HGG Advances

ARTICLE

Critical reasoning on the co-expression module QTL in the dorsolateral prefrontal cortex

Alanna C. Cote,[1,*] Hannah E. Young,[1] and Laura M. Huckins[2,3,*]

Summary

Expression quantitative trait locus (eQTL) analysis is a popular method of gaining insight into the function of regulatory variation. While *cis*-eQTL resources have been instrumental in linking genome-wide association study variants to gene function, complex trait heritability may be additionally mediated by other forms of gene regulation. Toward this end, novel eQTL methods leverage gene co-expression (module-QTL) to investigate joint regulation of gene modules by single genetic variants. Here we broadly define a "module-QTL" as the association of a genetic variant with a summary measure of gene co-expression. This approach aims to reduce the multiple testing burden of a *trans*-eQTL search through the consolidation of gene-based testing and provide biological context to eQTLs shared between genes. In this article we provide an in-depth examination of the co-expression module eQTL (module-QTL) through literature review, theoretical investigation, and real-data application of the module-QTL to three large prefrontal cortex genotype-RNA sequencing datasets. We find module-QTLs in our study that are disease associated and reproducible are not additionally informative beyond *cis*- or *trans*-eQTLs for module genes. Through comparison to prior studies, we highlight promises and limitations of the module-QTL across study designs and provide recommendations for further investigation of the module-QTL framework.

Introduction

In the last decade, genome-wide association studies (GWASs) have identified common genetic variants associated with risk for complex disorders, but a majority of these variants are in noncoding regions of the genome and little is known regarding the molecular mechanisms underlying these associations. Expression quantitative trait locus (eQTL) analysis, the association of a genetic variant with expression of a gene, has become a popular method of gaining insight into the function of regulatory variation.[1,2] With the advent of large postmortem tissue paired genotype-RNA sequencing (RNA-seq) datasets, genetic loci affecting gene expression in *cis* (*cis*-eQTL) have been identified for almost all protein-coding genes.[3] While *cis*-eQTL resources have been instrumental in linking GWAS variants to gene function, recent evidence suggests that an average of only 11% of disease heritability is mediated by *cis* regulation of gene expression.[4] One proposed explanation for this incomplete characterization of GWAS loci is that complex trait heritability may additionally be mediated by eQTL in *trans*. While it is thought that *trans*-eQTLs can provide insight into disease,[5] they are difficult to identify due to their small effect size and the severe multiple testing burden of a whole-genome eQTL search.[6] In an effort to improve *trans*-eQTL discovery, researchers have recently proposed novel eQTL methods that leverage gene co-expression (module-QTL). This approach aims to reduce the multiple testing burden of a *trans*-eQTL search through the consolidation of gene-based testing and provides biological context to eQTLs shared between genes.

Here we broadly define a "module-QTL" as the association of a genetic variant with a summary measure of gene co-expression. This co-expression summary measure varies by study but can include the average expression of genes within a module,[7] gene-gene correlation,[8] components derived from matrix factorization,[9–11] or the first principal component of expression of genes within a module,[9,12–15] among other measures. Module-QTL studies typically provide one of two justifications for the analysis. First, module-QTL analysis is a form of dimensionality reduction, reducing the multiple testing burden of a whole-genome eQTL search and uncovering novel *trans*-eQTL relationships. Second, regulatory molecules are not well represented in traditional eQTL studies, possibly due to their tightly regulated expression or the prior removal of biological factors and confounders from the dataset.[16,17] By incorporating novel measures of coordinated gene expression, module-QTL analysis uncovers intracellular mechanisms of genetic effects on gene co-expression and network properties not detectable through a standard eQTL search.

Among existing module-QTL studies, weighted gene co-expression network analysis (WGCNA) module detection and use of the module eigengene as the module-QTL phenotype is the most common approach for calculation of the module-QTL statistic (we review existing studies in Table S1). Previous module-QTL studies are more often performed in brain tissue, whole blood, or blood cell types, possibly due to the comprehensiveness of these gene expression resources. In most studies, significant

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [2]Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06511, USA
[3]Lead contact
*Correspondence: alanna.cote@icahn.mssm.edu (A.C.C.), laura.huckins@yale.edu (L.M.H.)
https://doi.org/10.1016/j.xhgg.2024.100311.

module-QTLs are identified for a small proportion of tested co-expression modules.[7–10,13,15,18] Module-QTLs are most often compared with standard eQTLs through eSNP overlap, and most previous studies report either no replication or partial replication testing of module-QTL findings.[10,11,13–15,18] Common follow-up analyses to a module-QTL search include module gene set enrichment,[7,10–15,18] association of module summary measures to other traits,[10,13–15,18] and comparison of module-QTLs to GWASs through overlap of lead variants or colocalization.[7,9,11,13,18]

Although module-QTL studies appear effective at face value, most studies do not include one or more standard avenues of follow-up after a novel eQTL search, including QTL replication, functional genomic annotation, disease association, and experimental validation. These analyses are essential to understand whether QTLs are valid, reproducible, and likely to recapitulate biology. The limited follow-up in most studies raises the question of to what extent the module-QTL provides novel information regarding genomic function. To our knowledge there has been no comprehensive examination of module-QTL studies, even though this analysis has grown in popularity over the past 10 years.[7,9,11,12,14] In this paper, we perform our own module-QTL search using postmortem frontal cortex paired genotype-RNA-seq datasets from three large cohorts. We find that the module-QTL framework provides limited additional insight beyond a standard eQTL search in our study. Module-QTLs that are reproducible and disease associated tend to be redundant with *cis*- or *trans*-eQTLs for a gene within the module. We argue that a reduction in multiple testing burden is not sufficient justification for the use of co-expression summary measures as valid biological variables in an eQTL search. We caution researchers in the interpretation of module-QTL associations as biological genetically regulated co-expression and highlight areas for further investigation of the module-QTL framework.

Methods

We conducted analyses in three cohorts with paired dorsolateral prefrontal cortex RNA-seq and genotypes from individuals of genotype-derived European ancestry: the CommonMind Consortium (CMC)[19] (n = 488), Religious Order and Memory and Aging Project (ROSMAP)[20] (n = 453), and CMC Human Brain Collection Core (HBCC)[19] (n = 119). Following all preprocessing steps, 20,395 genes and 6,139,876 SNPs were included in CMC module-QTL and standard eQTL analysis. A total of 17,957 genes and 6,548,566 SNPs were included in ROSMAP analysis, and 18,275 genes and 6,138,559 SNPs were included in HBCC analysis.

Genotype preprocessing

CMC and HBCC

Information regarding DNA preparation and genotyping is available from Hoffman et al.[19] Samples were imputed to the TOPMed reference panel using the Michigan Imputation Server v1.3.3 and phasing performed with Eagle v2.4. SNPs with imputation Rsq <0.3 or minor allele frequency <0.05 were removed.

ROSMAP

Whole-genome sequencing data were filtered for SNPs with minor allele frequency >0.05. Information regarding prior sequencing and processing of ROSMAP WGS is available at https://www.synapse.org/#!Synapse:syn10901595.

Computational cell-type deconvolution

Differences in cellular composition between bulk tissue samples can induce patterns of co-expression between genes with similar cell-type-specific expression.[21] These co-expression signals can obscure intra-cell-type co-expression and can be correlated with common genetic variation.[21,22] Because we are interested in studying the ability of the module-QTL statistic to identify intracellular effects of variants on co-expression, we chose to correct for cell-type composition effects on gene expression by (1) estimating cell-type proportions from bulk RNA-seq using a computational deconvolution method, and (2) adjusting the gene expression data including the cell-type proportion estimates as covariates.

We chose to estimate cell-type proportions per sample from bulk RNA-seq using CIBERTSORT,[23] due to the good performance of CIBERSORT in recent benchmarking studies of computational cell-type deconvolution methods.[24–26] Bulk RNA-seq from each cohort and a PsychENCODE resource of single-cell expression reference in units of TPM[27] were provided as input, subset for 406 brain cell-type marker genes shared between the single-cell and bulk RNA-seq resources. This single-cell dataset was merged and uniformly processed from PsychENCODE-generated data and data from three previous studies: Darmanis et al.[28] and Lake et al.[29,30] The final resource included expression of eight excitatory and eight inhibitory neuronal cell types, two developmental cell types, one adult neuronal cell type, astrocytes, endothelial cells, microglia, oligodendrocytes, and oligodendrocyte progenitor cells (OPCs). We filtered the data to remove cell markers for developmental cell types, since all three bulk RNA-seq datasets in this study include only samples from adults. We performed the deconvolution using the CIBERSORTx interactive website, specifying 500 permutations for significance analysis, no quantile normalization, and run mode = relative (imputing cell fractions).

RNA-seq preprocessing

CMC

Information regarding RNA preparation and sequencing is available from Hoffman et al.[19] Preprocessing of RNA-seq was as follows: (1) TMM normalization; (2) removal of genes with <0.5 CPM in <30% of samples; (3) winsorization of gene counts, setting values that deviate >3 standard deviations from mean count to a 3 standard deviation limit; (4) sample outlier removal through visual inspection of the PC biplot; and (5) sample outlier removal through interarray correlation, removing samples with whole transcriptome correlation less than 3 standard deviations below the mean. Finally, using variancePartition[31] we calculated the contribution of technical and biological sample variables (including estimated cell-type proportions) to the expression variation of each gene. Effects of technical and demographic factors on expression were thoroughly investigated in the flagship CommonMind eQTL paper.[32] In line with Fromer et al., we included diagnosis, sex, age of death, institution, PMI, RIN, RIN^2, clustered library batch, and five genotype-derived ancestry PCs as

covariates. Additionally, cell-type proportions that explained ≥1% of variation in gene expression across ≥10% of genes were included as covariates: Ex1, Ex2, Ex3, In1, In6, astrocytes, endothelial cells, neurons, OPC, and oligodendrocytes. We adjusted for covariates in each gene expression dataset using voom[33] and obtained gene expression residuals.

