nature portfolio

https://doi.org/10.1038/s41593-022-01128-z

Supplementary information

Cell-type-specific *cis*-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders

In the format provided by the authors and unedited

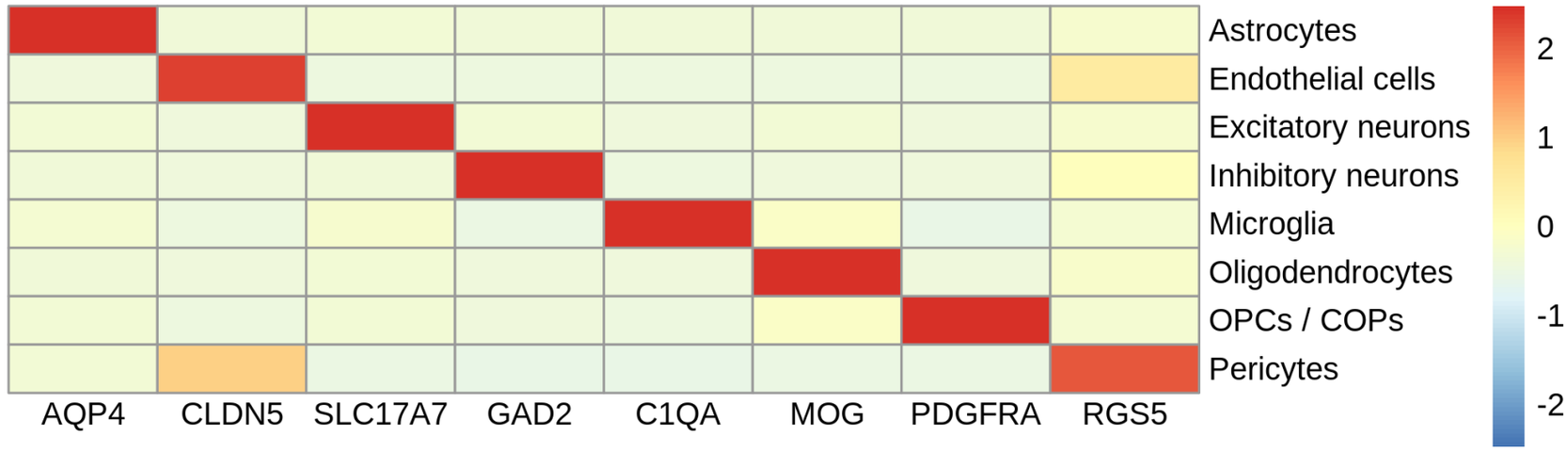


Figure S1: Mean marker expression across sample
The mean count per million (CPM) per cluster is shown for canonical markers of cell type identity (z-scaled for each cell type).

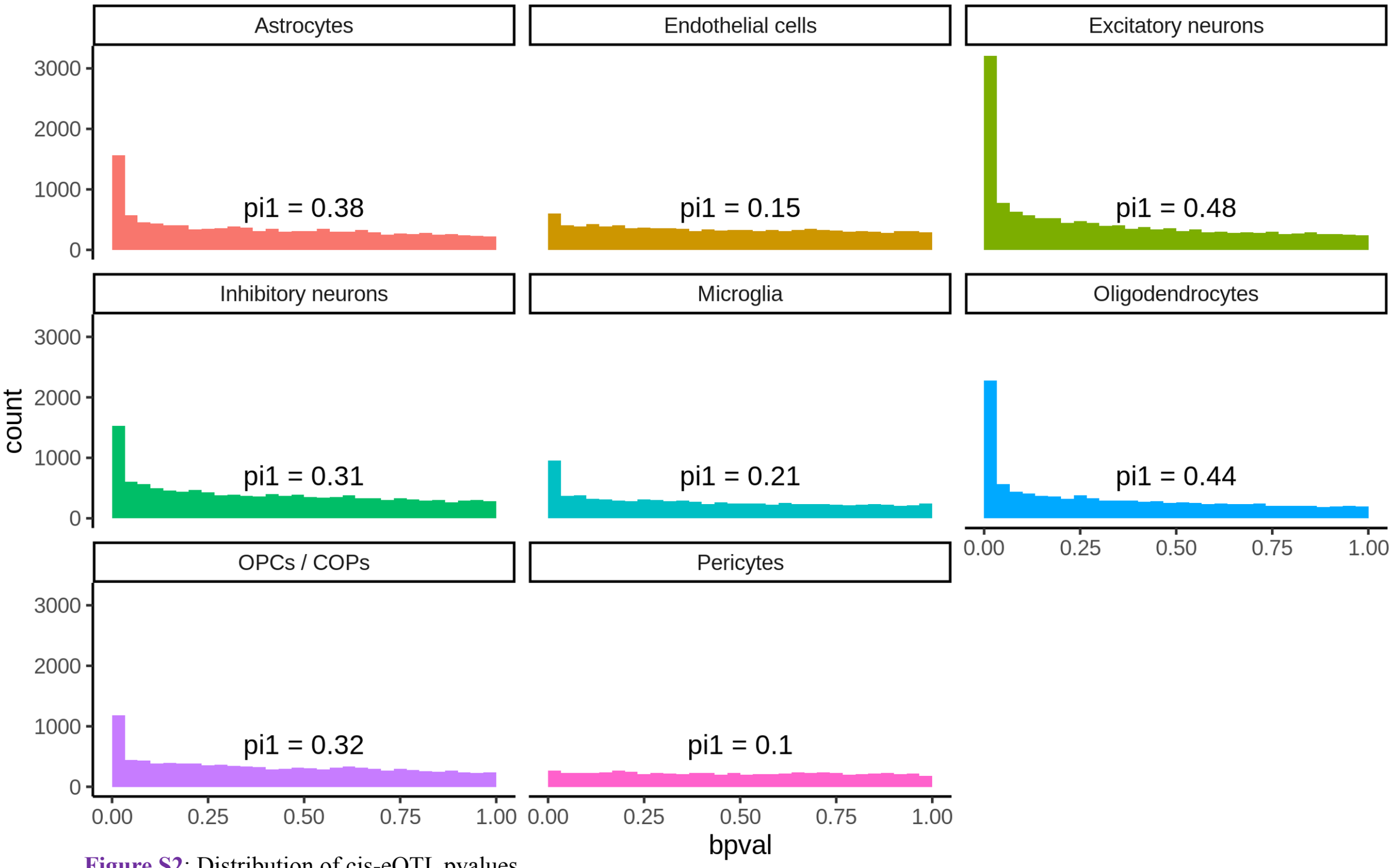


Figure S2: Distribution of cis-eQTL pvalues
Histograms of the cis-eQTL gene-level pvalues for the different cell types. The pi1 statistic estimates the proportion of true alternative hypothesis (i.e. the proportion of genes with a cis-eQTL).

