Analyses
For our co-localization analyses, we used summary statistics and genomic intervals from the 2014 Psychiatric Genomics Consortium (PGC) SCZ GWAS.35 We included 217 loci at a p value threshold of 1 × 10^{-6} (excluding the MHC locus), defined these loci by their LD r^2 ≥ 0.6 with the lead SNP, and then merged overlapping loci. GWAS and eQTL signatures were qualitatively compared using p value-p value (P-P) plots, rendered in R, and LocusZoom.36

Multiple methods that aim to identify GWAS-eQTL co-localized loci are currently available.37-42 We chose to further develop coloc for our co-localization analyses for several reasons: (1) it uses data from all SNPs within a locus; (2) it avoids the computational burden or approximate results of Bayesian inferential methods for causal variants,31,42 which rely on reference panel estimates of linkage disequilibrium (LD); and (3) and it has been widely used43-46 including in direct comparisons of GWAS-eQTL co-localization methods.42,46 We tested for co-localization using an updated version of coloc R functions, which we name coloc2 (see Web Resources), and incorporated several improvements to the method. First, coloc2 pre-processes data by aligning eQTL and GWAS summary statistics for each eQTL cis-region. Second, coloc2 model optionally incorporates changes implemented in gwas-pw.43 Briefly, we implemented likelihood estimation of mixture proportions of five hypotheses (H_0, no association; H_1, GWAS association only; H_2, eQTL association only; H_3, both but not co-localized; and H_4, both and co-localized) from genome-wide data. Coloc2 uses these proportions as priors (or optionally, coloc default or user-specified priors) in the empirical Bayesian calculation of the posterior probability of co-localization for each locus (eQTL cis-region). Coloc2 averages per-SNP Wakefield asymptotic Bayes factors (WABF)47 across three different values for the WABF prior variance term, 0.01, 0.1, and 0.5, and provides options for specifying phenotypic variance, estimating it from case-control proportions or estimating it from the data.

Results

Identification of eQTL
Primary and conditional eQTL were identified using genotype and RNA-seq data from the CommonMind Consortium post-mortem DL-PFC samples (467 European-ancestry case and control subjects).16 We identified 12,813 primary and 16,082 conditional eQTL, totaling 28,895 independent eQTL. Of the genes tested, 81% (12,813 of 15,817 autosomal genes) had at least one eQTL and 63% of these (51% of all genes) also had at least one conditional eQTL, with an average of 1.83 independent eQTL per gene (2.26 among those with at least one eQTL) (Figure 1A). Conversely, when examining the distributions for the number of genes whose expression was affected by each eQTL (Table S1), the majority of eQTL were specific for a single gene, and only a small fraction of eQTL, 1.47%, affected more than one gene, with a maximum of six genes affected by a single eQTL.

We tested for replication of conditional eQTL in two independent datasets, the National Institute of Mental Health’s Human Brain Collection Core (HBCC, n = 279, microarray expression data) and the Religious Orders Study/Memory and Aging Project22 (ROSMAP, n = 494, RNA-seq expression). For each gene the same models were evaluated that were identified in forward-stepwise conditional analysis in the CMC data. We observed significant evidence of replication for both primary and conditional eQTL in the HBCC and ROSMAP post-mortem brain cohorts (Table S2). The estimated proportion of true associations (π_1) in ROSMAP was 0.57 and 0.26 for primary and conditional eQTL, respectively; in HBCC π_1 was 0.46 and 0.20 for primary and conditional eQTL. Therefore, replication was stronger for primary than for conditional eQTL, as expected given their stronger effect sizes. Replication rates were somewhat higher in the RNA-seq ROSMAP data than in HBCC.

Genomic Characterization of Primary and Conditional eQTL
The features for which primary and conditional eQTL and their respective eGenes displayed identifiable differences included distance from eQTL to its gene’s transcription start site (TSS), gene length, LD blocks per genic cis-region, genic constraint score, and genic cis-SNP-heritability. According to prior results, eQTL that are shared across tissues and cell types tend to be located closer to transcription start sites than context-specific eQTL;13,14 we therefore first examined the relationship between primary or conditional eQTL status and distance to genic TSS. Primary eQTL fall closer to the TSS than conditional eQTL (Figure 1C): primary eQTL occur at a median distance of 70.4 kb from the TSS versus a median distance of 302 kb for conditional eQTL. This difference holds true even more proximally to the TSS (Figure S2); 8.1% and 2.5% of primary and conditional eQTL, respectively, fall within 3 kb of the TSS. We next characterized the relationship between the number of independent eQTL per gene and three different genomic features: gene length, number of LD blocks46 in the gene’s cis-region (± 1 Mb), and Exome Aggregation Consortium (ExAC) genic constraint score,49 including possible interactions. The best multivariate model for eQTL number included gene length, number of LD blocks, and genic constraint as predictors, as well as a gene length-LD blocks interaction (Table 1). The number of independent eQTL was positively correlated with gene length and number of LD blocks and negatively correlated with genic constraint score (Figure S3). We then examined the variance of gene expression explained by cis-region SNPs, or cis-SNP-heritability, estimated by linear mixed model variance component analysis25 (Figure S4). We found a strong effect of estimated cis-heritability on number of independent eQTL (Table 1, Figure S5). In a joint model with
cis-SNP-heritability, the main effects of gene length, number of LD blocks, and genic constraint on eQTL number remained at least nominally significant.