ROSMAP

RNA-seq preprocessing was as follows: (1) TMM normalization; (2) removal of genes with <0.5 CPM in <30% of samples; (3) winsorization of gene counts, setting values that deviate >3 standard deviations from mean count to a 3 standard deviation limit; (4) sample outlier removal through visual inspection of the PC biplot; and (5) sample outlier removal through interarray correlation, removing samples with whole transcriptome correlation less than 3 standard deviations below the mean. Gene expression was adjusted for sex, age of death, diagnosis, PMI, RIN, library batch, four genotype-derived ancestry PCs and cell-type proportions that explained ≥1% of variation in gene expression across ≥10% of genes: Ex1, Ex3, Ex4, Ex5, In1, In2, In6, In8, neurons, astrocytes, endothelial cells, microglia, and oligodendrocytes.

HBCC

Information regarding RNA preparation and sequencing is available from Hoffman et al.[19] RNA-seq preprocessing was as follows: (1) TMM normalization; (2) removal of genes with <0.5 CPM in <30% of samples; (3) winsorization of gene counts, setting values that deviate >3 standard deviations from mean count to a 3 standard deviation limit; (4) sample outlier removal through visual inspection of the PC biplot; and (5) sample outlier removal through interarray correlation, removing samples with whole transcriptome correlation less than 3 standard deviations below the mean. Gene expression was adjusted for diagnosis, sex, age of death, PMI, RIN, RIN^2, library batch, five genotype-derived ancestry PCs, and cell-type proportions that explained ≥1% of variation in gene expression across ≥10% of genes: Ex1, Ex2, Ex3, Ex5, In6, neurons, astrocytes, endothelial cells, microglia, OPCs, and oligodendrocytes.

We note that it is also common practice to adjust RNA-seq data for hidden confounds before eQTL analysis.[34,35] Hidden confounds are factors derived empirically from the gene expression data and represent both technical and biological patterns of correlated expression. In this study we adjust only for known covariates, to preserve patterns of biological co-expression and avoid overcorrecting the expression dataset.[32,36]

Module detection

Co-expression module detection was performed using a clustering approach through Multiscale Embedded Gene Co-expression Network Analysis (MEGENA).[37] We applied all default settings, including the construction of an unsigned, undirected network and a minimum module size of 10 genes.

Module characteristics

Below we define measures used to characterize a module in this study. All statistics are calculated per module . In line with Horvath and Dong, we define the co-expression similarity s_{ij} as the absolute value of the Pearson correlation coefficient between two gene expression profiles, x_i and x_j (Equation 1).

$$s_{ij} = \left|cor\left(x_i, x_j\right)\right| \quad \text{(Equation 1)}$$

For weighted networks, it is recommended to transform the similarity matrix to some power β in order to emphasize strong correlations and down weight weak correlations, and this is a standard practice for networks constructed using the popular co-expression method WGCNA.[38] For this study we define $\beta = 1$, and our final score of connection strength (adjacency, a_{ij}, Equation 2) is equal to the co-expression similarity.

$$a_{ij} = s_{ij}^{\beta} = \left|cor\left(x_i, x_j\right)\right| \quad \text{(Equation 2)}$$

Intramodular density

Module density (Equation 3) measures whether a module is tight or cohesive, giving information about whether or not all genes within a module show high connectedness.[39] It is a similar score to mean module connectivity, but less influenced by differences in module size.

$$Connectivity\ of\ Gene\ i = \Sigma_{j \neq i}\alpha_{ij}$$

$$Module\ Density = \frac{\Sigma_i \Sigma_{j \neq i}\alpha_{ij}}{n(n-1)} \text{ where } n = number\ of\ module\ genes \quad \text{(Equation 3)}$$

Module-PC (module eigengene)

The first principal component of module co-expression (E) is a linear projection of the expression of genes within a co-expression module, sometimes also referred to as the "module eigengene."[39] In this study, each module-PC was calculated by singular value decomposition of the module gene expression matrix using the prcomp() function in R with scale = TRUE, meaning each gene's expression was scaled to unit variance before PCA calculation.

cor.kME

The strength of membership of a gene within a module can be quantified as the correlation between the expression of a gene and a module eigengene (*kME*, Equation 4).[40] For a given module, correlation between *kME* values in a reference and test dataset (*cor.kME*, Equation 5) can be used to understand how well the structure of a co-expression module is preserved between datasets.[40]

$$kME_i = cor(x_i, E) \quad \text{(Equation 4)}$$

$$cor.kME = cor_{i \in M_q}\left(kME_i^{|ref|}, kME_i^{|test|}\right) \quad \text{(Equation 5)}$$

Module-QTL power analysis

We performed a module-QTL power analysis using simple linear regression implemented in the powerEQTL.SLR function of the *powerEQTL* R package v0.3.4.[41] The average standard deviation of module-PC values from the CMC primary dataset was included as the standard deviation of the outcome in the power calculation. With an alpha = 0.05, sample size = 500, number of tests = 6,139,877 SNPs × 736 modules, regression slope = 5, and standard deviation = 4.85, the estimated power is greater than 0.70 for module-QTL detection for a causal eSNP with MAF >0.05 (Figure S1). We note reported effect sizes of previous module-QTL studies vary considerably in scale,[13–15] therefore in this power analysis we chose to test regression slope values that seemed reasonable given the distribution of PC values and effect sizes we observe in our own results. We do not have a confident understanding of the expected true module-QTL effect sizes in our study, and our power calculation should be interpreted cautiously.

Module-QTL search

In this analysis we define a module-QTL as the association of a SNP with the first principal component of gene expression within a co-expression module (E) (Equation 6). SNPs genome-wide were tested for association with each module-PC using the MatrixEQTL v2.3 R package,[42] specifying a standard additive linear model (*useModel* = *modelLINEAR*) with five genotype-derived ancestry PCs included as covariates.

$$module - QTL : E \sim SNP\ dosage + Genotype\ PC[1 - 5] \quad \text{(Equation 6)}$$

Module preservation

The preservation of CMC modules in ROSMAP and HBCC cohorts was tested using the modulePreservation function in the WGCNA R package v1.71,[38] specifying 100 permutations. Only modules with at least 50% of genes shared between the two datasets were tested for preservation. We calculated two module preservation statistics: the Z summary score and the median rank score. The Z summary score is a composite statistic based on the Z scores of four density and three connectivity-based preservation statistics. Permutation testing, where the module membership of each gene is randomly assigned, is used to calculate a Z statistic for each individual preservation score. We considered a Z summary score >10 as strong evidence of preservation and a Z summary score >2 and <10 as moderate evidence of preservation.[40] The medianRank score is a composite score based on the ranks of modules on each observed density or connectivity statistic. The medianRank score is less affected by module size than the Z summary score and is therefore a useful measure when interested in comparing preservation of differently sized modules. Further details of the module preservation statistics can be found in Langfelder et al.[40]

Module-QTL replication

We tested for replication of all CMC module-QTL meeting the following criteria: (1) the SNP is present in the replication dataset, (2) at least 50% of genes in the CMC module were present in the replication dataset, and (3) at least 10 genes could be used for calculation of the module-PC in the replication dataset. We calculated the module-PC in each replication cohort using CMC-defined module membership. For each replication calculation, only independent eQTLs were included by selecting the most significant eSNP per module-QTL, to prevent overrepresentation of any single locus in replication statistics based on patterns of linkage disequilibrium. We note that because we constructed unsigned co-expression networks, modules contain genes that are either negatively or positively correlated with each other, and negatively or positively correlated with the module-PC. To ensure consistent direction of effect between module-QTLs compared in the primary and secondary dataset, we compared the direction of correlation between the module-PC and module genes in both datasets. If fewer than 50% of genes shared a direction of correlation between the primary and secondary dataset, we flipped the sign of the Z scores for module-QTLs in the secondary dataset before calculation of replication statistics.

Standard eQTL search

We performed standard eQTL searches (*cis* + *trans*) using the same base genetic and gene expression datasets as for the module-QTL search. SNPs genome-wide were tested for association with expression of each gene using the MatrixEQTL v2.3 R package,[42] specifying a standard additive linear model (*useModel* = *modelLINEAR*) with five genotype-derived ancestry PCs included as covariates.

Module-QTL-GWAS colocalization

We tested for colocalization between module-QTLs and a range of neuropsychiatric GWASs: cortical thickness,[43] cortical surface area,[43] attention-deficit hyperactivity disorder (ADHD),[44] autism spectrum disorder (ASD),[45] Alzheimer disease,[46] bipolar disorder,[47] anorexia nervosa,[48] major depressive disorder (MDD),[49] posttraumatic stress disorder (PTSD), schizophrenia,[50] and neuroticism.[51] In each test of colocalization, 500 kB on either side of each lead module-QTL variant was provided as input.

Coloc

We tested for module-QTL-GWAS colocalization using coloc[52] with default parameters. A PPH4 of greater than 0.80 was considered significant evidence of colocalization.

eCAVIAR

We tested for module-QTL-GWAS colocalization using eCAVIAR v2.2,[53] specifying a maximum of three causal SNPs. Linkage disequilibrium was estimated from the CMC TOPMed-imputed dosage data of the 488 individuals used in the primary module-QTL analysis. A colocalization posterior probability (CLPP) of greater than 0.01 was considered significant evidence of colocalization.

Module-QTL-eQTL colocalization

In each test of colocalization, 500 kB on either side of each lead module-QTL variant was provided as input. Module-QTLs were tested for colocalization with *cis*- and *trans*-eQTLs for genes within the module-QTL module.