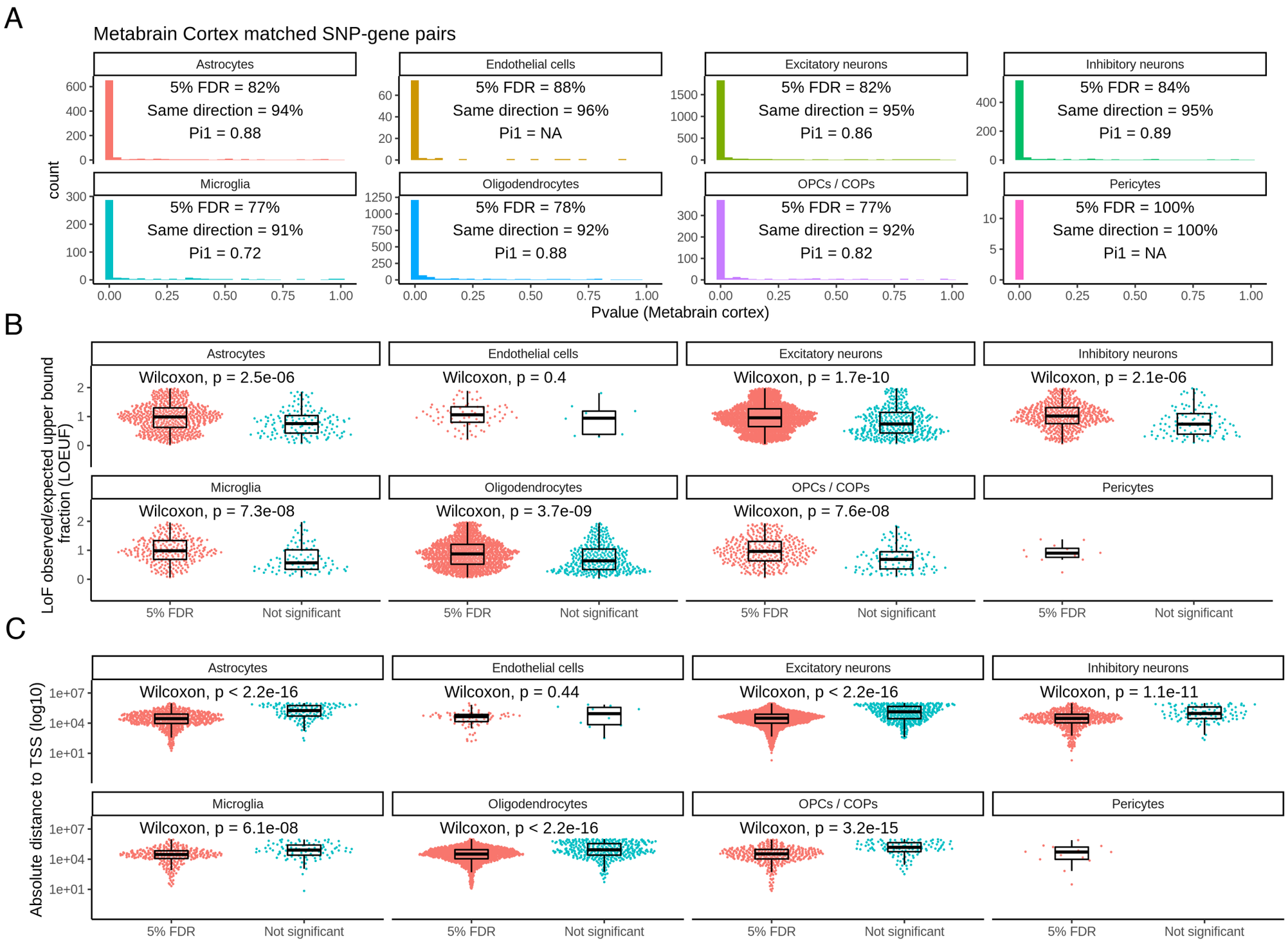


Figure S3: Replication of cis-eQTLs in a large cortical eQTL study

A) Histograms of cortical SNP-gene pvalues matching our cis-eQTLs. The '5% FDR' label indicates the proportion of pvalues that replicate at a 5% FDR, while the 'pi1' label indicates an estimate of the proportion of cis-eQTLs that would replicate with a larger sample size. **B**) LOEUF score (a measure of whether genes are constrained) for replicating cis-eQTLs (FDR <5%, n=654/76/1845/558/286/1249/367/12 genes with a LOEUF score for Astrocytes/Endothelial cells/Excitatory neurons/Inhibitory neurons/Microglia/ Oligodendrocytes/OPCs COPs/Pericytes respectively) and non-replicating cis-eQTLs (FDR >5%, n= 146/9/404/102/87/358/111 genes with a LOEUF score for Astrocytes/Endothelial cells/Excitatory neurons/Inhibitory neurons/Microglia/ Oligodendrocytes/OPCs COPs respectively). Lower scores indicate that genes are more constrained. **C**) Absolute distance to the TSS for replicating (FDR <5%, n=668/76/1882/569/292/1264/377/13 genes for Astrocytes/Endothelial cells/Excitatory neurons/Inhibitory neurons/Microglia/Oligodendrocytes/ OPCs COPs/Pericytes respectively) and non-replicating cis-eQTLs (FDR >5%, n=148/10/415/107/88/364/113 genes for Astrocytes/Endothelial cells/Excitatory neurons/Inhibitory neurons/Microglia/ Oligodendrocytes/OPCs COPs respectively). Boxplots show the median, first and third quartile, and whiskers extend up to 1.5 times the interquartile range. The two-sided pvalues were computed using a Wilcoxon test and are not corrected for multiple testing.

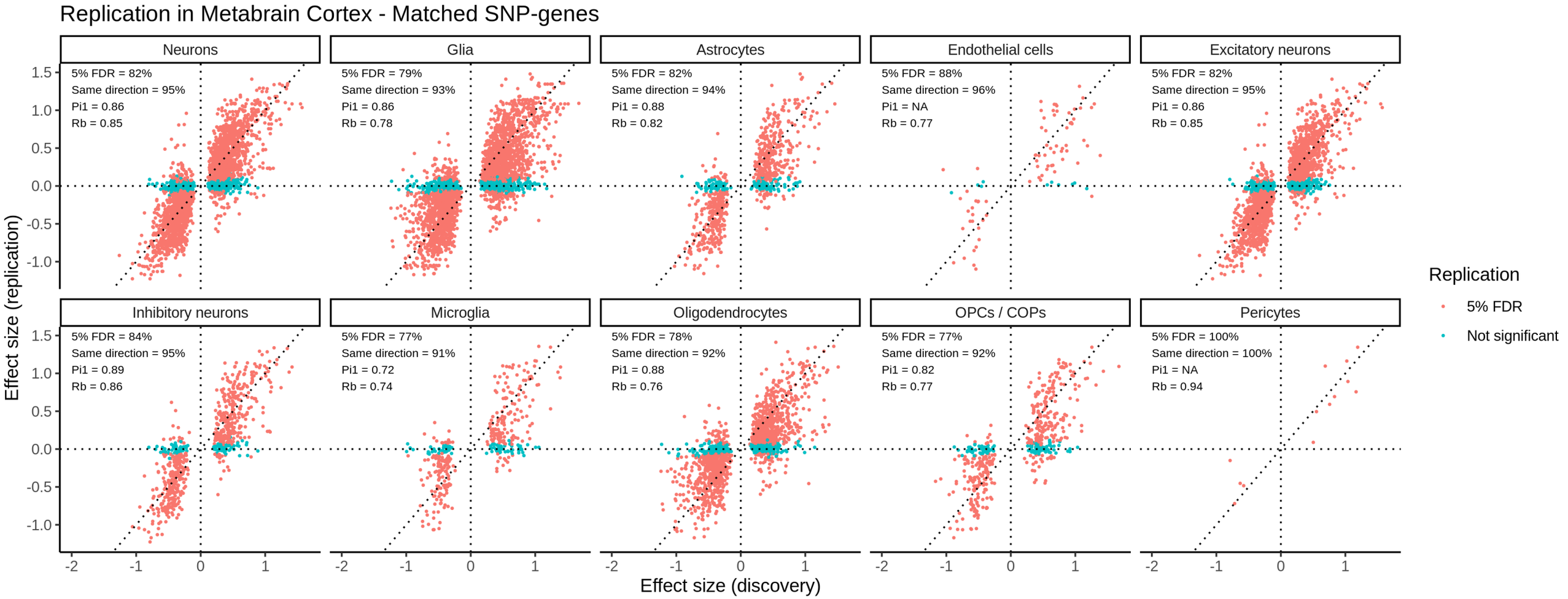


Figure S4: Effect sizes of eQTL in discovery vs replication

Effect sizes (beta) for the eQTL discovered in our study and the same SNP-gene pairs in the Metabrain study. SNP-gene pairs that replicate at 5% FDR are colored in red. The label '5%FDR' indicates the percentage of SNP-gene pairs that replicate at 5% FDR, while the 'Same direction' label indicates the percentage of the replicated SNP-gene pairs that have an effect size in the same direction. The pi1 statistic estimates the proportion of replicated signal (without accounting for the direction of effect), while the R_b statistic quantifies the correlation of the effect sizes. The 'Neurons' and 'Glia' subpanels show the aggregate of the neuronal cell types and glial cell types respectively.

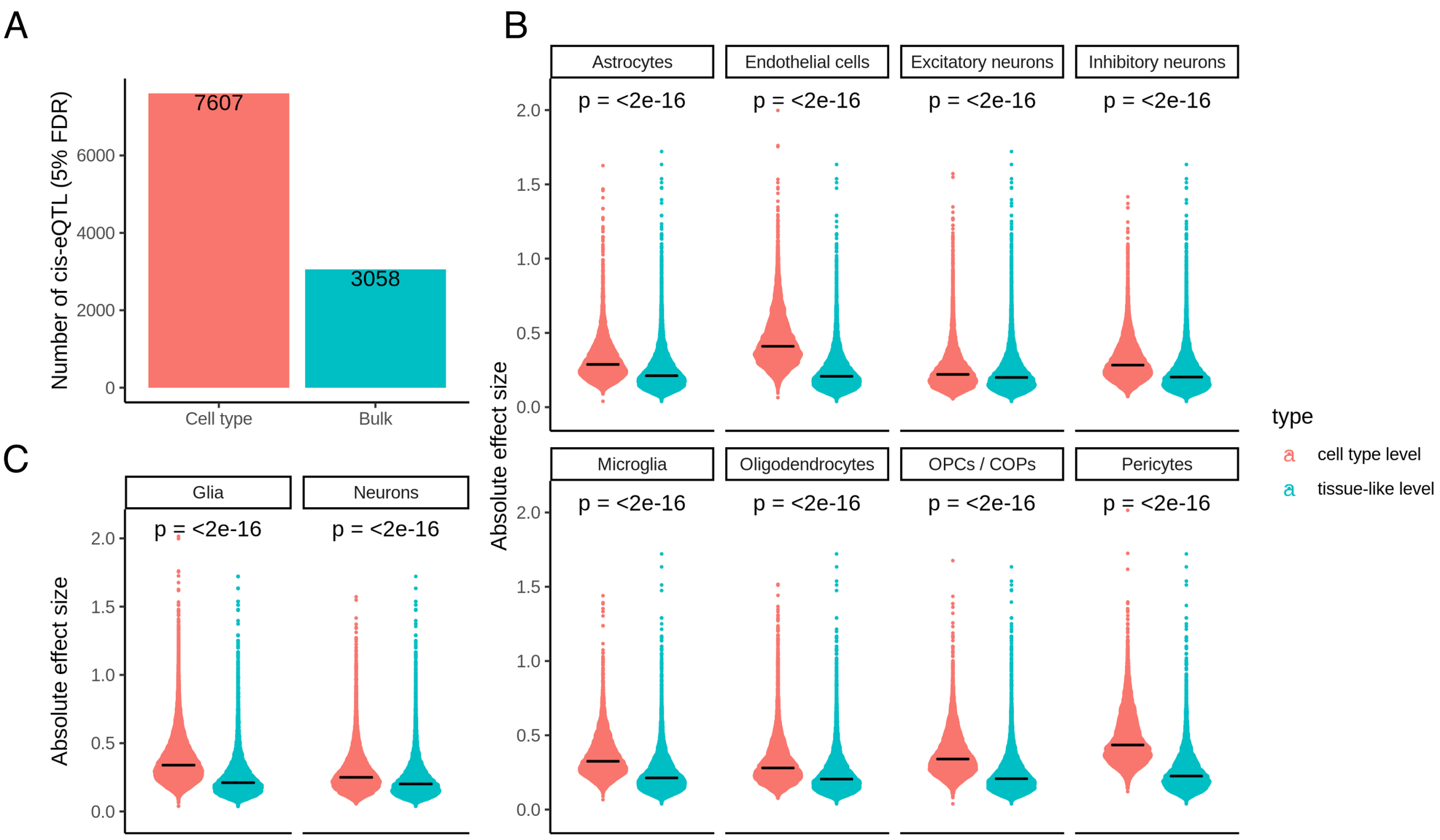


Figure S5: Cell type level vs 'tissue-like' eQTL discoveries and effect sizes
A) Number of cis-eQTLs discoveries (5% FDR) using pseudo-bulk gene expression data at a cell type level or at a 'tissue-like' level. **B)** Distribution of the effect sizes for the top cis-eQTL per gene (for all genes) using a cell type level approach or a 'tissue-like' approach for each cell type, and **C**) aggregated for all glial cell types and all neuronal cell types. Pvalues were obtained using two-sided Wilcoxon tests and are not corrected for multiple testing.