We then addressed whether genes with conditional eQTL exhibit greater context specificity as measured by the robust expression specificity metric Tau. We calculated Tau across 53 tissues from the Genotype-Tissue Expression (GTEx) project, across 6 DLPFC cell types (astrocytes, endothelial cells, microglia, neurons, oligodendrocytes, and oligodendrocyte progenitor cells) from single-cell RNA-seq, and across 8 developmental periods (early prenatal, early mid-prenatal, late mid-prenatal, late prenatal, infant, child, adolescent, and adult) from the BrainSpan atlas DLPFC RNA-seq data. We confirmed that higher values of Tau reflect expression specificity by comparing the distributions of all three Tau measures for all genes with the distributions for a subset of housekeeping genes (Figure S6). We found positive correlations between eQTL number and tissue, cell type, and developmental time point specificities (Figure 1B, Table 1, Table S3, Figure S7). In a joint model, the strongest correlation was with DLPFC cell type Tau, which is consistent with previous data demonstrating tissue-specific, cell type-dependent expression in blood; however, we note that all three Tau sets were inter-correlated (Table S3).

Epigenetic Enrichment Analyses

One way in which eQTL may affect gene expression is through alteration of cis-regulatory elements such as promoters and enhancers. Putative causal eSNPs have been shown to be enriched in genomic regions containing functional annotations such as DNase hypersensitive sites, transcription factor binding sites, promoters, and enhancers. Our observation that conditional eQTL fall farther from transcription start sites than primary eQTL led us to hypothesize that primary eQTL may affect transcription levels by altering functional sites in promoters whereas conditional eQTL may do so by altering more distal regulatory elements such as enhancers. We therefore assessed enrichment of primary and conditional eQTL in brain active promoter (TssA) and enhancer (merged Enh and EnhG) states derived from the NIH Roadmap Epigenomics Project, and in H3K4me3 and H3K27ac neuronal (NeuN+) and non-neuronal (NeuN-) ChIP-seq peaks from a subset of the CMC post-mortem DLPFC samples. The overlap of H3K4me3 and H3K27ac ChIP-seq peaks was used as a proxy for active promoters, and H3K27ac peaks that do not overlap H3K4me3 peaks were used as a (relatively non-specific) proxy for enhancers.

We performed logistic regression of SNP status (eQTL versus random matched SNP) on overlap with functional
annotations, separately for each eQTL order (primary, secondary, and greater than secondary).

Primary and conditional eQTL were significantly enriched in both promoter and enhancer chromatin states from REMC brain and CMC DLPFC tissues, with greatest enrichments overall observed in PFC neuronal (NeuN) promoters and enhancers (Figure 2, Table S4). We found that whereas active promoter enrichments in all tissue/cell types markedly decreased with higher conditional order of eQTL, enhancer enrichments either only slightly decreased (REMC brain and PFC NeuN, Figures 2A and 2C) or remained level (REMC brain-specific, Figure 2B). This greater decrease in enrichment for promoters compared to enhancers with increasing eQTL order was not confounded by an excess of eQTL near brain-expressed genes in comparison to matched SNPs (Figure S8, Table S5) and furthermore was not an artifact of varying effect size with eQTL order; the same overall pattern was observed when stratifying eQTL by variance in expression explained (R²) and comparing enrichment across eQTL order, within each R² bin (Figures S9–S12, Table S6).

**eQTL Co-localization with SCZ GWAS**

We performed co-localization analyses in order to evaluate the extent of overlap between eQTL and GWAS signatures in schizophrenia and to identify putative causal genes from GWAS associations. Considering 217 loci (Table S7) with lead SNPs reaching a significance threshold of p < 1 × 10⁻⁶ from the 2014 Psychiatric Genomics Consortium (PGC) schizophrenia GWAS, we tabulated the number of primary and conditional eQTL falling within GWAS loci. A total of 114 out of 217 loci contained primary and/or conditional eQTL for 346 genes; 110 of these genes had one eQTL only and 236 genes had more than one independent eQTL.

To quantitatively compare the SCZ GWAS and eQTL association signatures, we modified the R package coloc for Bayesian inference of co-localization between the two sets of summary statistics across each gene’s cis-region. Coloc2, our modified implementation of coloc, analyzes the hierarchical model of gwas-pw, with likelihood-based estimation of dataset-wide probabilities of five hypotheses (H₀, no association; H₁, GWAS association only; H₂, eQTL association only; H₃, both but not co-localized; and H₄, both and co-localized). We then used these probabilities as priors to calculate empirical Bayesian posterior probabilities for the five hypotheses for each locus, in particular PPH₄ for co-localization.