Coloc

We tested for module-QTL-eQTL colocalization using coloc[52] with default parameters. A PPH4 of greater than 0.80 was considered significant evidence of colocalization.

eCAVIAR

We tested for module-QTL-eQTL colocalization using eCAVIAR v2.2,[53] specifying a maximum of three causal SNPs. Linkage disequilibrium was estimated from the CMC TOPMed-imputed dosage data of the 488 individuals used in the primary module-QTL analysis. A colocalization posterior probability (CLPP) of greater than 0.01 was considered significant evidence of colocalization. eCAVIAR, unlike coloc, requires the locus Z scores as input. To ensure consistent direction of effect between each module-QTL and eQTL being compared, we flipped the sign of the Z scores for any eQTL where the eGene was negatively correlated with the module-PC.

Results

CMC module-QTL search results

Using the MEGENA module detection method with all default settings, we identified 736 CMC co-expression modules, including 315 parent and 421 child modules. Modules contained an average of 98.96 genes (SD = 361.62) with a total of 19,750 assigned genes. PC1 explains an average of 45% variance in module gene expression (Figure S2).

A total of 6,139,876 SNPs and 736 modules were tested in a whole-genome module-QTL search. We identified

3,135 SNP-module associations across 24 modules at a per-module Bonferroni threshold of $p < 8.14 \times 10^{e-9}$ (Table S2), corresponding to 26 independent module-QTLs when we restrict to the most significant SNP per 1 MB locus. Module assignments and module-QTL results are provided in additional files 1 and 2. The average minor allele frequency of module-QTL eSNPs is 0.201 (Median = 0.181). A total of 3,053/3,135 module-QTL associations correspond to SNP associations in the 17q21.31 locus with module 493, which we explore in greater detail in later sections.

Evaluation of the module-QTL statistical framework

We conducted a series of analyses to understand what the module-QTL statistic may represent biologically, including examination of the module-PC phenotype and comparison of module-QTLs to single eQTL findings within the CommonMind dataset.

Comparison of the module-PC to individual gene expression

We examine specifically the first principal component of a gene co-expression module (module-PC) as a module-QTL summary measure due to its popularity in prior module-QTL studies[9,12–15] and as a summary measure of gene co-expression. Also referred to as the "module eigengene," the module-PC was proposed as a representation of module gene expression by the creators of the popular co-expression network method WGCNA,[39] and correlation between the module eigengene and a trait is often used to relate modules to disease or other phenotypes.[14,39,54]

To understand the module-PC phenotype in greater detail, we first compared module-PC scores to single gene expression for the same individuals in the same dataset. We calculated pairwise correlations for 5,000 randomly selected genes in the CMC expression dataset (Figure S3A). Since only a small number of gene pairs are expected to show functional relationships, we expect this plot to be normally distributed and centered around zero. Only 2.19% of gene-gene correlations exceed an absolute Pearson r of 0.50. In comparison, we find a strong association between module-PCs and individual module genes (Figure S3B; average absolute Pearson correlation = 0.91).

Transferability of the module-PC across cohorts

Because co-expression summary measures are independently calculated across cohorts, we next investigated whether the module-PC represents a similar gene expression phenotype across datasets. We first calculated the module-PC in each replication cohort using CMC-defined module membership, only considering modules where greater than 50% of module genes are present in the replication data. Then, we measured the correlation of kME module membership scores between CMC and each replication dataset. While overall cor.kME values are high (average cor.kME: ROSMAP = 0.638, HBCC = 0.786), we also observe variation in how well the structure of a module is preserved across datasets (Figure S4A). There is no significant difference in cor.kME values between modules with and without a significant module-QTL (Figures S4B and S4D), even when we restrict to modules with evidence of preservation in the secondary dataset ($Z_{Summary} > 2$) (Figures S4C and S4E).

Characteristics of modules with module-QTLs

Modules with a significant module-QTL are smaller in size than modules without a significant module-QTL (W = 11,250, p value = 0.00613), but show no significant difference in intramodular density (t = −1.87, p value = 0.0622) (Figures 1A and 1B). Modules with a significant module-QTL have greater variance in gene expression explained by PC1 (t = −2.16, p value = 0.0313) (Figure 1C). There is no significant difference in medianRank values between modules with and without a significant module-QTL (ROSMAP: W = 6942.5, p value = 0.835, HBCC: W = 10038, p value = 0.104) (Figures 1D and 1E). There is a significant difference in $Z_{Summary}$ scores between modules with and without a module-QTL when testing module preservation in ROSMAP (W = 8750, p value = 0.022), but not when testing module preservation in HBCC (W = 9743, p value = 0.183) (Figures 1D and 1E). We note that $Z_{Summary}$ scores are often influenced by module size, with larger modules showing higher $Z_{Summary}$ values, which makes comparison of $Z_{Summary}$ scores difficult to interpret.

Comparison of module-QTLs to standard eQTLs

We next determined how many module-QTLs are also detectable as *cis*- or *trans*-eQTLs for at least one module gene (Figure 2A). When we restrict to the most significant SNP per module-QTL locus, over half of module-QTL SNPs are detectable as standard eQTLs at a per-gene Bonferroni threshold of $p < 8.14 \times 10^{-9}$, and considerably less at a genome-wide correction of $p < 3.99 \times 10^{-13}$, meaning module-QTLs capture *trans*-relationships likely not detectable as standard eQTLs when considering the multiple testing burden of a *trans*-eQTL search. Next, we conducted formal tests of module-QTL-eQTL colocalization using coloc[52] and eCAVIAR,[53] testing 1-MB region around each lead module-QTL variant and any significant *cis*- or *trans*-eQTL for genes within the module ($p < 8.14 \times 10^{-9}$ for significant eQTL). Fifteen of 26 independent module-QTLs were also a *cis*- or *trans*-eQTL for at least one module gene, and all 15 of these module-QTLs demonstrated significant colocalization (PPH4 > 0.80 and/or CLPP > 0.01) with a *cis*- or *trans*-eQTL for a gene within the module (Figure 2B). When we restrict to the four module-QTLs with evidence of replication (replication $p < 0.05$), we observe that colocalizing standard eQTLs for module genes are highly significant. p values for the association of each lead module-QTL SNP with module genes are provided in Table S3.

Investigation of the module-QTL using standard eQTL follow-up approaches

Module-QTL replication

We tested for preservation of CMC co-expression modules in ROSMAP and HBCC (Figure S5). Ninety-two modules show strong evidence of preservation in ROSMAP, while 467 modules show moderate evidence. Module preservation is slightly improved in HBCC, with 147 modules

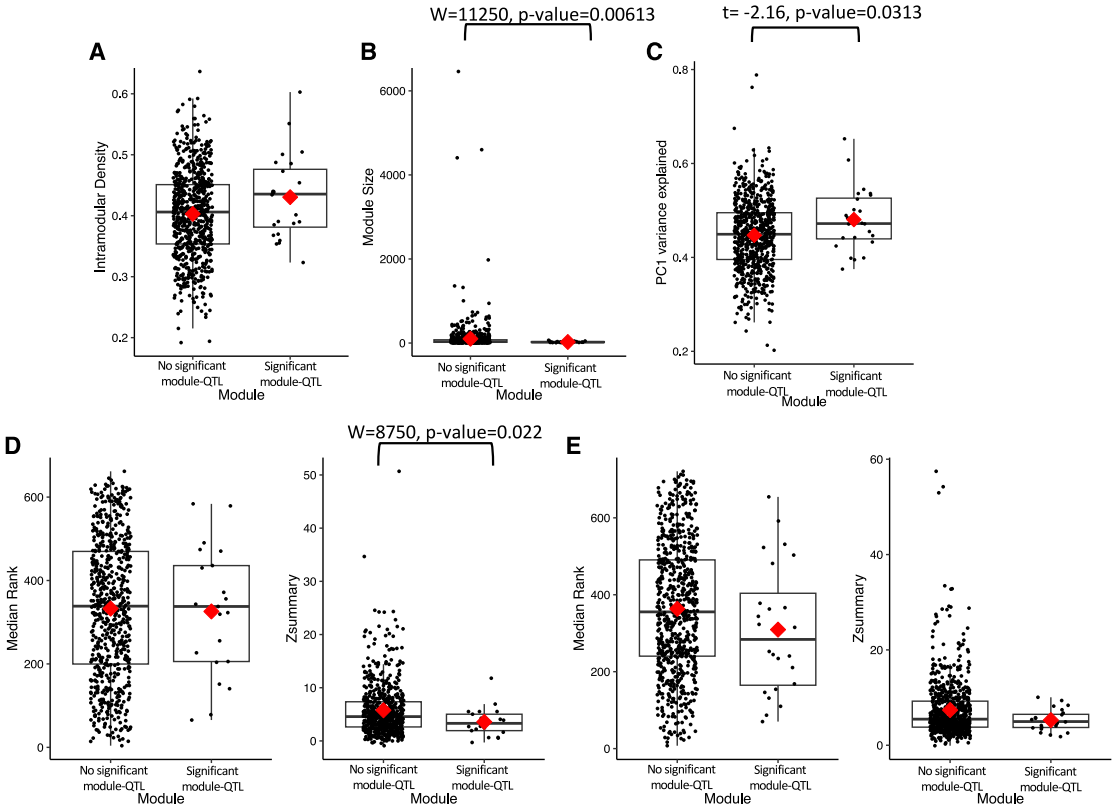

Figure 1. Characteristics of modules with and without module-QTLsFootnoteView Edit Log23
Distribution of (A) intramodular density, (B) module size, and (C) PC1 variance explained for co-expression modules derived from the CommonMind Consortium dataset with and without a significant module-QTL association. Distribution of medianRank and $Z_{Summary}$ scores of CMC module preservation in (D) ROSMAP and (E) HBCC datasets between modules with and without a significant module-QTL association. Red diamonds indicate mean value.

showing strong evidence and 542 modules showing moderate evidence of preservation. Sixty-four modules show strong evidence of preservation in both secondary datasets. Fifteen of 21 CMC module-QTL modules testable in ROSMAP and 23 of 24 CMC module-QTL modules testable in HBCC show evidence of preservation in ROSMAP and HBCC, respectively ($Z_{Summary} > 2$).