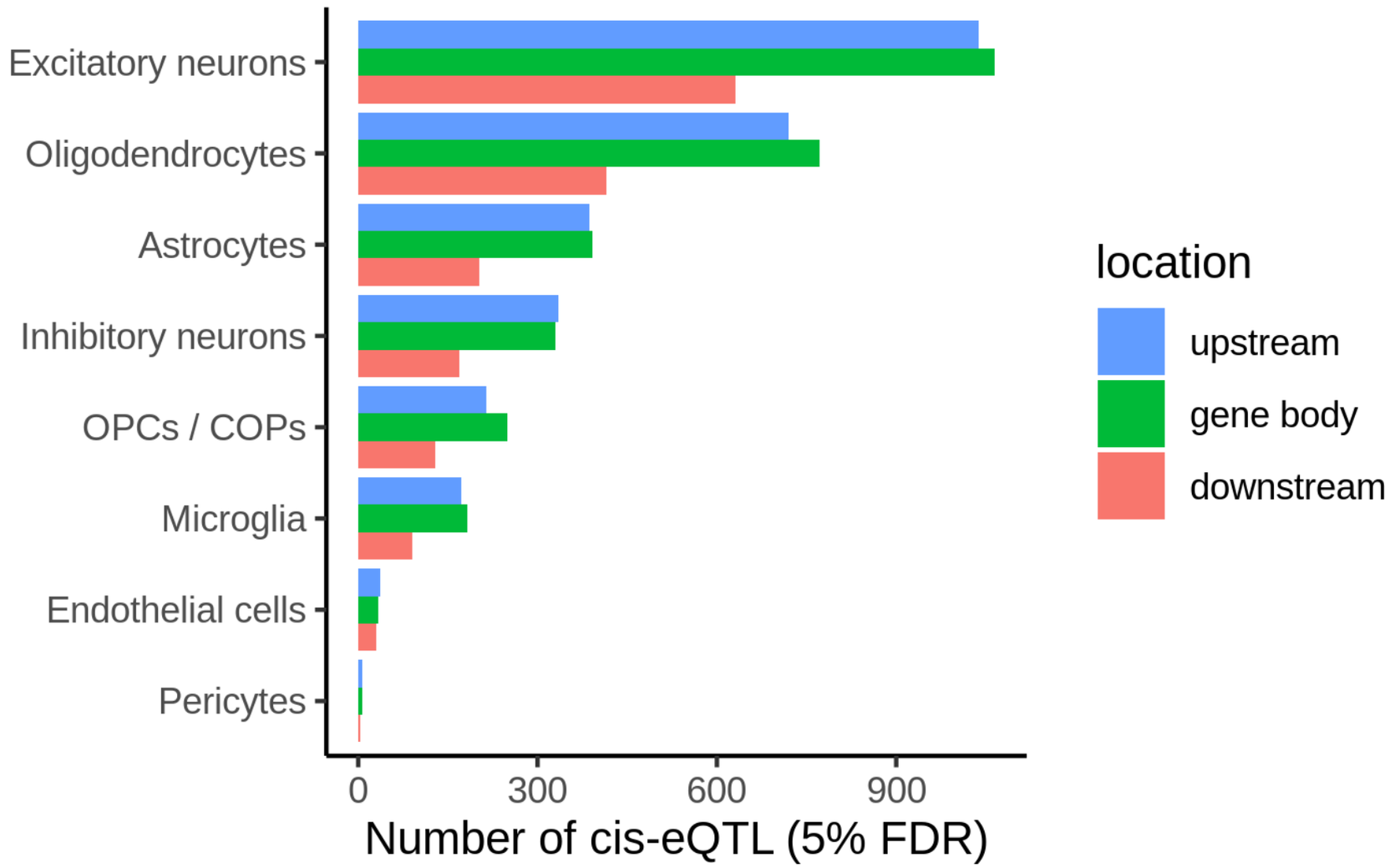


Figure S6: Location of cis-eQTLs
The number of cis-eQTLs that are located upstream, downstream or within the gene body of the affected gene are shown.

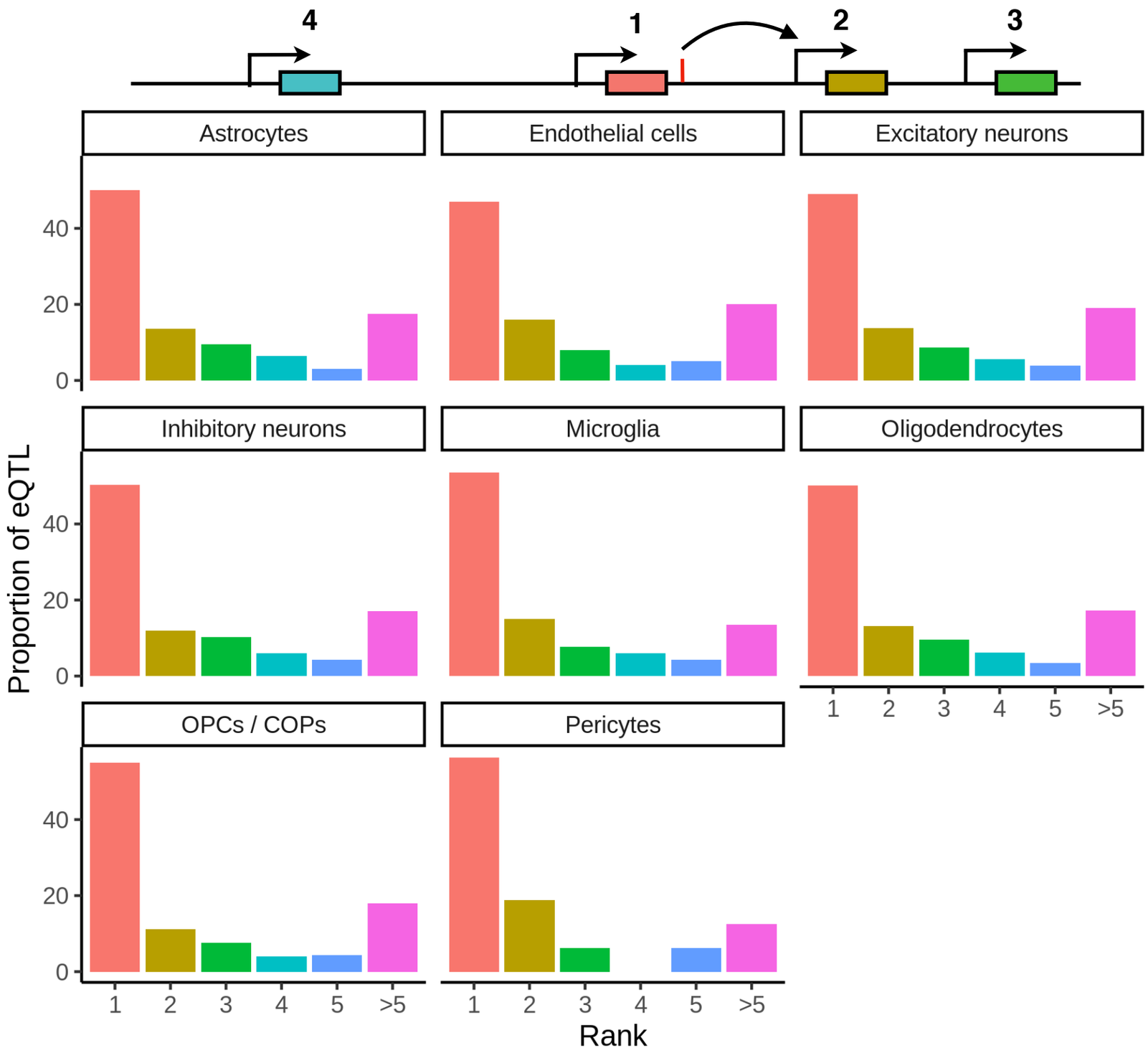


Figure S7: Rank of affected cis-eQTL genes.
The proportion of cis-eQTLs affecting the closest gene (rank =1), second closest gene (rank=2), third closest gene (rank=3), fourth closest gene (rank=4), fifth closest gene (rank=5), or further (rank>5) are shown for each cell type.