For genes with conditional eQTL overlapping SCZ GWAS loci, summary statistics from all-but-one conditional eQTL analyses were assessed for co-localization with the GWAS signature (Figure 3). To illustrate this analytical strategy, we show eQTL results for the iron responsive element binding protein 2 gene IREB2 (MIM: 147582, chr15:78729773–78793798) as an example (Figure 4). Forward stepwise selection analysis identified two independent cis-eQTL for IREB2. In order to generate summary statistics for each eQTL in isolation, we conducted two all-but-one conditional analyses, in each analysis conditioning on all but a focal independent eQTL (for IREB2 this entailed conditioning on only one eQTL per conditional analysis, but involved conditioning on up to six eQTL per gene across all genes considered in the SCZ co-localization analysis). We then tested for co-localization between the GWAS and all of the eQTL summary statistics resulting from the above conditioning analysis using coloc2 (Table S12). In the case of IREB2, the conditional eQTL (rs7171869) was implicated as co-localized with the GWAS signal at this locus with a posterior probability for co-localization (PPH₄) of 0.94. A qualitative examination of the IREB2 locus supported the coloc2 results: the correlation between the GWAS p values and conditional eQTL p values was higher than that between the GWAS and primary eQTL p values (Figure 4A). In addition, the GWAS signature for the locus more closely resembled the
conditional eQTL signature than either the non-conditional eQTL signature or the primary eQTL signature (Figure 4B).

We found that 40 loci contained genes with strong evidence of co-localization between eQTL and GWAS signatures, with posterior probability of $H_4$ (PPH$_4$) $\geq 0.8$ (Table 2). When restricting to genome-wide significance for the GWAS, we found co-localization in 24 of the 108 loci. Given the correlations between number of independent eQTL and expression specificity scores (Tau) across tissues, cell types, and development, we tabulated the reported genes’ Tau percentiles and expression levels, to highlight contexts in which the genes are specifically expressed (Table 2, Table S8). We acknowledge that while posterior probability PPH$_4$ $\geq 0.8$ demonstrates strong Bayesian evidence for co-localization, it is an arbitrary threshold for characterizing loci as GWAS-eQTL co-localized; we find that many loci with PPH$_4$ $\geq 0.5$ appear qualitatively consistent with co-localization.

Importantly, for 6 of the 40 co-localizing loci, a conditional rather than primary eQTL co-localized with the GWAS with compelling qualitative support (Table 2, Figure 4, Table S11, Figures S13–S17). The genes showing strong evidence for conditional eQTL co-localization include SLC35E2, PROX1-AS1 (MIM: 601546), PPM1M (MIM: 608979), SDAD1P1, STAT6 (MIM: 601512), and IREB2. Also notable are the occurrences of complex patterns of co-localization for some loci; for example, three loci showed evidence for co-localization with a primary eQTL for one gene and a conditional eQTL for another.

Comparison with Previous Co-localization Analyses
In the prior CMC study, a GWAS-eQTL co-localization analysis implemented in Sherlock and using non-conditional eQTL summary statistics reported a total of 18 co-localized loci, representing 17% of the 108 genome-wide significant loci examined. Through our all-but-one conditional co-localization analysis, we replicate the majority of their findings and detect an additional 13 instances of co-localization, bringing the total number of co-localizations when considering only the genome-wide significant (and not including the MHC) loci up to 24 (representing 22% of these 108 loci) (Table S9). These 13 comprise instances of conditional eQTL co-localization (for genes SLC35E2 and IREB2) and improved detection of primary eQTL co-localization due to isolation of independent eQTL signatures and our choice of co-localization software (coloc2). Of the six co-localized loci identified in the previous but not current analysis, three resulted from differences in study design such as GWAS locus definition and...
**Discussion**

We utilized genotype and expression data from 467 human post-mortem brain samples from the DLPFC to conduct eQTL mapping analyses, to characterize both primary and conditional eQTL. We then identified co-localization between SCZ GWAS and eQTL association signals, comprising both primary and conditional eQTL. Our principal findings include four major observations. First, we detect that conditional eQTL are widespread in the brain tissue samples we investigated. In 63% of genes with at least one eQTL, we found multiple statistically independent eQTL (representing 8,136 genes). In addition, conditional eQTL make substantial contributions to regulatory genetic variation, as there is a strong association between

---

**Figure 3. All-but-One Conditional Analysis to Isolate Independent eQTL Signatures**

(A) Hypothetical GWAS signature (top, green) at a given locus and an overlapping hypothetical eQTL signature (bottom, purple), which comprises two independent eQTL.

(B) 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 conditional (secondary) eQTL and GWAS signatures.

---

**eQTL overlap criteria, and two were suggestive in the current analysis (0.65 < PPH4 < 0.8).** The one remaining discrepant locus (chr8:143302933–143403527) was found to co-localize with TSNARE1 eQTL previously (Sherlock p = 8.24 × 10⁻⁷) but not here (coloc2 primary eQTL PPH4 = 0.074, PPH3 = 0.93). A qualitative comparison of the eQTL and GWAS data (Figure S18) did not appear to support co-localization; while the strongest GWAS association and the strongest eQTL are in close physical proximity, the LD between the two index SNPs is low (r²~0.2–0.4). Additionally, our attempts to disentangle independent eQTL signal via conditional analysis do not reveal the GWAS index SNP to be in high LD with any of the conditionally independent eQTL peaks.