Next, we tested for replication of CMC module-QTLs in replication datasets. To interpret module-QTL replication with appropriate context we performed three sets of replication testing (Table 1; Figures S6 and S7): (1) module-QTL replication, (2) standard eQTL replication for module-QTL SNP-gene pairs, and (3) standard eQTL replication for module-QTL module genes. Twelve of 26 and 25 of 26 independent module-QTLs were available for replication testing in ROSMAP and HBCC, respectively. CMC module-QTLs replicate poorly in both ROSMAP and HBCC, with low allelic concordance (ROSMAP: 66.67%; HBCC: 52%) and no significant correlation of module-QTL Z scores between the primary and replication datasets (ROSMAP: rho = 0.231; HBCC: rho = 0.301, both $p > 0.05$). Four module-QTL loci are detectable in at least one replication dataset at a replication $p < 0.05$: two loci of chromosome 8 associated with module 437 in the ROSMAP cohort, one locus on chromosome 5 associated with module 790 in HBCC, and one locus on chromosome 17 associated with module 493 in both replication cohorts. These four module-QTLs are the only module-QTLs out of the 25 tested for replication that represent either *cis* or *cis* and *trans* SNP-gene relationships, while the 21 module-QTLs without evidence of replication represent only trans SNP-gene relationships. Module-QTLs with evidence of replication are explored in more detail in the next section. Modules for replicating module-QTLs show moderate evidence of preservation in HBCC (module 493: Z summary = 5.77, medianRank = 70.25; module 790: Z summary = 3.41, medianRank = 244.50), and low to moderate evidence of preservation in ROSMAP (module 493: Z summary = 5.02, medianRank = 65.50; module 437: Z summary = 1.93, medianRank = 319.25).

Next, we compared module-QTL replication with eQTL replication for the same SNPs and genes. We performed standard *cis*- and *trans*-eQTL searches in the same datasets, identified significant eQTLs in the primary CMC cohort at each p value threshold, and retained the top *cis*- and *trans*-eQTL per chromosome per gene to limit the impact of linkage disequilibrium patterns on replication results. eSNPs and eGenes for module-QTLs show stronger *cis*- and *trans*-eQTL replication than module-QTL replication (Figure S7; Table 1). Standard eQTLs for module-QTL eSNPs

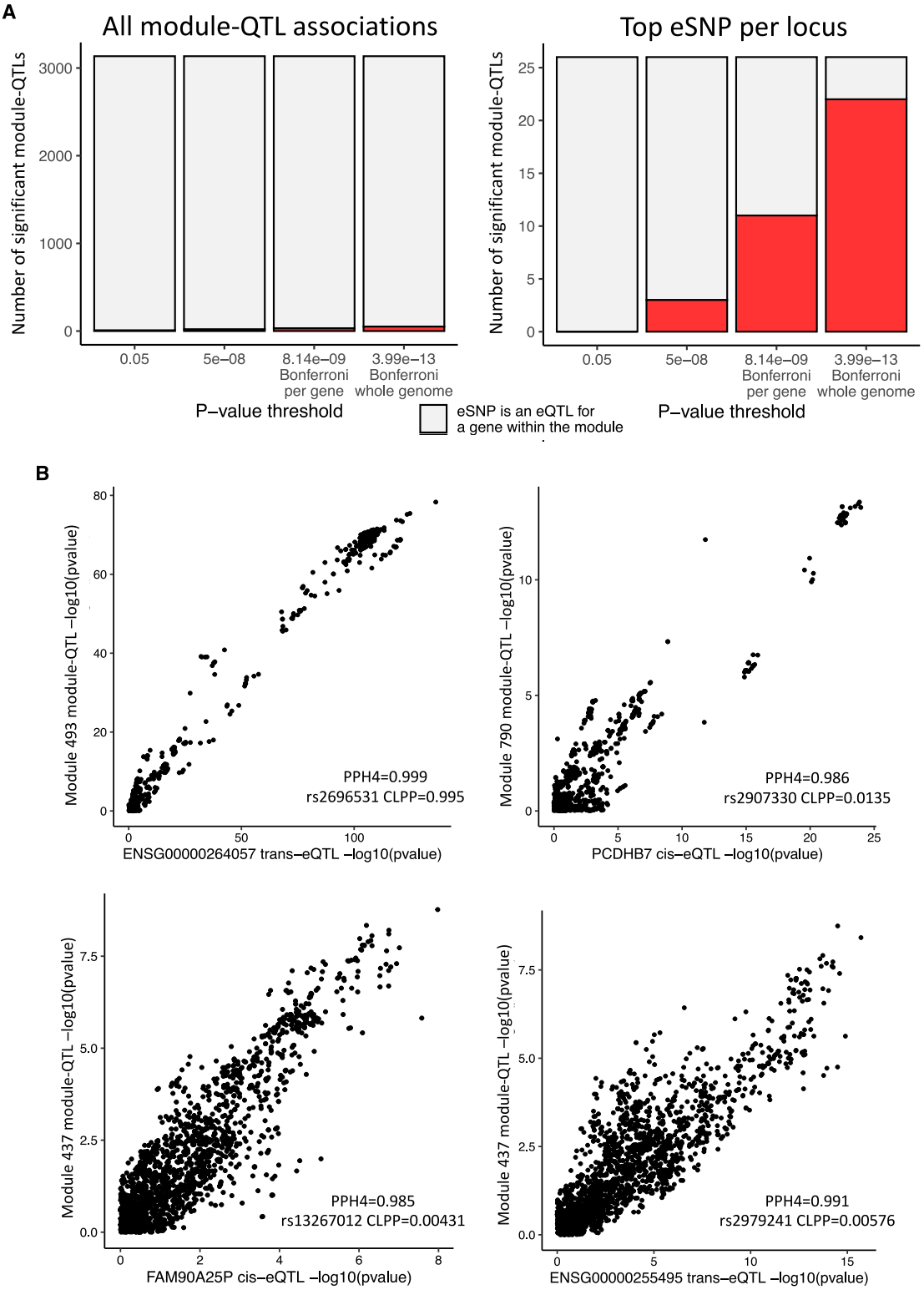

Figure 2. Comparison of module-QTLs to standard eQTLs for module genes

(A) Number of shared eSNPs between module-QTLs and *cis*- and/or *trans*-eQTLs for genes within the module at different p value thresholds for a significant eQTL of $p < 0.05$, $p < 5 \times 10^{-8}$, per-gene/per-module correction of $p < 8.14 \times 10^{-9}$ (0.05/6,139,876 SNPs), and a strict genome-wide threshold of $p < 3.99 \times 10^{-13}$ (0.05/6,139,876 SNPs × 20,395 genes).

(B) Example module-QTL-eQTL colocalization for module-QTLs with evidence of replication in an independent dataset (replication $p < 0.05$). eCAVIAR posterior probability (CLPP) is reported for the most significant eSNP in each module-QTL.

Table 1. Replication of CMC module-QTL and standard eQTL in ROSMAP and HBCC

		CMC QTLs tested in replication	eSNPs tested in replication	eModules/eGenes tested in replication		
HBCC	Module-QTL	25/26	19	23		
	eQTL (module-QTL SNP-gene pairs)	37 (15 *cis*/22 *trans*)	29	30		
	eQTL (module-QTL eGenes)	111 (58 *cis*/53 *trans*)	54	82		
ROSMAP	Module-QTL	12/26	10	11		
	eQTL (module-QTL SNP-gene pairs)	32 (14 *cis*/18 *trans*)	23	25		
	eQTL (module-QTL eGenes)	78 (55 *cis*/23 *trans*)	69	69		
		eGenes/eModules with replication ($p < 0.05$)	QTLs significant in replication (Bonf)	QTLs significant in replication ($p < 0.05$)	*Z* score correlation (Spearman rho)	Allelic concordance
HBCC	module-QTL	2/23 (8.69%)	1/25 (4%)	2/25 (8%)	0.301 ($p > 0.05$)	52%
	eQTL (module-QTL SNP-gene pairs)	16/30 (53.33%)	10/37 (27.03%)	21/37 (56.76%)	0.913 ($p < 0.05$)	75.68%
	eQTL (module-QTL eGenes)	56/82 (68.29%)	18/111 (16.22%)	61/111 (54.95%)	0.873 ($p < 0.05$)	76.58%
ROSMAP	module-QTL	2/11 (18.18%)	2/12 (16.67%)	3/12 (25%)	0.231 ($p > 0.05$)	66.67%
	eQTL (module-QTL SNP-gene pairs)	16/25 (64%)	13/32 (40.63%)	22/32 (68.75%)	0.887 ($p < 0.05$)	81.25%
	eQTL (module-QTL eGenes)	50/69 (71.01%)	31/77 (40.26%)	53/77 (68.83%)	0.859 ($p < 0.05$)	87.01%
	cis	eGenes with replication ($p < 0.05$)	QTLs significant in replication (Bonf)	QTLs significant in replication ($p < 0.05$)	*Z* score correlation (Spearman rho)	Allelic concordance
HBCC	eQTL (module-QTL SNP-gene pairs)	13/15 (86.67%)	13/15 (86.67%)	5/15 (33.33%)	0.918 ($p < 0.05$)	93.33%
	eQTL (module-QTL eGenes)	51/58 (87.93%)	14/58 (24.14%)	51/58 (87.93%)	0.911 ($p < 0.05$)	96.55%
ROSMAP	eQTL (module-QTL SNP-gene pairs)	13/14 (92.86%)	8/14 (57.14%)	13/14 (92.86%)	0.938 ($p < 0.05$)	85.71%
	eQTL (module-QTL eGenes)	47/55 (85.45%)	28/55 (51.01%)	47/55 (85.45%)	0.811 ($p < 0.05$)	94.55%
	trans	eGenes with replication ($p < 0.05$)	QTLs significant in replication (Bonf)	QTLs significant in replication ($p < 0.05$)	*Z* score correlation (Spearman rho)	Allelic concordance
HBCC	eQTL (module-QTL SNP-gene pairs)	8/22 (36.36%)	4/22 (18.18%)	8/22 (36.36%)	0.733 ($p < 0.05$)	63.63%
	eQTL (module-QTL eGenes)	10/53 (18.87%)	4/53 (7.55%)	10/53 (18.87%)	0.456 ($p < 0.05$)	54.72%
ROSMAP	eQTL (module-QTL SNP-gene pairs)	9/18 (50%)	5/18 (27.78%)	9/18 (50%)	0.917 ($p < 0.05$)	77.78%
	eQTL (module-QTL eGenes)	9/23 (39.13%)	5/23 (21.74%)	9/23 (39.13%)	0.887 ($p < 0.05$)	69.57%

and eGenes show a greater proportion of significant replication, greater correlation of effect estimates, and greater allelic concordance.