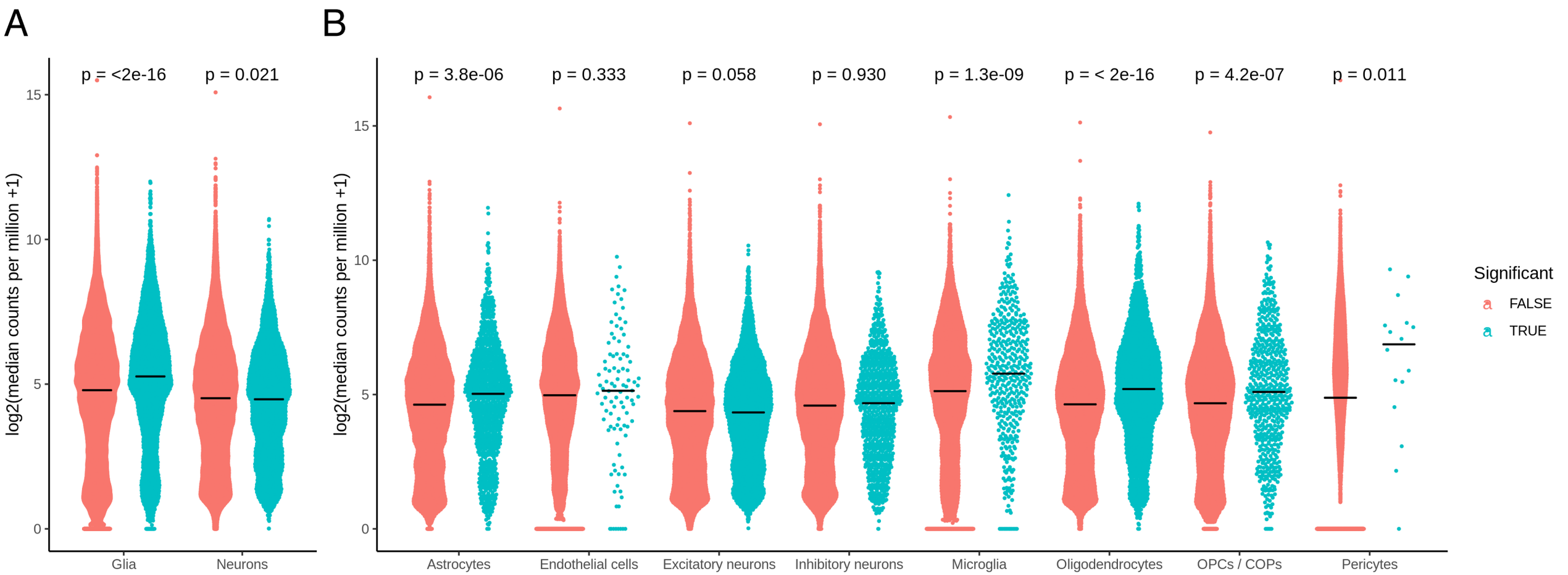


Figure S8: Expression level of cis-eQTL genes.
A) Median CPM (log2 +1) are shown for genes with a cis-eQTL and genes without a cis-eQTLs (for all glial or neuronal cell types).
B) Median CPM (log2 +1) are shown for genes with a cis-eQTL and genes without a cis-eQTLs (for all cell types). The two-sided pvalues were obtained using a Wilcoxon test and are not corrected for multiple testing.

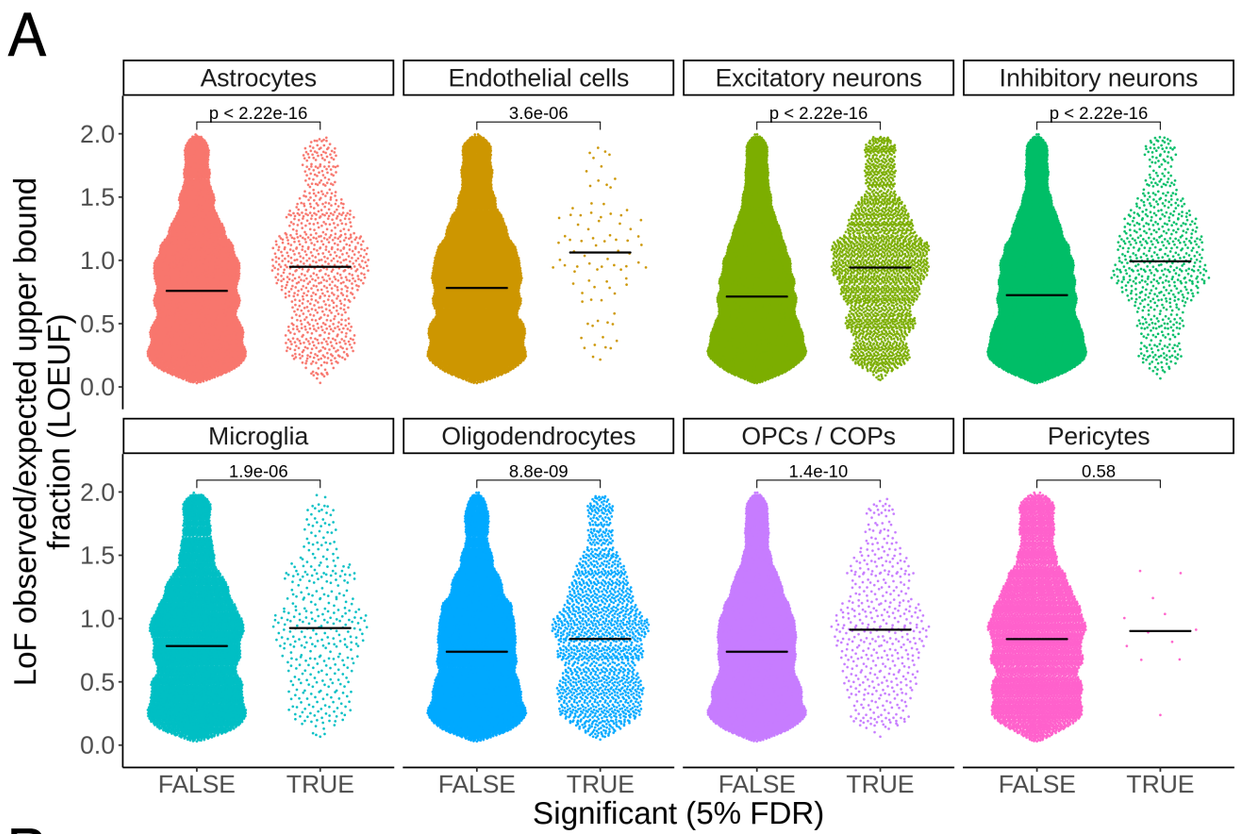
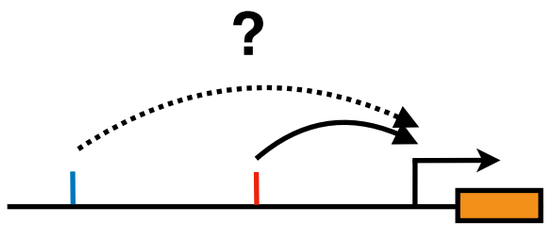


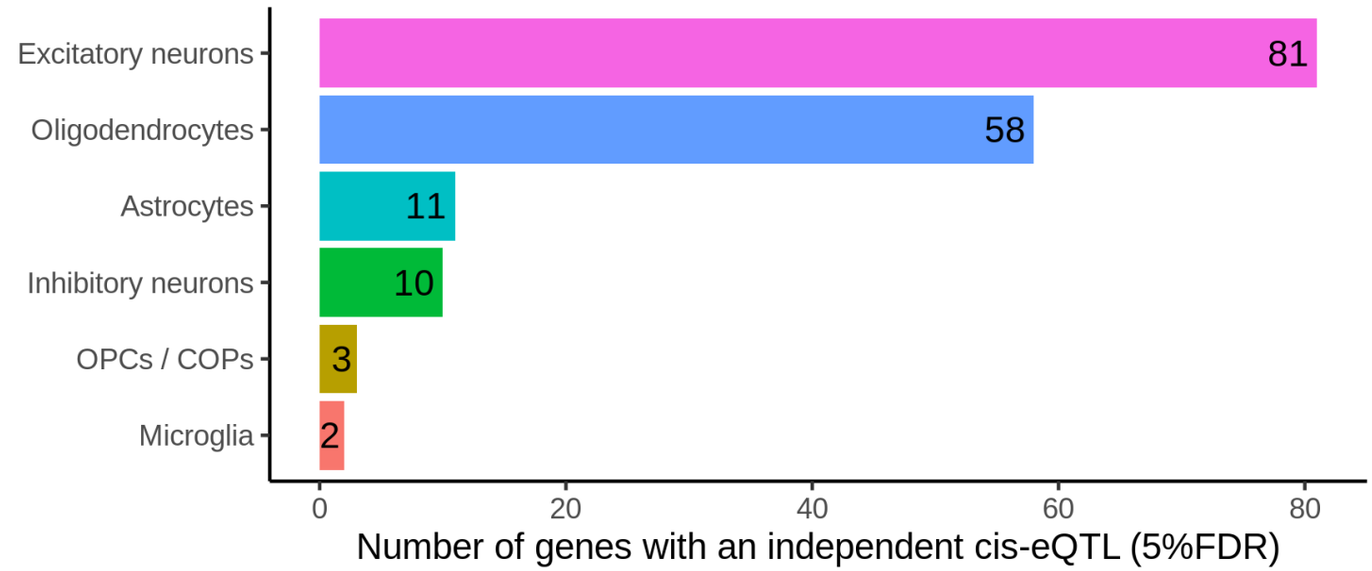
Figure S9: Constraint of genes with a cis-eQTL

A) LOEUF scores for genes with and without a cis-eQTL (5% FDR) in each cell type, and **B**) separated for genes with a cis-eQTL detected only in the cell-type level analysis, only in the 'tissue-like' level eQTL analysis and in both. The two-sided pvalues were obtained using a Wilcoxon test and are not corrected for multiple testing. More constrained genes have lower LOEUF scores.