We also compared our conditional co-localization results with those from non-conditional eQTL analysis, using coloc2 and the same SCZ GWAS loci (Table S10). Conditional and non-conditional coloc2 results were highly concordant, with slightly higher PPH4s resulting from the same WABFs due to a higher prior probability of co-localization estimated in the non-conditional coloc2 analysis. Thirty-five loci were co-localized in both analyses; five loci that were co-localized in the non-conditional analysis only were highly suggestive in the conditional analysis (0.65 < PPH4 < 0.8), and the five loci that were co-localized only in the conditional coloc2 analysis involved conditional eQTL, emphasizing the utility of the conditional analysis. This conditional eQTL co-localization represents a substantial proportion (~15%) of all instances of co-localization, and furthermore could reflect context-specific differential expression that has the potential to implicate cell types, tissue types, and developmental stages that are relevant to disease etiology.

---

We also compared our conditional co-localization results with those from non-conditional eQTL analysis, using coloc2 and the same SCZ GWAS loci (Table S10). Conditional and non-conditional coloc2 results were highly concordant, with slightly higher PPH4s resulting from the same WABFs due to a higher prior probability of co-localization estimated in the non-conditional coloc2 analysis. Thirty-five loci were co-localized in both analyses; five loci that were co-localized in the non-conditional analysis only were highly suggestive in the conditional analysis (0.65 < PPH4 < 0.8), and the five loci that were co-localized only in the conditional coloc2 analysis involved conditional eQTL, emphasizing the utility of the conditional analysis. This conditional eQTL co-localization represents a substantial proportion (~15%) of all instances of co-localization, and furthermore could reflect context-specific differential expression that has the potential to implicate cell types, tissue types, and developmental stages that are relevant to disease etiology.

---

We also compared our conditional co-localization results with those from non-conditional eQTL analysis, using coloc2 and the same SCZ GWAS loci (Table S10). Conditional and non-conditional coloc2 results were highly concordant, with slightly higher PPH4s resulting from the same WABFs due to a higher prior probability of co-localization estimated in the non-conditional coloc2 analysis. Thirty-five loci were co-localized in both analyses; five loci that were co-localized in the non-conditional analysis only were highly suggestive in the conditional analysis (0.65 < PPH4 < 0.8), and the five loci that were co-localized only in the conditional coloc2 analysis involved conditional eQTL, emphasizing the utility of the conditional analysis. This conditional eQTL co-localization represents a substantial proportion (~15%) of all instances of co-localization, and furthermore could reflect context-specific differential expression that has the potential to implicate cell types, tissue types, and developmental stages that are relevant to disease etiology.
This demonstrates that genetic variation affecting RNA abundance is incompletely characterized by focusing on only one primary eQTL per gene, which is the case currently for most eQTL studies.