We also subsequently tested two additional significance thresholds for replication testing: a false discovery rate (FDR) 1% threshold using the Benjamini-Hochberg procedure, and a genome-wide Bonferroni correction of $p < 6.79 \times 10^{-11}$ for module-QTLs (0.05/1,000,000/736 modules) and $p < 2.45 \times 10^{-12}$ for genes (0.05/1,000,000/20,395 genes), incorporating the standard GWAS threshold and accounting for all SNPs by gene/module tests (Tables S4 and S5). At an FDR 1%, we see a slight improvement in module-QTL replication but still observe stronger replication of standard eQTLs for module-QTL eSNPs and eGenes. At a genome-wide Bonferroni correction, we identify two significant module-QTLs, one of which is significant in both replication cohorts, and observe greater or comparable module-QTL replication as standard eQTL replication.

Finally, while we adjusted for effect of diagnosis on gene expression in preprocessing of our RNA-seq data, we considered the possibility that individuals with Alzheimer disease in the ROSMAP cohort may exhibit disrupted or differing genetically regulated co-expression compared with those without neurodegenerative disease. We find there is no considerable difference in module-QTL replication when restricting the ROSMAP cohort to individuals without a diagnosis of Alzheimer disease (ROSMAP $n = 342$) (Table S6).

Module-QTL-GWAS colocalization

To investigate the potential association of module-QTL regulation with genetic risk for complex brain-related traits, we tested for colocalization of the 26 independent CMC module-QTLs with GWAS using coloc and eCAVIAR. Tested GWAS traits included the following: cortical thickness,[43] cortical surface area,[43] ADHD,[44] ASD,[45] Alzheimer disease,[46] bipolar disorder,[47] anorexia nervosa,[48] MDD,[49] PTSD, schizophrenia,[50] and neuroticism.[51] A total of 500 KB on either side of a significant module-QTL variant was provided as the input region (overlapping genomic regions collapsed per module). We identified one instance of significant colocalization with a cortical surface area GWAS locus (PPH4 = 0.863) (Figure 4).

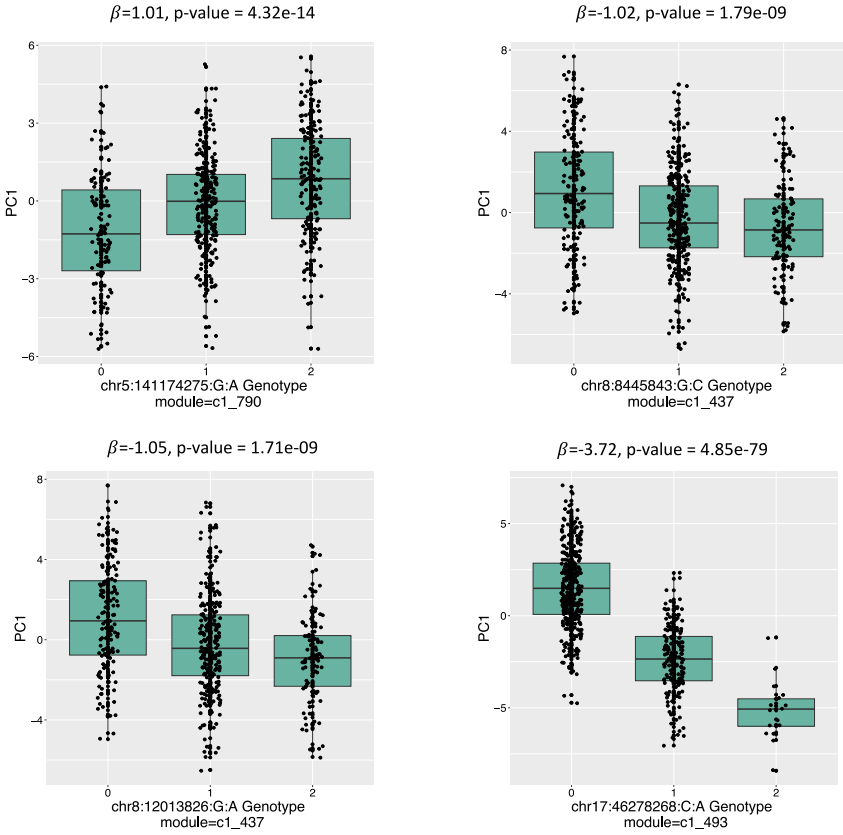

Figure 3. Scatterplots of SNP-module associations for module-QTL loci with evidence of replication in at least one cohort (replication $p < 0.05$)

Dosage was converted to hard call genotype values using PLINK default parameters.

Further examination of module-QTLs with evidence of replication

We examined in more detail the four CMC module-QTLs with evidence of replication in a secondary dataset (replication $p < 0.05$; Figure 3; supplementary text).

rs2696531–module 493

The rs2696531–module 493 module-QTL was detectable in both ROSMAP and HBCC replication datasets, and co-localizes with a cortical surface area GWAS locus (Figures 4 and S8). Module 493 contains 15 genes all located on chromosome 17, including *LRRC37A, ARL17A, ARL17B, LRRC37A2, KANSL1-AS1, KANSL1, LRR37A3, FAM215B, LRRC37A4P, RDM1,* and five pseudogenes. The first principal component of module 493 (the module-PC) explains 53.6% of variance in module gene expression and correlates strongly with many genes within the module, with all but two genes positively correlated, and 11 of 15 genes showing an absolute Pearson correlation greater than 0.70. The rs2696531–module 493 module-QTL represents 11 *cis* and four *trans* SNP-gene relationships. Within our processed CMC data, the lead module-QTL eSNP rs2696531 is associated with all but two individual module gene at a per-gene Bonferroni threshold of $p < 8.14 \times 10^{-9}$. eQTLs for four of these genes (*KANSL*, *KANSL1-AS1*, *RDM1*, and *ENSG00000264057*) colocalize with the module-QTL (coloc PPH4>0.80 or eCAVIAR CLPP >0.01).

This module-QTL lies within the 17q21.31 locus, a 1.5-MB region containing a megabase-long inversion and other structural variation[55,56] previously associated with neurodegenerative disorders like Parkinson disease,[57,58] progressive supranuclear palsy,[59,60] and Alzheimer disease.[61] This region holds two major haplotypes, an H1 haplotype with a frequency of about 0.8 in people of European ancestry, and the inverted H2 haplotype that is lower frequency and rare or absent in people of African and East Asian ancestry. Nine of 15 module genes are located in the 17q21.31 locus (*LRRC37A4P, ENSG00000131484, ENSG00000280022, KANSL1-AS1, LRRC37A, ARL17B, LRRC37A2, ARL17A, FAM215B*). There is a strong module-QTL association at the H1/H2 tag SNP rs8070723[61,62] ($\beta = -3.43$, t-statistic $= -21.05$, p value $= 3.40e{-}70$), where the H2-haplotype tagging G allele is negatively associated with the module-PC. 82.91% of significant module-QTL SNPs at this locus ($p < 8.14e{-}09$) are in high LD with the tag SNP ($r^2 > 0.9$). In addition, this module-QTL shows the same direction of effect as structural variant-eQTLs previously associated with expression of module genes at the 17q21.31 locus, including higher expression of *LRRC37A, LRRC37A2*, and *ARL17A* and reduced expression of *LRRC37A4P* with the H2 haplotype.[59] We also observe a strong positive correlation between module-QTL p values and strength of LD with an H1/H2 proxy structural variant for SNPs in the module 493 module-QTL region (Pearson $r = 0.979$, p value $<2.2e{-}308$; Figure S9). For these reasons we believe it is difficult to distinguish the potential effects of SNPs and structural variation on module expression in the 17q21.31 region, and further research is needed to understand the distinct regulatory effects of genetic variants at these loci.