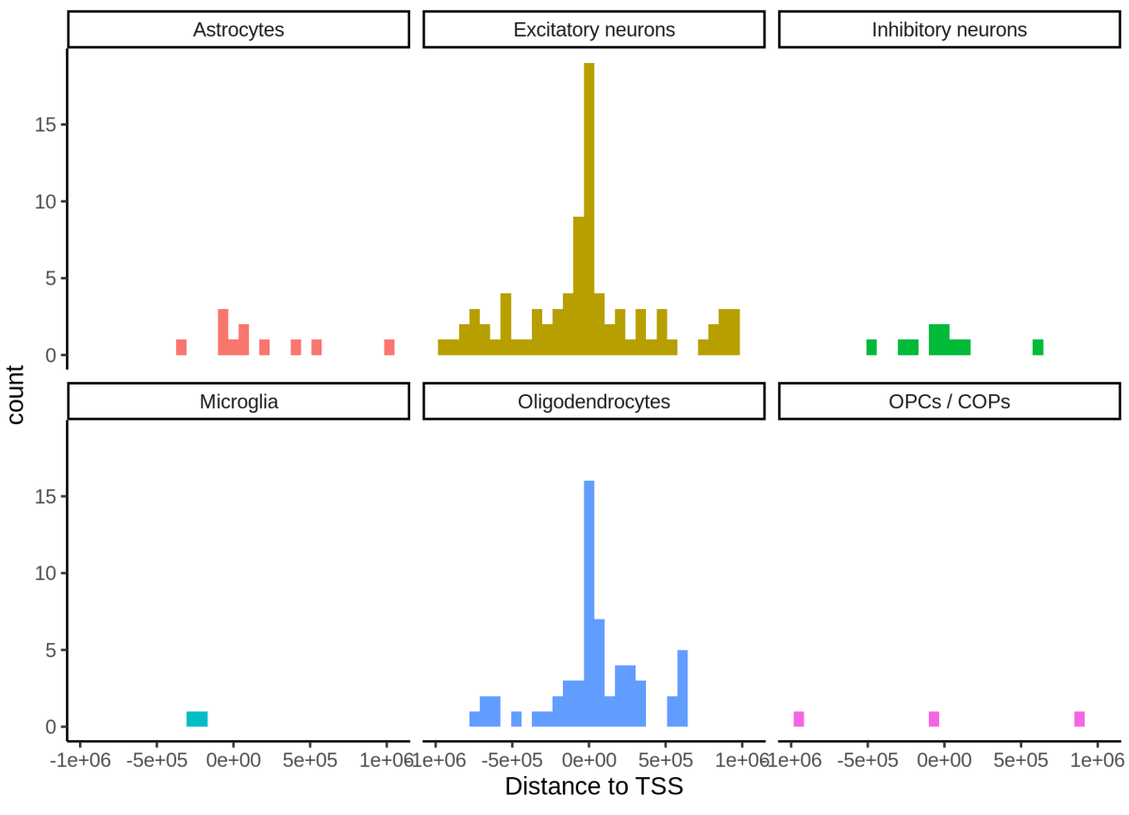
A



B



C



D

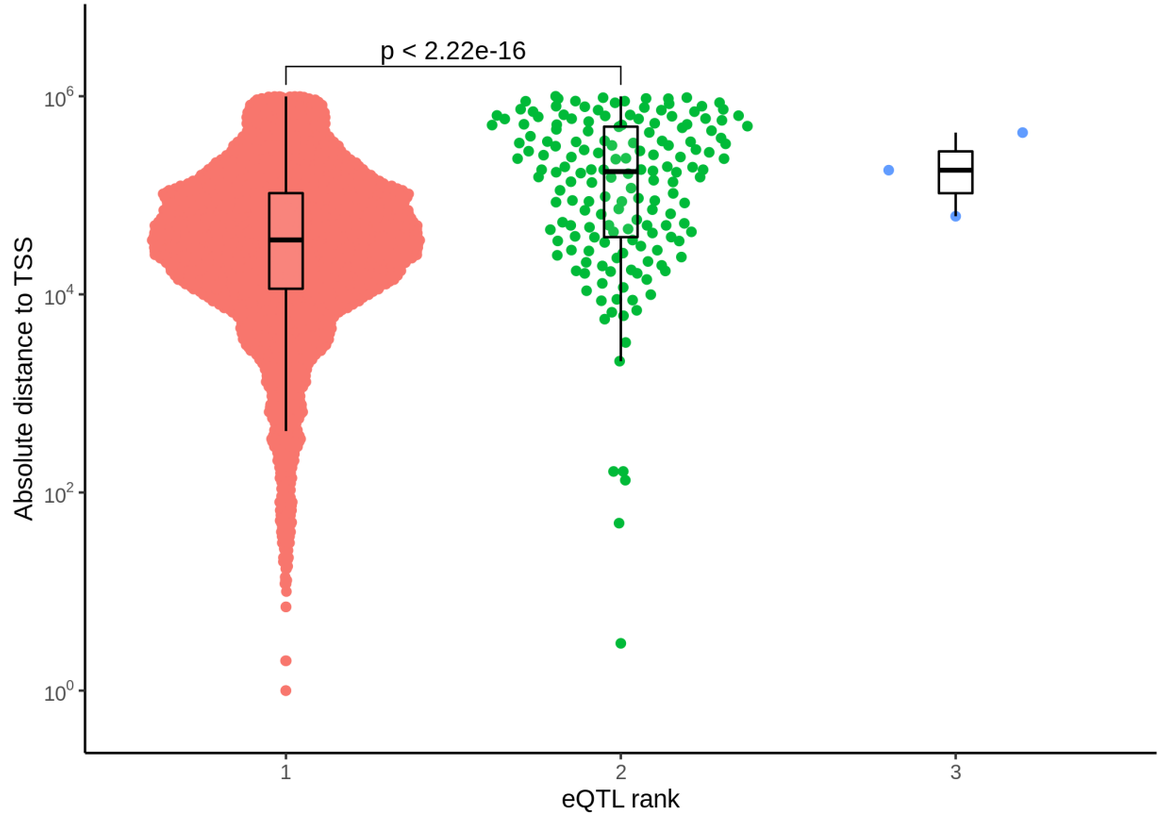


Figure S10: Independent cis-eQTL discoveries and properties.
A) Schematic representation of the effect of an additional independent SNP on gene expression. **B**) Number of independent cis-eQTLs (5% FDR) for each cell type. **C**) Distribution of independent cis-eQTLs around the TSS. **D**) Absolute distance to the TSS for the main cis-eQTL, the secondary cis-eQTL and the third cis-eQTL. The Wilcoxon test two-sided pvalue is shown (n=6173 primary eQTLs, n=168 secondary eQTLs, n=3 third eQTLs). Boxplots show the median, first and third quartile, and whiskers extend up to 1.5 times the interquartile range.

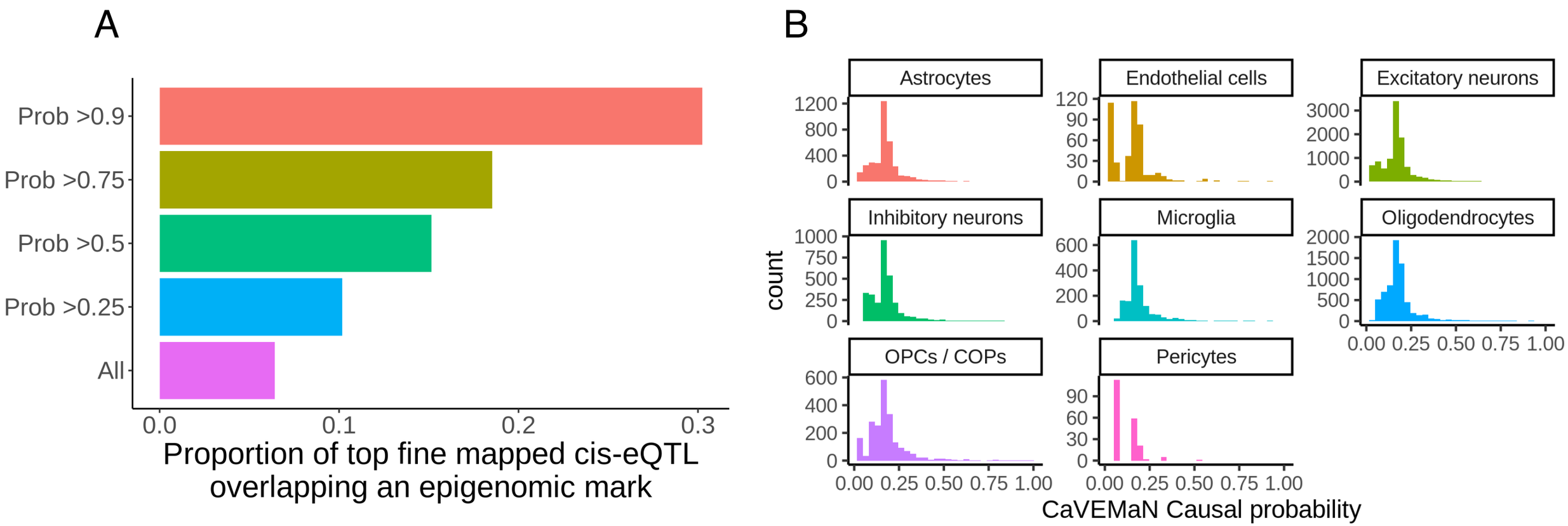


Figure S11: Fine-mapping of cis-eQTLs.
A) Proportion of fine-mapped SNPs overlapping and epigenomic mark for different causal probability thresholds.
B) Distribution of causal probabilities for cis-eQTLs in the different cell types.