Second, we find the genomics of conditional eQTL and their genes are consistent with complex, context-specific regulation of gene expression, which may be conferred through overlap with distal regulatory elements. Genes with more independent eQTL tend to be larger and span multiple recombination hotspot intervals, and tend to be less constrained at the protein level. While these associations may reflect in part greater power to detect independent eQTL that are not in linkage disequilibrium and explain more phenotypic variance, they are also consistent with more complex regulation and greater potential for regulatory genetic variation. Context-specific genetic regulation of expression could manifest as conditional eQTL signal in the analysis of expression from a heterogeneous source. For example, eQTL in naive and stimulated (LPS, IFN) monocytes may occur as either primary or conditional eQTL in our CMC data, due to related microglial cells.
<table>
<thead>
<tr>
<th>Chr</th>
<th>GWAS Locus Start</th>
<th>GWAS Locus End</th>
<th>GWAS Lead SNP</th>
<th>GWAS p Value</th>
<th>eSNP</th>
<th>eSNP p Value</th>
<th>Primary/Conditional</th>
<th>PPM</th>
<th>Gene</th>
<th>Relevant Tissue/Cell Type/Developmental Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2372401</td>
<td>2402501</td>
<td>rs4688845</td>
<td>4.03E−09</td>
<td>rs12037821</td>
<td>4.9E−04</td>
<td>conditional</td>
<td>0.87</td>
<td>SLC35E2</td>
<td>–/–/early mid-prenatal</td>
</tr>
<tr>
<td>1</td>
<td>8355697</td>
<td>8638984</td>
<td>rs301797</td>
<td>2.03E−09</td>
<td>rs138050288</td>
<td>1.8E−04</td>
<td>primary</td>
<td>0.95</td>
<td>RERE</td>
<td>–/–/–</td>
</tr>
<tr>
<td>1</td>
<td>30412551</td>
<td>30443951</td>
<td>rs1498232</td>
<td>1.28E−09</td>
<td>rs2015244</td>
<td>1.8E−08</td>
<td>primary</td>
<td>0.99</td>
<td>PTIPRU</td>
<td>–/neurons/early mid-prenatal</td>
</tr>
<tr>
<td>1</td>
<td>163582923</td>
<td>163766623</td>
<td>rs7521492</td>
<td>5.64E−07</td>
<td>rs10799961</td>
<td>3.18E−11</td>
<td>primary</td>
<td>0.91</td>
<td>PRX1</td>
<td>–/–/early prenatal</td>
</tr>
<tr>
<td>1</td>
<td>205015255</td>
<td>205189455</td>
<td>rs16937</td>
<td>8.69E−07</td>
<td>rs12724651</td>
<td>7.31E−07</td>
<td>primary</td>
<td>0.89</td>
<td>TMEM81</td>
<td>–/neurons/–</td>
</tr>
<tr>
<td>1</td>
<td>214137889</td>
<td>214163689</td>
<td>rs7529073</td>
<td>9.69E−07</td>
<td>rs12031350</td>
<td>8.15E−06</td>
<td>conditional</td>
<td>0.87</td>
<td>RBP5</td>
<td>–/–/–</td>
</tr>
<tr>
<td>2</td>
<td>73194203</td>
<td>73900439</td>
<td>rs56145559</td>
<td>8.42E−08</td>
<td>rs11679809</td>
<td>1.85E−34</td>
<td>primary</td>
<td>0.86</td>
<td>ALMS1P</td>
<td>testis/–/–</td>
</tr>
<tr>
<td>2</td>
<td>110262036</td>
<td>110398236</td>
<td>rs9330316</td>
<td>7.69E−08</td>
<td>rs892464</td>
<td>2.35E−26</td>
<td>primary</td>
<td>0.92</td>
<td>SEPT10</td>
<td>–/–/late prenatal</td>
</tr>
<tr>
<td>2</td>
<td>198148577</td>
<td>198835577</td>
<td>rs6434928</td>
<td>1.48E−11</td>
<td>rs12621129</td>
<td>6.06E−12</td>
<td>primary</td>
<td>0.94</td>
<td>SF3B1</td>
<td>–/–/–</td>
</tr>
<tr>
<td>2</td>
<td>200715237</td>
<td>201247789</td>
<td>rs281768</td>
<td>1.78E−14</td>
<td>rs35220450</td>
<td>3.46E−14</td>
<td>primary</td>
<td>0.95</td>
<td>FTCDNL1, AC073043.2</td>
<td>–/–/adult</td>
</tr>
<tr>
<td>2</td>
<td>208371631</td>
<td>208531731</td>
<td>rs2709410</td>
<td>8.77E−04</td>
<td>rs186546506</td>
<td>8.77E−04</td>
<td>conditional</td>
<td>0.83</td>
<td>LINC01792, AC007163.3</td>
<td>putamen (basal ganglia)/ –/adult</td>
</tr>
<tr>
<td>2</td>
<td>220033801</td>
<td>220071601</td>
<td>rs6707588</td>
<td>9.51E−07</td>
<td>rs13404754</td>
<td>1.08E−09</td>
<td>primary</td>
<td>0.92</td>
<td>CREB1</td>