Although not facilitated by the module-QTL analysis, our study identifies novel colocalization between genetic effects on gene expression in the 17q21.31 locus and a cortical surface area GWAS. This result is aligned with previous studies implicating this locus in cortical morphology. In individuals of European ancestry, the 17q21.31 inversion has previously been associated with decreased total surface area and increased cortical thickness.[63] Transcriptome-wide association studies using developing cortex eQTL models have identified associations between genes in this region (*LRRC37A, LRRC37A2, LRC37A17P,* etc.) and intracranial volume.[64]

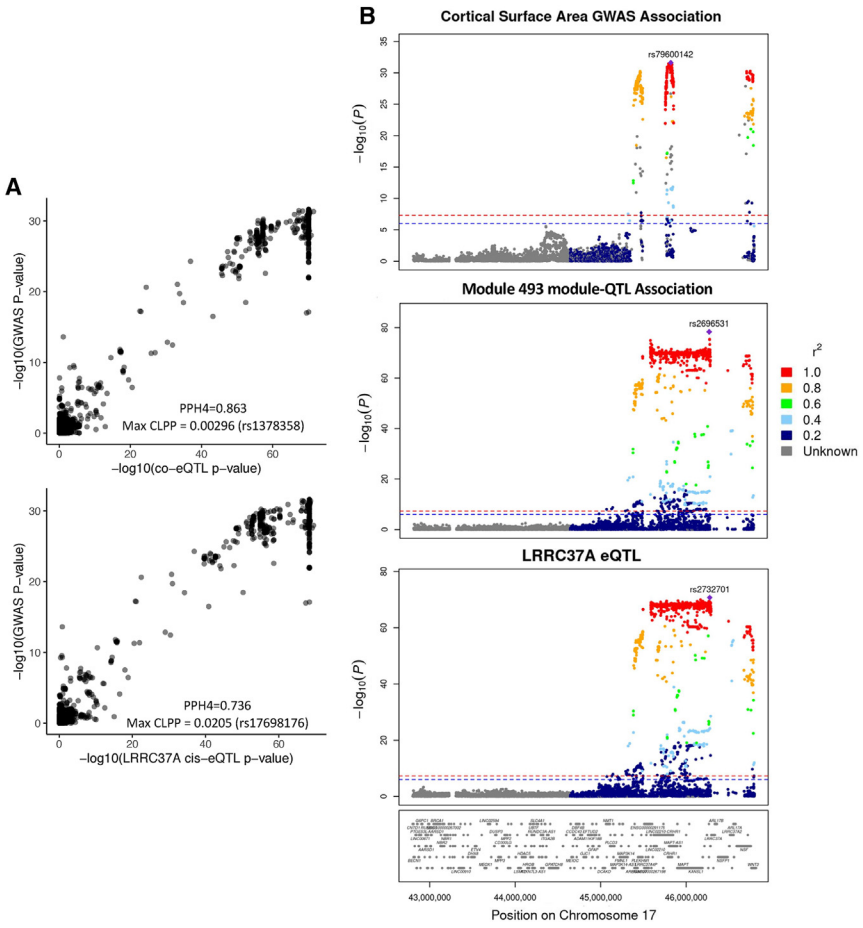

Figure 4. Colocalization of module 493 module-QTL and *LRRC37A* eQTL with a cortical surface area GWAS at the 17q21.31 locus

(A) Comparison of -log10(*p* value) for a cortical surface area GWAS region and module 493 module-QTL or standard eQTL at the 17q21.31 locus.

(B) LocusZoom-like plots of the 17q21.31 locus for cortical surface area GWAS, module-QTL, and *LRRC37A* eQTL associations.

Effects of RNA-seq misalignment on module-QTL results

The mis-mapping of RNA-seq short reads between genes with sequence similarity is an important potential source of false positive co-expression and *trans*-eQTL findings.[65] Using a resource that quantified the potential for misalignment between gene pairs, we summarized the cross-mappability between genes within each co-expression module in our study. We find that modules 493, 790, and 437 contain genes with high cross-mappability scores as compared with genes within most other modules (average mappability score: module 493 = 847.18, module 790 = 368.96, module 437 = 448.18; proportion of cross-mappable gene pairs: module 493 = 0.362, module 790 = 1, module 437 = 0.133; Figure S12), indicating that RNA-seq misalignment may impact the co-expression relationships we see for these modules, but we note not all instances of sequence similarity indicate false positive co-expression relationships, and genes with sequence similarity can also share true functional relationships. For example, the protocadherin subfamily genes in module 790 share both high sequence similarity and specific biological functions.

Discussion

The module-QTL framework has gained popularity in recent years because it addresses broader goals currently of great importance to the functional genomics community: the optimization of *trans*-eQTL discovery and the translation of disease-relevant genes to coherent biological pathways. Here, we evaluate the module-QTL methodology using three large cohorts of paired genotype and RNA-seq data in the prefrontal cortex. In our study, module-QTLs show similarity with *cis*- or *trans*-eQTLs within the same dataset but replicate poorly across datasets. Overall modules with and without a module-QTL largely do not show significant differences in measures of module preservation, indicating that there is no relationship between reproducibility of a module and the likelihood of that module having a module-QTL association. Module-QTLs with evidence of replication are more likely to include at least one *cis* SNP-gene relationship, and colocalize with at least one *cis*- or *trans*-eQTL for a gene within the module. Lead SNPs for replicating module-QTLs also show sufficiently strong association with the expression of single module genes to likely be detectable as standard eQTLs, despite the multiple testing burden of a whole-genome eQTL search (Table S3). Two of four replicating module-QTLs may partly represent structural variant effects on gene expression, but further investigation is needed to parse the regulatory effects of genetic variants at these loci. Last, module-QTLs identified in this study show no novel association with disease when tested for colocalization with GWAS of brain-related traits.

Most previous module-QTL studies that we reviewed report either no replication or incomplete replication testing of module-QTL findings.[10,11,13–15,18] To address the poor module-QTL replication observed in the present study, we conducted follow-up analyses to further understand expected eQTL replication rates and the portability of the module-QTL phenotype across cohorts. First, we compared module-QTL replication to standard eQTL replication within the same dataset, for multiple reasons. It has been shown that hidden confound adjustment can improve eQTL detection,[34,35] so it is possible that

adjustment only for known covariates reduced our ability to detect QTLs. In addition, module-QTLs represent both *cis*- and *trans*-SNP-gene regulatory relationships. *Trans*-eQTLs tend to exhibit smaller effect sizes and lower replication rates than *cis*-eQTLs,[22,32] suggesting we may expect reduced replication rates for module-QTLs representing primarily *trans*- vs. *cis*-eQTL relationships. Last, we were interested in understanding the potential impact of the module-QTL framework on replication rate, as compared with traditional eQTL analysis for the same SNP-gene relationships. By calculating standard eQTL replication for module-QTL SNP-gene pairs and module-QTL eGenes, we can better interpret module-QTL replication results in the context of the specific power of our primary dataset and the true effect sizes captured by module-QTL associations. We find overall that eSNPs and eGenes show stronger *cis*- and *trans*-eQTL replication than module-QTL replication in our secondary datasets.

The improved replication rates using a standard eQTL search suggested to us that it was likely the module-QTL statistic, and not the reduced replication rate of *trans*-eQTLs or the specific processing of our dataset, that contributed to low module-QTL replication. It is possible that, because the module-PC is independently calculated across cohorts, it does not represent the same gene expression phenotype across datasets, contributing to poor replication of module-QTL results. We observe this in our own data, where there is considerable variation in how well module structure is preserved between datasets (Figure S4). In addition, leveraging information from multiple datasets in a module-QTL meta-analysis could improve power and fine-mapping resolution, but it is not entirely clear how modules should be combined across cohorts if the module summary measure is calculated separately in each dataset. It is common practice following dimensionality reduction to investigate the relationship between the original variables and resulting factors,[66] and module-QTL studies using the "module eigengene" or any other summary measure of co-expression should always consider the magnitude and direction of association between the summary measure and individual module genes to ensure a complete understanding of their findings.

We also suggest a purely theoretical inconsistency in the module-QTL framework. Although intramodular connectivity alone is not sufficient to evaluate module validity, we would expect in theory that "tighter" modules with highly connected genes are more likely to share biological function. But the higher the intramodular connectivity, the more redundant the module-PC becomes with the expression of individual module genes. We observe this in our own data, where we find a significant correlation between intramodular density and the average or maximum absolute correlation between a module-PC and module genes (Figure S14).

We note several limitations of this study. First, results of this analysis may not extend to module-QTL studies of different designs, including different summary measures of co-expression or module detection methods, or RNA-seq in different tissues or cell types. We use the module-PC due to its popularity in previous module-QTL studies[9,12–15] and as a summary measure of gene co-expression, but acknowledge that some module-QTL studies apply different measures of co-expression, and results of our analysis may not extrapolate to module-QTL studies of different designs. Additionally, choice of module detection method has been shown in prior module-QTL and co-expression studies to have a substantial effect on the resulting co-expression modules, with some approaches identifying largely complementary and non-overlapping sets of modules even when applied to the same dataset.[9,67] Finally, although we adjusted gene expression for estimated cell-type proportions within samples, cellular composition may impact co-expression in our analysis and obscure cell-type-specific co-expression otherwise detectable in single-cell data.[21] For example, some recent publications have leveraged scRNA-seq data to identify associations between SNPs and the co-expression of gene pairs at the level of individual donors.[8,68] Researchers found that these associations are not the result of direct eQTL effects of an SNP on each gene, and validate these QTLs through tests of replication and comparison to existing CRISPR datasets. It is possible that individual network construction may improve power to detect co-expression patterns representing intracellular biology, despite the sparsity of single-cell data, or that a focus on co-expressed gene pairs, as opposed to larger groups of genes, is more suited to study using the module-QTL or similar statistic.

We also highlight an interesting example of a module-QTL that provided novel insight into gene expression regulation beyond a standard eQTL search, identified in two prior studies of monocyte gene expression.[9] This module-QTL, identified in monocytes stimulated with LPS for 24 h and located near the *SLC39A8* gene, was associated with a module of five metallothionein genes (*MT1A, MT1F, MT1G, MT1H, MT1M*). The module-QTL did not colocalize with *cis*-eQTL for *SLC39A8* in monocytes stimulated for 24 h, but did colocalize with the *cis*-eQTL 90 min after LPS stimulation, suggesting that the module-QTL framework led to identification of a transient *SLC39A8 cis*-eQTL regulating expression of this group of metallothionein genes. Genes involved in expression regulation, such as transcription factors or microRNAs, have historically been underrepresented in co-expression and eQTL studies, possibly due to their transient, low, or strictly regulated expression.[16,17] In this case, the module-QTL statistic may have facilitated the identification of context-dependent gene expression regulation without the explicit requirement of associating a variant with expression of a regulatory molecule.