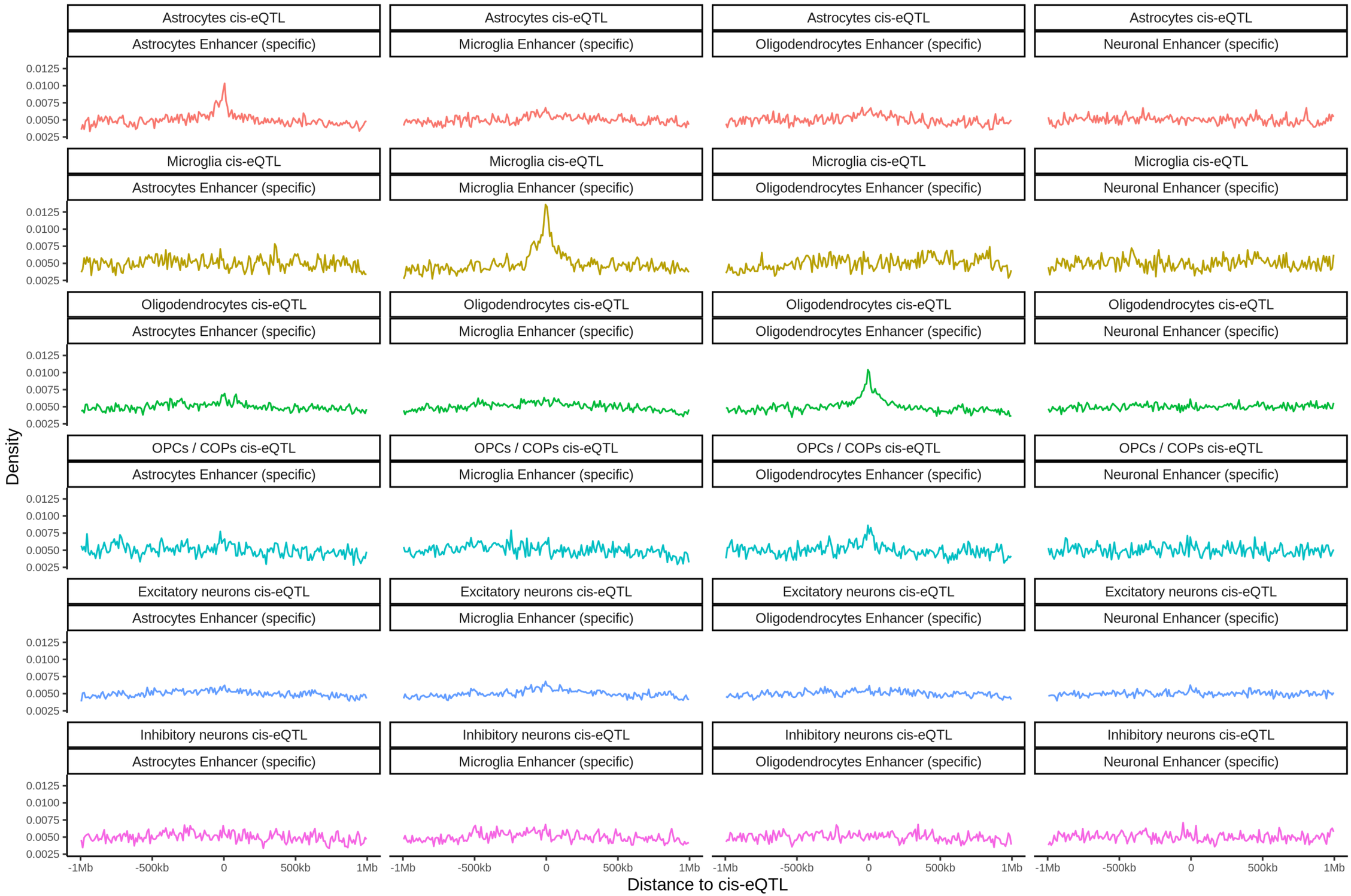


Figure S12: Enrichment of cell type specific epigenomic marks around cis-eQTLs (Nott)
Enrichment of cell-type specific epigenomic marks (CHIP-seq of H3K4me3 and H3K27ac) around cis-eQTLs for each cell type. The number of cell type specific epigenomic marks was computed for each 10kb around the cis-eQTLs in a 2MB window.

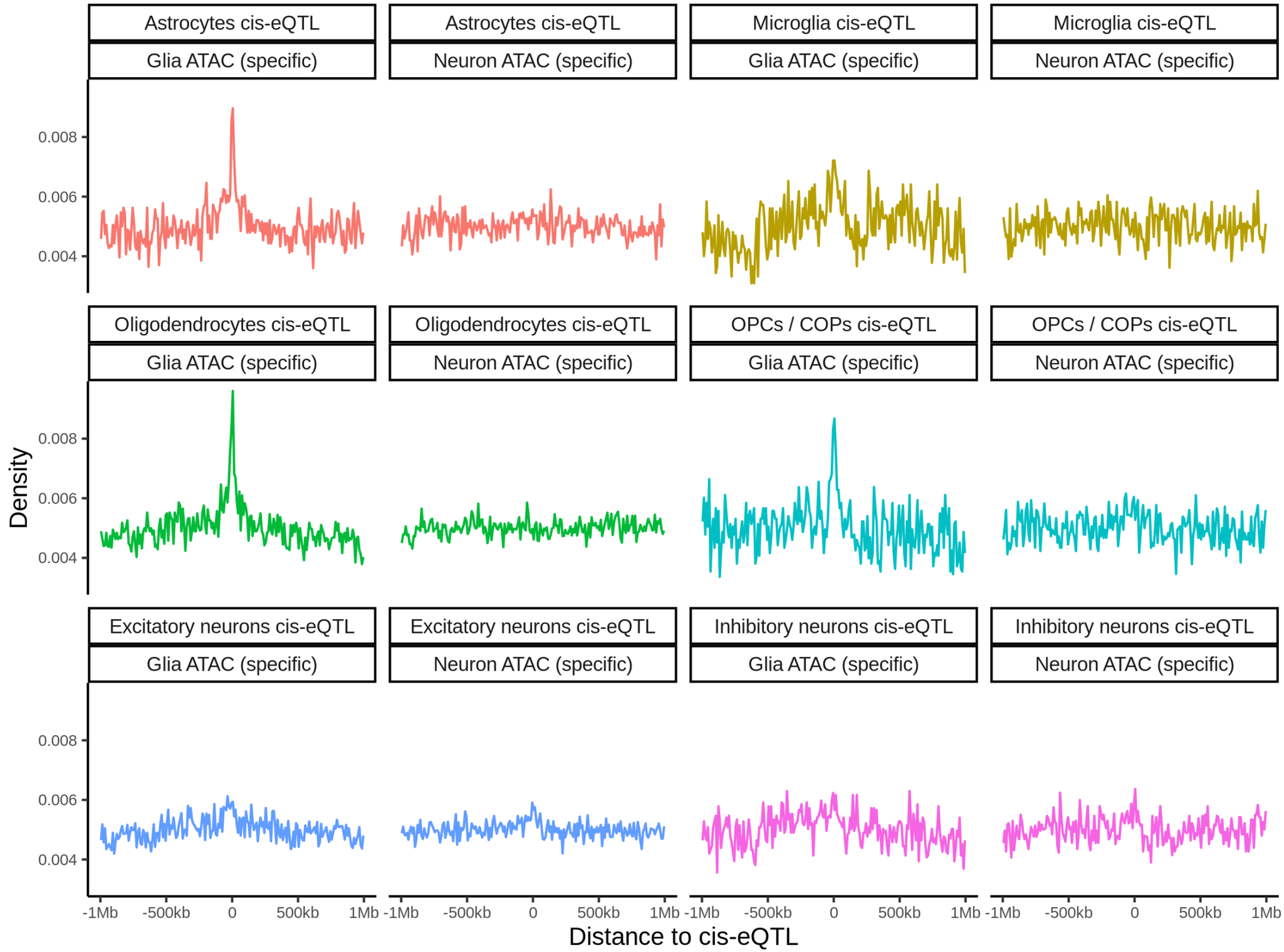


Figure S13: Enrichment of cell type specific epigenomic marks around cis-eQTLs (Fullard)
Enrichment of cell-type specific epigenomic marks (ATAC-seq on NeuN+ and NeuN- sorted nuclei) around cis-eQTLs for each cell type. The number of cell type specific epigenomic marks was computed for each 10kb around the cis-eQTLs in a 2MB window.