<td>–/–/early prenatal</td>
</tr>
<tr>
<td>3</td>
<td>36843183</td>
<td>36945783</td>
<td>rs7596809</td>
<td>3.39E−12</td>
<td>rs9834970</td>
<td>1.88E−05</td>
<td>primary</td>
<td>0.94</td>
<td>DCLK3</td>
<td>nerve - tibial/neurons/infant</td>
</tr>
<tr>
<td>3</td>
<td>52281078</td>
<td>53539269</td>
<td>rs2535627</td>
<td>3.96E−11</td>
<td>rs6801235</td>
<td>2.81E−08</td>
<td>conditional</td>
<td>0.86</td>
<td>PPM1M</td>
<td>–/neurons/late prenatal</td>
</tr>
<tr>
<td>3</td>
<td>63792650</td>
<td>64004050</td>
<td>rs832187</td>
<td>2.58E−08</td>
<td>rs113386200</td>
<td>1.95E−12</td>
<td>primary</td>
<td>0.98</td>
<td>THOC7</td>
<td>–/–/–</td>
</tr>
<tr>
<td>3</td>
<td>135807405</td>
<td>136615405</td>
<td>rs7432375</td>
<td>5.27E−11</td>
<td>rs10935184</td>
<td>7.71E−25</td>
<td>primary</td>
<td>0.93</td>
<td>PCCB</td>
<td>–/–/–</td>
</tr>
<tr>
<td>4</td>
<td>170357552</td>
<td>170646052</td>
<td>rs10520163</td>
<td>1.02E−09</td>
<td>rs7438</td>
<td>1.02E−09</td>
<td>primary</td>
<td>0.97</td>
<td>CLCN3</td>
<td>–/–/–</td>
</tr>
<tr>
<td>5</td>
<td>45291475</td>
<td>46404116</td>
<td>rs1501357</td>
<td>1.24E−08</td>
<td>rs9292918</td>
<td>4.45E−05</td>
<td>primary</td>
<td>0.94</td>
<td>BRCAT54, RP11-53O19.1</td>
<td>–/–/adult</td>
</tr>
<tr>
<td>6</td>
<td>83779798</td>
<td>84407274</td>
<td>rs739869</td>
<td>8.57E−10</td>
<td>rs2016358</td>
<td>1.19E−09</td>
<td>primary</td>
<td>0.90</td>
<td>SNAP91</td>
<td>cerebellar hemisphere/–/–</td>
</tr>
<tr>
<td>6</td>
<td>108875527</td>
<td>109019327</td>
<td>rs9398171</td>
<td>3.37E−09</td>
<td>rs111727905</td>
<td>3.84E−06</td>
<td>primary</td>
<td>0.97</td>
<td>ZNF259P1</td>
<td>–/–/early mid-prenatal</td>
</tr>
<tr>
<td>7</td>
<td>21485312</td>
<td>21545712</td>
<td>rs73060317</td>
<td>6.60E−07</td>
<td>rs14984481</td>
<td>3.59E−05</td>
<td>primary</td>
<td>0.92</td>
<td>SP4</td>
<td>–/–/early prenatal</td>
</tr>
<tr>
<td>8</td>
<td>8088038</td>
<td>10056127</td>
<td>rs2945232</td>
<td>2.03E−08</td>
<td>rs2980441</td>
<td>7.68E−09</td>
<td>primary</td>
<td>0.82</td>
<td>FAM86B3P</td>
<td>–/–/adolescent</td>
</tr>
<tr>
<td>8</td>
<td>26181524</td>
<td>26279124</td>
<td>rs1042992</td>
<td>2.27E−07</td>
<td>rs17055186</td>
<td>3.06E−24</td>
<td>conditional</td>
<td>0.91</td>
<td>SDAD1P1</td>
<td>testis/–/adult</td>
</tr>
<tr>
<td>8</td>
<td>38020424</td>
<td>38310924</td>
<td>rs57799857</td>
<td>3.23E−07</td>
<td>rs201999999</td>
<td>1.70E−07</td>
<td>primary</td>
<td>0.88</td>
<td>WHSC1L1</td>
<td>–/–/early prenatal</td>
</tr>
<tr>
<td>8</td>
<td>144822546</td>
<td>144871746</td>
<td>rs11784536</td>
<td>1.83E−07</td>
<td>rs12541792</td>
<td>6.45E−35</td>
<td>primary</td>
<td>0.90</td>
<td>FAM83H</td>
<td>esophagus - mucosa/oligodendrocytes/adolescent</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Chr</th>
<th>GWAS Locus Start</th>
<th>GWAS Locus End</th>
<th>GWAS Lead SNP</th>
<th>GWAS p Value</th>
<th>eSNP</th>
<th>eSNP p Value</th>
<th>Primary/Conditional</th>
<th>PPH4</th>
<th>Gene</th>
<th>Relevant Tissue/Cell Type/Developmental Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>26839508</td>
<td>26909408</td>
<td>rs10967586</td>
<td>4.75E−07</td>
<td>rs12345197</td>
<td>3.90E−06</td>
<td>primary</td>
<td>0.80</td>
<td>IFT74</td>
<td>--/--</td>
</tr>
<tr>
<td>11</td>
<td>46340213</td>
<td>46751213</td>
<td>rs7951870</td>
<td>1.97E−11</td>
<td>rs16938506</td>
<td>5.08E−05</td>
<td>primary</td>
<td>0.88</td>
<td>MDK</td>
<td>--/--early mid-prenatal</td>
</tr>
<tr>
<td>12</td>
<td>57428314</td>
<td>57497814</td>
<td>rs324017</td>
<td>2.13E−07</td>
<td>rs4559</td>
<td>2.02E−05</td>
<td>conditional</td>
<td>0.91</td>
<td>STAT6</td>
<td>--/microglia/adolescent</td>
</tr>
<tr>
<td>14</td>