In summary, while organizing disease-relevant genes by shared functional relationships can further understanding of the biological pathways underlying a complex disorder, the module-QTL framework may be an example of the tendency for researchers to revert to studying gene groups in

null or underpowered genomic studies.[69] While this study and prior module-QTL studies do identify *trans* SNP-gene relationships not detectable through standard eQTL searches,[10,11] we do not believe this is sufficient to justify the use of co-expression summary scores as measures of biological gene expression, especially in an analysis intended to yield specific mechanistic insight into gene expression regulation. A significant module-QTL association does not necessarily indicate that all or even most module genes undergo regulation by that locus, and findings should not be interpreted without also considering each SNP's association at the gene level. Through review of previous studies, theoretical investigation, and real-data application of the module-QTL to three large prefrontal cortex genotype-RNA-seq datasets, we provide an in-depth examination of the module-QTL framework that we hope can inform further design and interpretation.

Data and code availability

The code generated during this study is available at https://github.com/accote45/module-QTL. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNA-seq), R01AG17917 (MAP), R01AG30146, R01AG36836 (RNA-seq), U01AG32984 (genomic and whole exome sequencing), (whole-genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu. CommonMind Consortium datasets supporting this article are available via the CMC Knowledge Portal: https://www.synapse.org/#!Synapse:syn2759792/wiki/69613. Data were generated as part of the CommonMind Consortium supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd, and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881, AG02219, AG05138, MH06692, R01MH110921, R01MH109677, R01MH109897, and U01MH103392, and contract HHSN271201300031C through IRP NIMH. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories, and the NIMH Human Brain Collection Core. CMC Leadership: Panos Roussos, Joseph Buxbaum, Andrew Chess, Schahram Akbarian, and Vahram Haroutunian (Icahn School of Medicine at Mount Sinai); Bernie Devlin and David Lewis (University of Pittsburgh); Raquel Gur and Chang-Gyu Hahn (University of Pennsylvania); Enrico Domenici (University of Trento); Mette A. Peters and Solveig Sieberts (Sage Bionetworks); and Thomas Lehner, Stefano Marenco, and Barbara K. Lipska (NIMH). Data were generated as part of the PsychENCODE Consortium. Visit https://doi.org/10.7303/syn26365932 for a complete list of grants and PIs.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2024.100311.

Acknowledgments

We thank Drs. Alison Goate, Michael Breen, and Paul O'Reilly for their insightful feedback. This work was funded by NIMH (R01MH118278). This work was supported in part through the computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences. Research reported in this publication was also supported by the Office of Research Infrastructure of the National Institutes of Health under award number S10OD026880 and S10OD030463. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

A.C. and L.H. conceived the study and designed the experiments. A.C. performed analysis and wrote the paper with input from all co-authors. All authors read and approved the final manuscript.

Declaration of interests

The authors declare no competing interests.

Received: October 23, 2023
Accepted: May 16, 2024

Web resources

CIBERSORTx, https://cibersortx.stanford.edu/

CMC Portal, CMC Portal, https://www.synapse.org/#!Synapse:syn2759792/wiki/69613

Coloc, https://chr1swallace.github.io/coloc/

eCAVIAR, http://genetics.cs.ucla.edu/caviar/

MatrixEQTL, https://cran.r-project.org/web/packages/MatrixEQTL/index.html

MEGENA, https://cran.r-project.org/web/packages/MEGENA/index.html

powerEQTL, https://cran.r-project.org/web/packages/powerEQTL/index.html

PsychENCODE Portal, https://www.synapse.org/#!Synapse:syn26365932

ROSMAP Portal, https://adknowledgeportal.org

variancePartition, https://bioconductor.org/packages/release/bioc/html/variancePartition.html

WGCNA, WGCNA, https://cran.r-project.org/web/packages/WGCNA/index.html

References

1. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLoS Genet. *6*, e1000888. https://doi.org/10.1371/journal.pgen.1000888.
2. Young, H., Cote, A., and Huckins, L.M. (2022). Integration with systems biology approaches and -omics data to characterize risk variation. In Psychiatric Genomics (Elsevier), pp. 289–315. https://doi.org/10.1016/B978-0-12-819602-1.00017-6.
3. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
4. Yao, D.W., O'Connor, L.J., Price, A.L., and Gusev, A. (2020). Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. *52*, 626–633. https://doi.org/10.1038/s41588-020-0625-2.
5. Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243. https://doi.org/10.1038/ng.2756.
6. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell *177*, 1022–1034.e6. https://doi.org/10.1016/j.cell.2019.04.014.
7. Yang, H.-S., White, C.C., Klein, H.-U., Yu, L., Gaiteri, C., Ma, Y., Felsky, D., Mostafavi, S., Petyuk, V.A., Sperling, R.A., et al. (2020). Genetics of Gene Expression in the Aging Human Brain Reveal TDP-43 Proteinopathy Pathophysiology. Neuron *107*, 496–508.e6. https://doi.org/10.1016/j.neuron.2020.05.010.
8. van der Wijst, M.G.P., Brugge, H., de Vries, D.H., Deelen, P., Swertz, M.A., LifeLines Cohort Study; and BIOS Consortium, and Franke, L. (2018). Single-cell RNA sequencing identifies cell-type-specific cis-eQTLs and co-expression QTLs. Nat. Genet. *50*, 493–497. https://doi.org/10.1038/s41588-018-0089-9.
9. Kolberg, L., Kerimov, N., Peterson, H., and Alasoo, K. (2020). Co-expression analysis reveals interpretable gene modules controlled by trans-acting genetic variants. Elife *9*, e58705. https://doi.org/10.7554/eLife.58705.
10. Hore, V., Viñuela, A., Buil, A., Knight, J., McCarthy, M.I., Small, K., and Marchini, J. (2016). Tensor decomposition for multiple-tissue gene expression experiments. Nat. Genet. *48*, 1094–1100. https://doi.org/10.1038/ng.3624.
11. Ramdhani, S., Navarro, E., Udine, E., Efthymiou, A.G., Schilder, B.M., Parks, M., Goate, A., and Raj, T. (2020). Tensor decomposition of stimulated monocyte and macrophage gene expression profiles identifies neurodegenerative disease-specific trans-eQTLs. PLoS Genet. *16*, e1008549. https://doi.org/10.1371/journal.pgen.1008549.
12. Pergola, G., Di Carlo, P., Jaffe, A.E., Papalino, M., Chen, Q., Hyde, T.M., Kleinman, J.E., Shin, J.H., Rampino, A., Blasi, G., et al. (2019). Prefrontal Coexpression of Schizophrenia Risk Genes Is Associated With Treatment Response in Patients. Biol. Psychiatry *86*, 45–55. https://doi.org/10.1016/j.biopsych.2019.03.981.
13. Gudmundsdottir, V., Pedersen, H.K., Mazzoni, G., Allin, K.H., Artati, A., Beulens, J.W., Banasik, K., Brorsson, C., Cederberg, H., Chabanova, E., et al. (2020). Whole blood co-expression modules associate with metabolic traits and type 2 diabetes: an IMI-DIRECT study. Genome Med. *12*, 109. https://doi.org/10.1186/s13073-020-00806-6.
14. Esmaili, S., Langfelder, P., Belgard, T.G., Vitale, D., Azardaryany, M.K., Alipour Talesh, G., Ramezani-Moghadam, M., Ho, V., Dvorkin, D., Dervish, S., et al. (2021). Core liver homeostatic co-expression networks are preserved but respond to perturbations in an organism- and disease-specific manner. Cell Syst. *12*, 432–445.e7. https://doi.org/10.1016/j.cels.2021.04.004.
15. Nath, A.P., Ritchie, S.C., Byars, S.G., Fearnley, L.G., Havulinna, A.S., Joensuu, A., Kangas, A.J., Soininen, P., Wennerström, A., Milani, L., et al. (2017). An interaction map of circulating metabolites, immune gene networks, and their genetic regulation. Genome Biol. *18*, 146. https://doi.org/10.1186/s13059-017-1279-y.
16. Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R., et al. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. *24*, 14–24. https://doi.org/10.1101/gr.155192.113.
17. Saha, A., Kim, Y., Gewirtz, A.D.H., Jo, B., Gao, C., McDowell, I.C., GTEx Consortium, Engelhardt, B.E., and Battle, A. (2017). Co-expression networks reveal the tissue-specific regulation of transcription and splicing. Genome Res. *27*, 1843–1858. https://doi.org/10.1101/gr.216721.116.
18. Rotival, M., Zeller, T., Wild, P.S., Maouche, S., Szymczak, S., Schillert, A., Castagné, R., Deiseroth, A., Proust, C., Brocheton, J., et al. (2011). Integrating Genome-Wide Genetic Variations and Monocyte Expression Data Reveals Trans-Regulated Gene Modules in Humans. PLoS Genet. *7*, e1002367. https://doi.org/10.1371/journal.pgen.1002367.
19. Hoffman, G.E., Bendl, J., Voloudakis, G., Montgomery, K.S., Sloofman, L., Wang, Y.-C., Shah, H.R., Hauberg, M.E., Johnson, J.S., Girdhar, K., et al. (2019). CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. Sci. Data *6*, 180. https://doi.org/10.1038/s41597-019-0183-6.
20. De Jager, P.L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B.N., Felsky, D., Klein, H.-U., White, C.C., Peters, M.A., Lodgson, B., et al. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci. Data *5*, 180142. https://doi.org/10.1038/sdata.2018.142.
21. Farahbod, M., and Pavlidis, P. (2020). Untangling the effects of cellular composition on coexpression analysis. Genome Res. *30*, 849–859. https://doi.org/10.1101/gr.256735.119.
22. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet. *53*, 1300–1310. https://doi.org/10.1038/s41588-021-00913-z.
23. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. *37*, 773–782. https://doi.org/10.1038/s41587-019-0114-2.