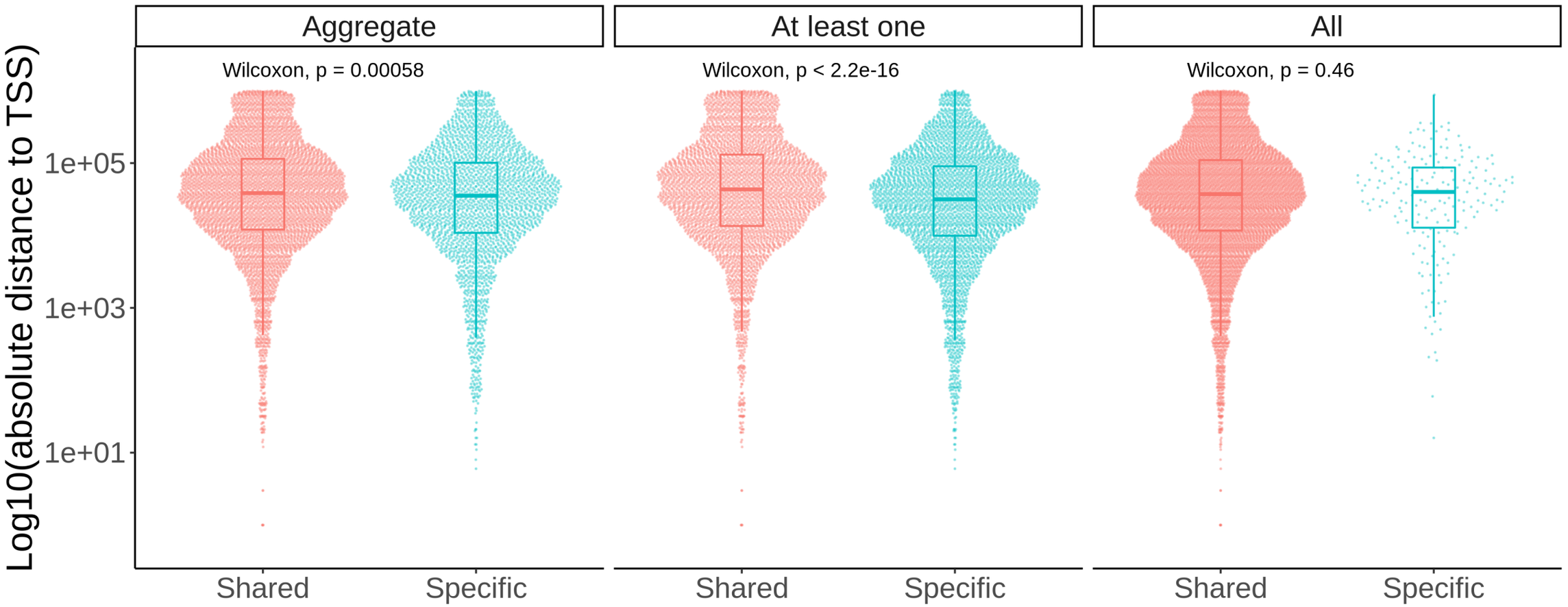


Figure S14: Absolute distance to the TSS for genes with a cell type specific genetic effect
Absolute distance to the TSS for cis-eQTL genes with (and without) a different genetic effect in the discovered cell type (5%FDR) than the: 1) aggregate genetic effect in all other cell types (N=4925/2661 shared/specific eQTLs respectively), 2) at least one other cell type (N=4040/3537 shared/specific eQTLs respectively), and 3) all other cell types (N=7401/185 shared/specific eQTLs respectively). Boxplots show the median, first and third quartile, and whiskers extend up to 1.5 times the interquartile range. The pvalues were obtained using a two-sided Wilcoxon test and are not corrected for multiple testing.

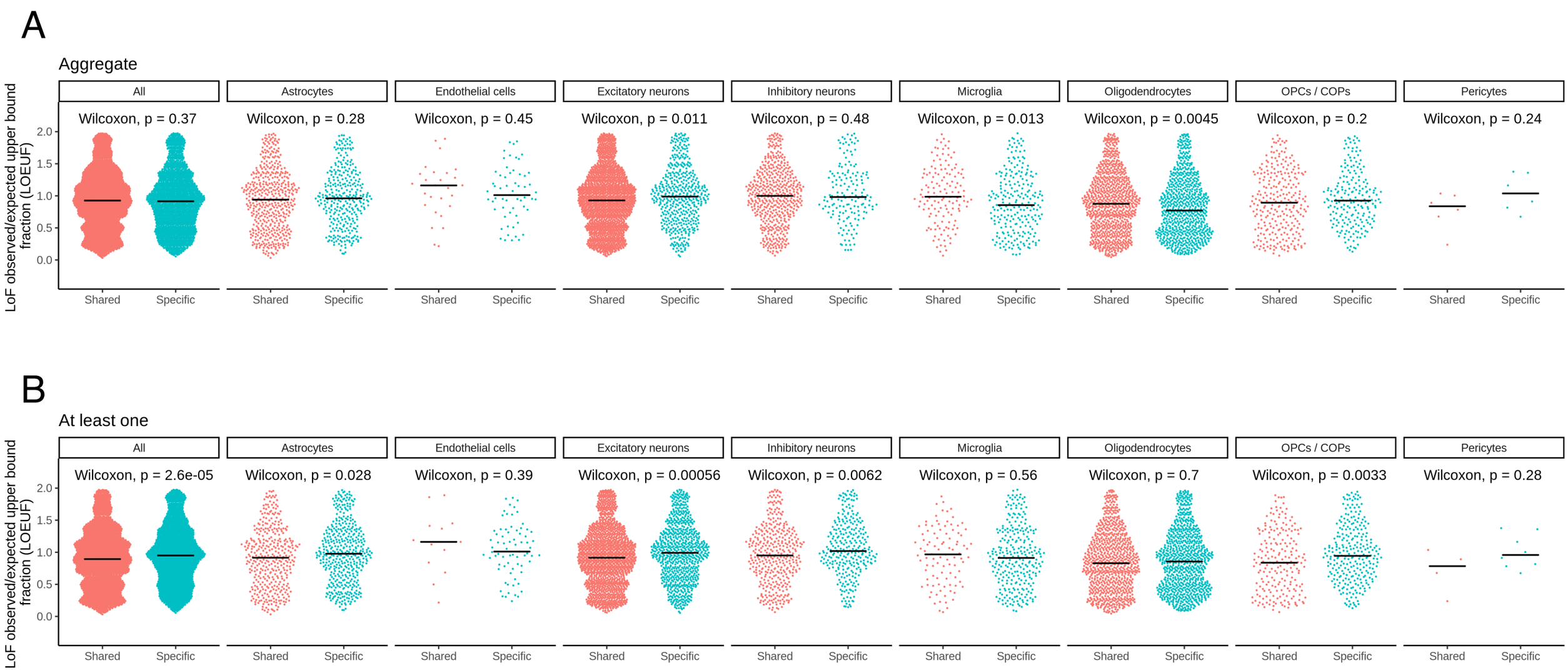


Figure S15: Constraint scores for cell type specific cis-eQTLs
A) LOEUF scores for genes with (and without) a cis-eQTL with a different genetic effect in the discovered cell type than the aggregate genetic effect in all other cell types (5%FDR). **B**) LOEUF scores for genes with and without a cis-eQTL with a different genetic effect in the discovered cell type than at least one other cell type (5%FDR). **C**) LOEUF scores for genes with and without a cis-eQTL with a different genetic effect in the discovered cell type than all other cell types (5%FDR). More constrained genes have lower LOEUF scores. The pvalues were obtained using a two-sided Wilcoxon test and are not corrected for multiple testing.

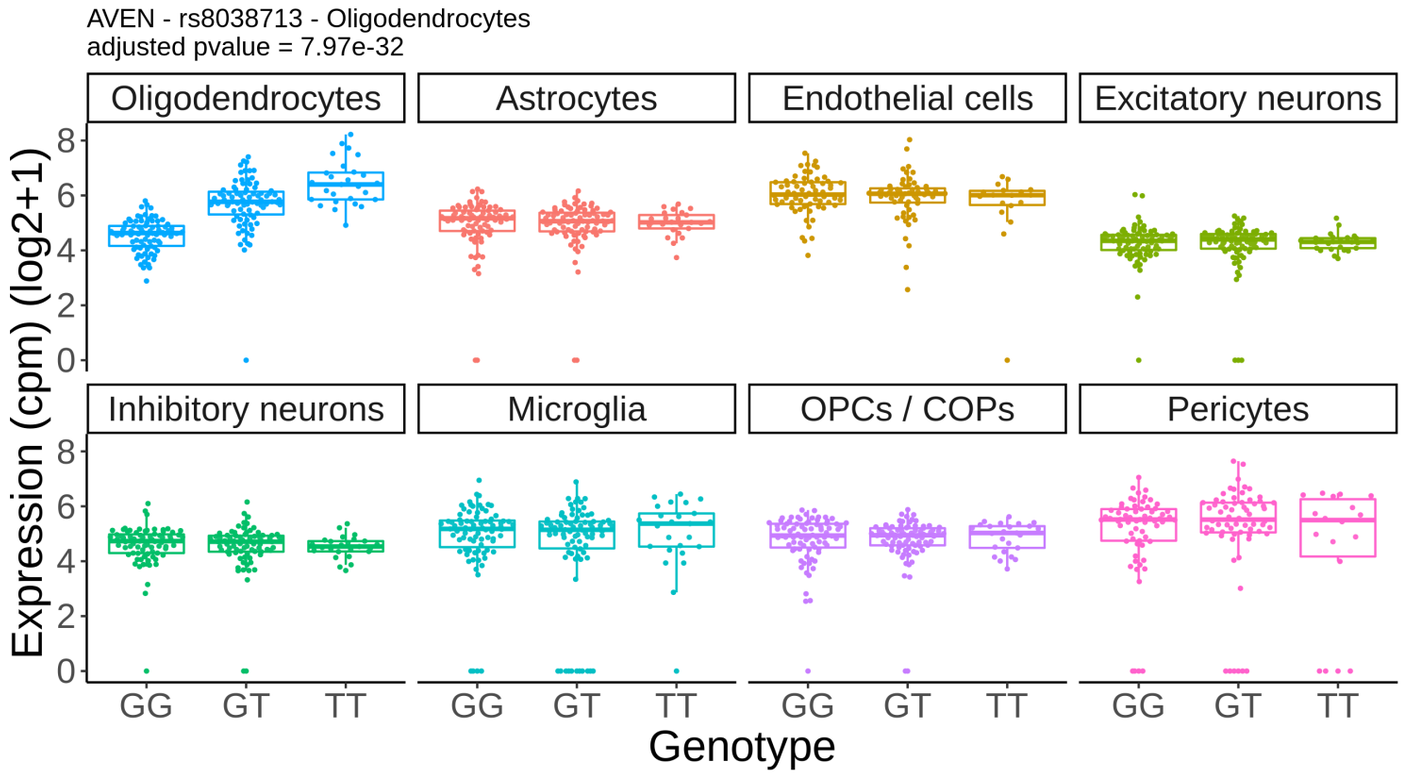


Figure S16: Examples of cell type specific cis-eQTLs.
The level of expression for different genotypes is shown for two different genes. Each dot represents one individual. The displayed pvalue is the interaction pvalue testing whether the genetic effect in the discovered cell type is different from the genetic effects in all other cell types (same SNP-gene pairs) (n = 192 biologically independent samples). Boxplots show the median, first and third quartile, and whiskers extend up to 1.5 times the interquartile range. The two-sided pvalues were obtained using a negative binomial mixed model (Wald Z statistic) and are corrected for multiple testing (Methods).