<td>35421614</td>
<td>35847614</td>
<td>rs77477310</td>
<td>1.52E−07</td>
<td>rs1028449</td>
<td>8.09E−04</td>
<td>primary</td>
<td>0.84</td>
<td>RP11-85K15.2</td>
<td>--/--</td>
</tr>
<tr>
<td>15</td>
<td>78803032</td>
<td>78926732</td>
<td>rs8042374</td>
<td>1.87E−12</td>
<td>rs7171869</td>
<td>1.44E−04</td>
<td>conditional</td>
<td>0.94</td>
<td>IREB2</td>
<td>--/--early prenatal</td>
</tr>
<tr>
<td>15</td>
<td>84661161</td>
<td>85153461</td>
<td>rs950169</td>
<td>7.62E−11</td>
<td>rs35677834</td>
<td>1.54E−34</td>
<td>primary</td>
<td>0.80</td>
<td>LOC101929479, RP11-561C5.3</td>
<td>ovary/--early mid-prenatal</td>
</tr>
<tr>
<td>15</td>
<td>91416560</td>
<td>91436560</td>
<td>rs4702</td>
<td>2.30E−12</td>
<td>rs4702</td>
<td>4.49E−13</td>
<td>primary</td>
<td>1.00</td>
<td>FURIN</td>
<td>--/endothelial cells/adolescent</td>
</tr>
<tr>
<td>16</td>
<td>4447751</td>
<td>4596451</td>
<td>rs6500602</td>
<td>2.79E−07</td>
<td>rs3747580</td>
<td>4.75E−16</td>
<td>primary</td>
<td>0.90</td>
<td>CORO7</td>
<td>--/--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs8046295</td>
</tr>
<tr>
<td>16</td>
<td>29924377</td>
<td>30144877</td>
<td>rs12691307</td>
<td>1.30E−10</td>
<td>rs4788203</td>
<td>1.95E−05</td>
<td>primary</td>
<td>0.88</td>
<td>TMEM219</td>
<td>--/--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs3935873</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs4787491</td>
</tr>
<tr>
<td>16</td>
<td>58669293</td>
<td>58691393</td>
<td>rs12325245</td>
<td>1.15E−08</td>
<td>rs11647976</td>
<td>4.83E−04</td>
<td>primary</td>
<td>0.94</td>
<td>CNOT1</td>
<td>--/--</td>
</tr>
<tr>
<td>17</td>
<td>17722402</td>
<td>18030202</td>
<td>rs8082590</td>
<td>6.84E−09</td>
<td>rs4072739</td>
<td>4.74E−13</td>
<td>primary</td>
<td>0.92</td>
<td>DRG2</td>
<td>--/--</td>
</tr>
<tr>
<td>19</td>
<td>11839736</td>
<td>11859736</td>
<td>rs72986630</td>
<td>4.64E−08</td>
<td>rs72986630</td>
<td>2.20E−14</td>
<td>primary</td>
<td>1.00</td>
<td>ZNF823</td>
<td>--/endothelial cells/early prenatal</td>
</tr>
<tr>
<td>19</td>
<td>19374022</td>
<td>19658022</td>
<td>rs2905426</td>
<td>6.92E−09</td>
<td>rs2965199</td>
<td>9.22E−36</td>
<td>primary</td>
<td>0.87</td>
<td>GATAD2A</td>
<td>--/--</td>
</tr>
<tr>
<td>19</td>
<td>50067499</td>
<td>50135399</td>
<td>rs56873913</td>
<td>2.19E−07</td>
<td>rs5023763</td>
<td>9.32E−05</td>
<td>primary</td>
<td>0.93</td>
<td>SNRNP70</td>
<td>--/--</td>
</tr>
<tr>
<td>22</td>
<td>41408556</td>
<td>42689414</td>
<td>rs9607782</td>
<td>6.76E−12</td>
<td>rs20044724</td>
<td>1.87E−04</td>
<td>primary</td>
<td>0.96</td>
<td>RANGAP1</td>
<td>--/--</td>
</tr>
</tbody>
</table>
being present in brain tissue homogenate. We found that 60 stimulation-specific eQTL (FDR < 0.01 in interferon or lipopolysaccharide stimulated monocytes, but FDR ≥ 0.05 in naive monocytes) were also conditional eQTL in DLPFC. Notably, rs7171787, a conditional (tertiary) eQTL in our DLPFC analysis, is a stimulation-specific monocyte eQTL for the neurodevelopmental gene CYFIP1. In our data, associations with specificity of expression across tissues, developmental periods, and cell types determined from single-cell RNA-sequencing data suggest that context specificity plays a role in the occurrence of multiple statistically independent eQTL. Cell type specificity is particularly strongly correlated with eQTL number, consistent with those cell types being present in the current tissue homogenate data. Since previous studies have shown the importance of developmental or cell-specific contributions to schizophrenia, interrogation of independent eQTL effects may elucidate developmental or tissue-specific effects obscured in whole-tissue eQTL studies.