24. Sutton, G.J., Poppe, D., Simmons, R.K., Walsh, K., Nawaz, U., Lister, R., Gagnon-Bartsch, J.A., and Voineagu, I. (2022). Comprehensive evaluation of deconvolution methods for human brain gene expression. Nat. Commun. *13*, 1358. https://doi.org/10.1038/s41467-022-28655-4.
25. Nadel, B.B., Oliva, M., Shou, B.L., Mitchell, K., Ma, F., Montoya, D.J., Mouton, A., Kim-Hellmuth, S., Stranger, B.E., Pellegrini, M., and Mangul, S. (2021). Systematic evaluation of transcriptomics-based deconvolution methods and references using thousands of clinical samples. Brief. Bioinform. *22*, bbab265. https://doi.org/10.1093/bib/bbab265.
26. Avila Cobos, F., Alquicira-Hernandez, J., Powell, J.E., Mestdagh, P., and De Preter, K. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat. Commun. *11*, 5650. https://doi.org/10.1038/s41467-020-19015-1.
27. Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*, eaat8464. https://doi.org/10.1126/science.aat8464.
28. Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. USA *112*, 7285–7290. https://doi.org/10.1073/pnas.1507125112.
29. Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.-L., Chen, S., et al. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science *352*, 1586–1590. https://doi.org/10.1126/science.aaf1204.
30. Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., and Zhang, K. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat. Biotechnol. *36*, 70–80. https://doi.org/10.1038/nbt.4038.
31. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinf. *17*, 483. https://doi.org/10.1186/s12859-016-1323-z.
32. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453. https://doi.org/10.1038/nn.4399.
33. Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29. https://doi.org/10.1186/gb-2014-15-2-r29.
34. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. *7*, 500–507. https://doi.org/10.1038/nprot.2011.457.
35. Leek, J.T., and Storey, J.D. (2007). Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genet. *3*, 1724–1735. https://doi.org/10.1371/journal.pgen.0030161.
36. Cote, A.C., Young, H.E., and Huckins, L.M. (2022). Comparison of confound adjustment methods in the construction of gene co-expression networks. Genome Biol. *23*, 44. https://doi.org/10.1186/s13059-022-02606-0.
37. Song, W.-M., and Zhang, B. (2015). Multiscale Embedded Gene Co-expression Network Analysis. PLoS Comput. Biol. *11*, e1004574. https://doi.org/10.1371/journal.pcbi.1004574.
38. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.
39. Horvath, S., and Dong, J. (2008). Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput. Biol. *4*, e1000117. https://doi.org/10.1371/journal.pcbi.1000117.
40. Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is My Network Module Preserved and Reproducible? PLoS Comput. Biol. *7*, e1001057. https://doi.org/10.1371/journal.pcbi.1001057.
41. Dong, X., Li, X., Chang, T.-W., Scherzer, C.R., Weiss, S.T., and Qiu, W. (2021). powerEQTL: an R package and shiny application for sample size and power calculation of bulk tissue and single-cell eQTL analysis. Bioinformatics *37*, 4269–4271. https://doi.org/10.1093/bioinformatics/btab385.
42. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.
43. Grasby, K.L., and Jahanshad, N. (2020). The Genetic Architecture of the Human Cerebral Cortex.
44. Demontis, D., Walters, G.B., Athanasiadis, G., Walters, R., Therrien, K., Nielsen, T.T., Farajzadeh, L., Voloudakis, G., Bendl, J., Zeng, B., et al. (2023). Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. Nat. Genet. *55*, 198–208. https://doi.org/10.1038/s41588-022-01285-8.
45. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444. https://doi.org/10.1038/s41588-019-0344-8.
46. Wightman, D.P., Jansen, I.E., Savage, J.E., Shadrin, A.A., Bahrami, S., Holland, D., Rongve, A., Børte, S., Winsvold, B.S., Drange, O.K., et al. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat. Genet. *53*, 1276–1282. https://doi.org/10.1038/s41588-021-00921-z.
47. Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Børte, S., Bryois, J., et al. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat. Genet. *53*, 817–829. https://doi.org/10.1038/s41588-021-00857-4.
48. Watson, H.J., Yilmaz, Z., Thornton, L.M., Hübel, C., Coleman, J.R.I., Gaspar, H.A., Bryois, J., Hinney, A., Leppä, V.M., Mattheisen, M., et al. (2019). Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat. Genet. *51*, 1207–1214. https://doi.org/10.1038/s41588-019-0439-2.
49. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. *50*, 668–681. https://doi.org/10.1038/s41588-018-0090-3.
50. Trubetskoy, V., Pardinas, A.F., Qi, T., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508. https://doi.org/10.1038/s41586-022-04434-5.

51. Nagel, M., Jansen, P.R., Stringer, S., Watanabe, K., de Leeuw, C.A., Bryois, J., Savage, J.E., Hammerschlag, A.R., Skene, N.G., Muñoz-Manchado, A.B., et al. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat. Genet. *50*, 920–927. https://doi.org/10.1038/s41588-018-0151-7.
52. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. *10*, e1004383. https://doi.org/10.1371/journal.pgen.1004383.
53. Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am. J. Hum. Genet. *99*, 1245–1260. https://doi.org/10.1016/j.ajhg.2016.10.003.
54. Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384. https://doi.org/10.1038/nature10110.
55. Boettger, L.M., Handsaker, R.E., Zody, M.C., and McCarroll, S.A. (2012). Structural haplotypes and recent evolution of the human 17q21.31 region. Nat. Genet. *44*, 881–885. https://doi.org/10.1038/ng.2334.
56. Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G., et al. (2005). A common inversion under selection in Europeans. Nat. Genet. *37*, 129–137. https://doi.org/10.1038/ng1508.
57. Skipper, L., Wilkes, K., Toft, M., Baker, M., Lincoln, S., Hulihan, M., Ross, O.A., Hutton, M., Aasly, J., and Farrer, M. (2004). Linkage Disequilibrium and Association of MAPT H1 in Parkinson Disease. Am. J. Hum. Genet. *75*, 669–677. https://doi.org/10.1086/424492.
58. Simón-Sánchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., et al. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. *41*, 1308–1312. https://doi.org/10.1038/ng.487.
59. Vialle, R.A., De Paiva Lopes, K., Bennett, D.A., Crary, J.F., and Raj, T. (2022). Integrating whole-genome sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the human brain. Nat. Neurosci. *25*, 504–514. https://doi.org/10.1038/s41593-022-01031-7.
60. PSP Genetics Study Group, Höglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M.A., Wang, L.-S., Klei, L., Rademakers, R., De Silva, R., Litvan, I., et al. (2011). Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat. Genet. *43*, 699–705. https://doi.org/10.1038/ng.859.
61. Allen, M., Kachadoorian, M., Quicksall, Z., Zou, F., Chai, H.S., Younkin, C., Crook, J.E., Pankratz, V.S., Carrasquillo, M.M., Krishnan, S., et al. (2014). Association of MAPT haplotypes with Alzheimer's disease risk and MAPT brain gene expression levels. Alzheimer's Res. Ther. *6*, 39. https://doi.org/10.1186/alzrt268.
62. Bowles, K.R., Pugh, D.A., Liu, Y., Patel, T., Renton, A.E., Bandres-Ciga, S., Gan-Or, Z., Heutink, P., Siitonen, A., Bertelsen, S., et al. (2022). 17q21.31 sub-haplotypes underlying H1-associated risk for Parkinson's disease are associated with LRRC37A/2 expression in astrocytes. Mol. Neurodegener. *17*, 48. https://doi.org/10.1186/s13024-022-00551-x.
63. Wang, H., Makowski, C., Zhang, Y., Qi, A., Kaufmann, T., Smeland, O.B., Fiecas, M., Yang, J., Visscher, P.M., and Chen, C.-H. (2023). Chromosomal inversion polymorphisms shape human brain morphology. Cell Rep. *42*, 112896. https://doi.org/10.1016/j.celrep.2023.112896.
64. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., De La Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. Cell *179*, 750–771.e22. https://doi.org/10.1016/j.cell.2019.09.021.
65. Saha, A., and Battle, A. (2018). False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. F1000Res. *7*, 1860. https://doi.org/10.12688/f1000research.17145.2.
66. Lever, J., Krzywinski, M., and Altman, N. (2017). Principal component analysis. Nat. Methods *14*, 641–642. https://doi.org/10.1038/nmeth.4346.
67. Saelens, W., Cannoodt, R., and Saeys, Y. (2018). A comprehensive evaluation of module detection methods for gene expression data. Nat. Commun. *9*, 1090. https://doi.org/10.1038/s41467-018-03424-4.
68. Li, S., Schmid, K.T., de Vries, D., Korshevniuk, M., Oelen, R., van Blokland, I., BIOS Consortiumsc-eQTLgen Consortium, Groot, H.E., Swertz, M., et al. (2022). Identification of genetic variants that impact gene co-expression relationships using large-scale single-cell data. Genetics *24*, 80. https://doi.org/10.1101/2022.04.20.488925.
69. Sedeño-Cortés, A.E., and Pavlidis, P. (2014). Pitfalls in the application of gene-set analysis to genetics studies. Trends Genet. *30*, 513–514. https://doi.org/10.1016/j.tig.2014.10.001.