This context specificity of expression regulation is potentially mediated through overlap of eSNPs with distal regulatory elements, such as enhancers. Conditional eQTL occur farther from transcription start sites than primary eQTL, consistent with effects on enhancers. In addition, while both primary and conditional eQTL are enriched in both active promoter and enhancer regions, 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.

Third, we have identified a number of candidate genes for which genetic variation for expression co-localizes with genetic variation for schizophrenia risk (Table 2), including cases of co-localization with conditional eQTL. Genetic co-localization is expected if gene expression causally mediates disease risk, although we recognize that co-localization could also result from pleiotropy or linkage, particularly in regions of extensive linkage disequilibrium and haplotype structure. We also note that several co-localization methods have recently been developed, and direct comparisons have found broad concordance among these methods and a high degree of specificity of positive results using colocal. However, some differences in results would likely be achieved using alternative co-localization methods.

Our analyses prioritize 27 genes within 24 genome-wide significant (GWAS p < 5 × 10^{-8}) SCZ loci and 19 genes in 17 suggestive (p < 1 × 10^{-6}) loci. In addition to a number of previously implicated SCZ risk genes, our findings include several genes not previously considered as candidates, in some cases—e.g., SLC35E2, PTPRU (MIM: 602454), LINCO1792, DCLK3, PPM1M, LOC101929479—because the genes themselves do not overlap the GWAS locus regions but their eQTL do. In examining these genes for expression specificity in GTEx tissues, brain sample cell types from single-cell RNA-seq, and in BrainSpan DLPFC developmental periods (Tables 2 and S8), we find that their expression contexts show a diversity of patterns and can provide clues to generate specific hypotheses for functional follow-up of their potential roles in SCZ. Interestingly, genes broadly expressed across cell types tend to show prenatal expression.

Fourth, we highlight the importance of examining conditional eQTL for co-localization with GWASs. In at least 6 out of 40 loci showing GWAS-eQTL co-localization, a conditional eQTL signal co-localizes with SCZ risk. This is likely to be a conservative estimate, as the smaller effect sizes of conditional eQTL results in bias against detection of conditional GWAS-eQTL co-localization. If we had considered only primary eQTL in the analyses, these instances of co-localization would not have been identified. Among our highlighted conditional eQTL-GWAS co-localized genes are IREB2, STAT6, and PROX1-AS1. IREB2 (iron regulatory element binding protein 2) is a key regulator of iron homeostasis, and has been previously implicated in neurodegenerative disorders. Mouse IREB2 homolog hp2 knockouts exhibit impairments in coordination and balance, exploration, and nociception. The immune-related transcription factor STAT6 induces interleukin 4 (IL-4)-mediated anti-apoptotic activity of T helper cells, and the locus is associated with migraine and brain glioma as well as several immune/inflammatory diseases. STAT6 also activates neuronal progenitor/stem cells and neurogenesis, making it intriguing as an immune-related SCZ candidate given recent observations about the role of complement factor 4 (C4) gene as a SCZ risk gene and prior work potentially implicating microglia. Consistent with a role in immune-mediated synaptic pruning, STAT6 expression is broadly postnatal and shows specificity for microglia (Table S8). PROX1-AS1 encodes a lncRNA that has been implicated as aberrantly expressed in several cancers, is upregulated in the cell cycle S-phase, and promotes G1/S transition in cell culture. As a potential regulator of the Prospero Homeobox 1 (PROX1) transcription factor, it could be involved in development and cell differentiation in several tissues, including oligodendrocytes and GABAergic interneurons in the brain. PROX1-AS1 expression is specific to neurons and mature oligodendrocytes and is expressed postnatally (Table S8).

In conclusion, we find that conditional eQTL are widespread and are consistent with complex and context-specific regulation. Accounting for conditional eQTL leads to new findings of GWAS-eQTL co-localization and generates specific hypotheses for the role of gene expression regulation in disease etiology. The analytical strategy presented here could be implemented as a means of identification of putatively causal genes for any phenotype in which GWAS summary statistics and expression and genotype data from the GWAS phenotype-relevant tissue are available. 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.

Supplemental Data
Supplemental Data include 18 figures and 12 tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.04.011.

Consortia
CMC leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffmann-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner, and Barbara Lipska (NIMH). Additional members of CMC: A. Arcument Ciccek, Cong Lu, Kathryn Roeder, Lu Xie (Carnegie Mellon Univ.); Konrad Talbot (Cedars-Sinai Medical Center); Scott E. Hemby (High Point Univ.); Patrick F. Sullivan (Univ. Pennsylvania); Andrew Browne, Andrew Chess, Aaron Topol, Alexander Charney, Amanda Dobryn, Ben Readhead, Bin Zhang, Dalila Pinto, David A. Bennett, David H. Kavanagh, Douglas M. Ruderfer, Eli A. Stahl, Eric E. Schadt, Gabriel E. Hoffman, Hardik R. Shah, Jun Zhu, Jessica S. Johnson, John F. Fullard, Joel T. Dudley, Kiran Girdhar, Kristen J. Brennand, Gabriel E. Hoffman, Hardik R. Shah, Jun Zhu, Jessica S. Johnson, John F. Fullard, Joel T. Dudley, Kiran Girdhar, Kristen J. Brennand, Laura G. Sloofman, Laura M. Hucksin, Menachem Fromer, Millind C. Mahajan, Panos Roussos, Schahram Akbarian, Shaun M. Purcell, Tymor Hamamsy, Towhique Raj, Vahram Haroutunian, Ying-Chih Wang, Zeynep H. Gu¨ mus¸ (Mount Sinai School of Medicine by which the gene contributes to disease risk.

Web Resources
AMP-AD Knowledge Portal, https://www.synapse.org/ampad
GTEx Portal, https://www.gtexportal.org/home/
LDetect LD blocks, https://bitbucket.org/nygcresearch/ldetect-data/overview
R statistical software, https://www.r-project.org/
SNPsnap, https://data.broadinstitute.org/mpg/snpssnap/

References