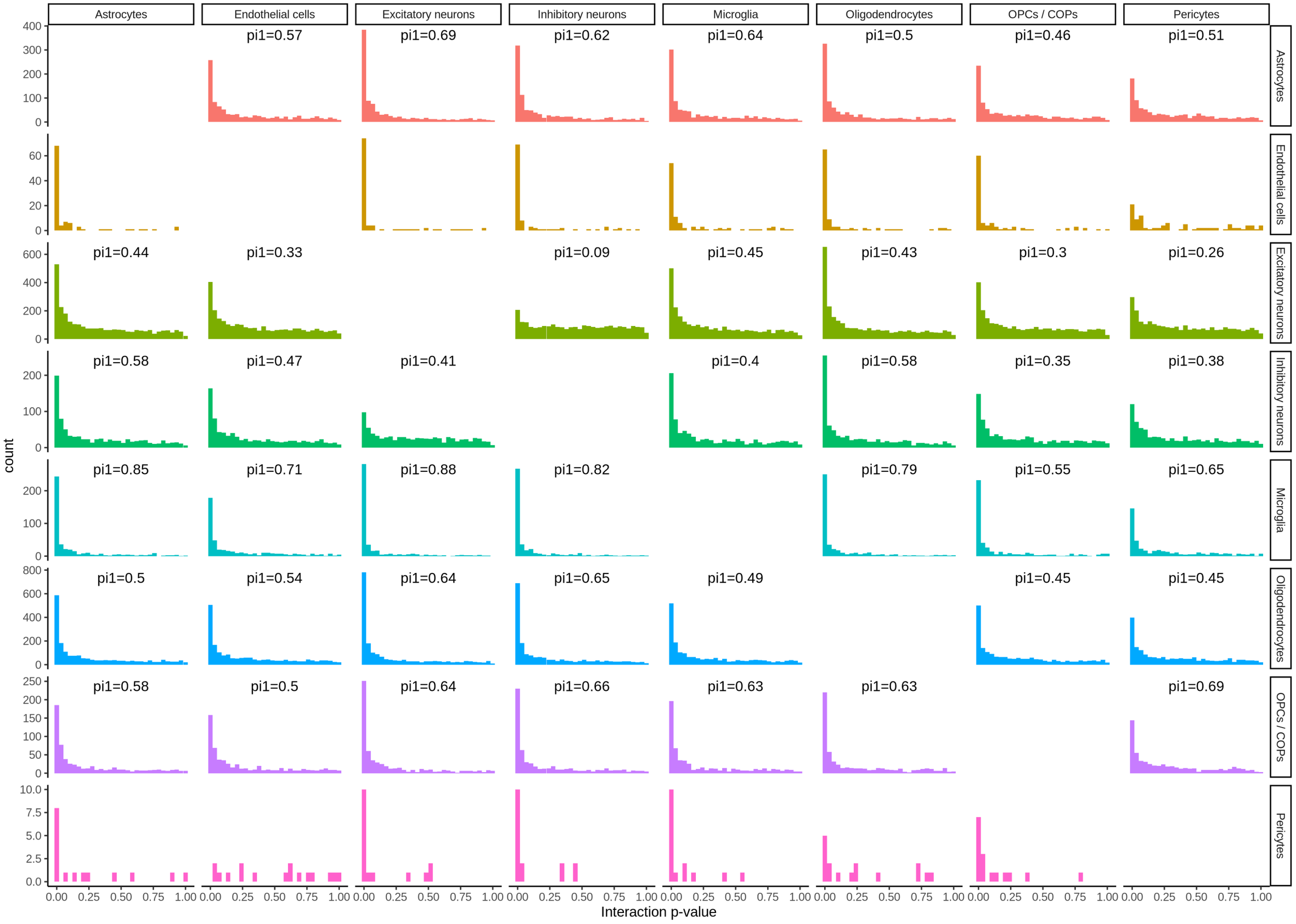


Figure S17: Distribution of interaction pvalues for cell type specific cis-eQTLs
Histograms of interaction pvalues testing for cell type specific effects. For cis-eQTLs (5% FDR) discovered in each cell type (rows), we tested whether the genetic effect for the same SNP-gene were significantly different in the other cell types (columns). The pi1 statistic estimated the proportion of true alternative hypotheses (i.e. the proportion of SNP-gene pairs with a different genetic effect in the tested cell type).

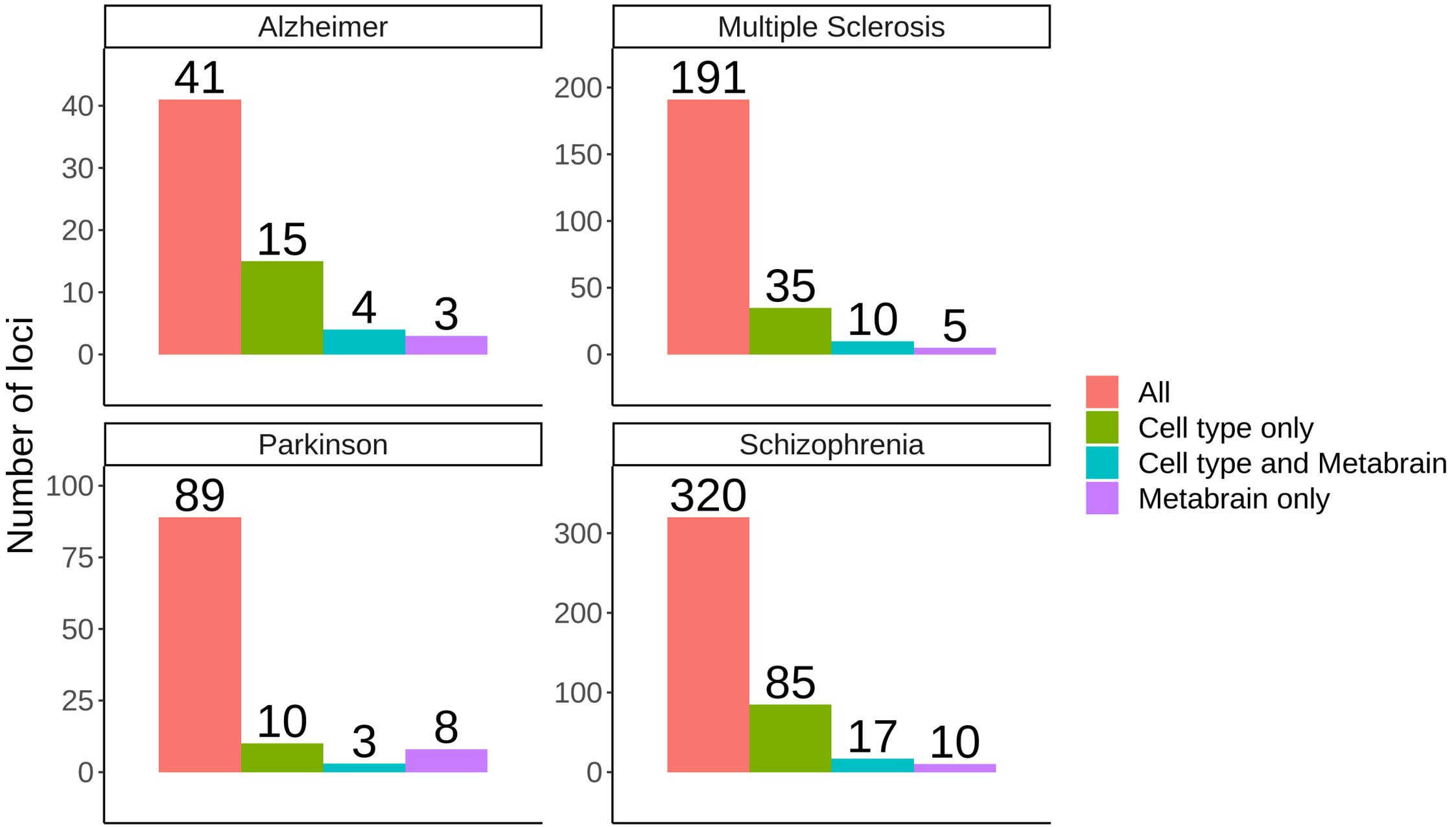


Figure S18: Number of colocalized loci
Number of GWAS loci which colocalize with a cell type level eQTL only (posterior probability >0.7), with both a cell type level eQTL and a bulk eQTL from the Metabrain study (using SMR), and loci that only colocalize in the Metabrain study.

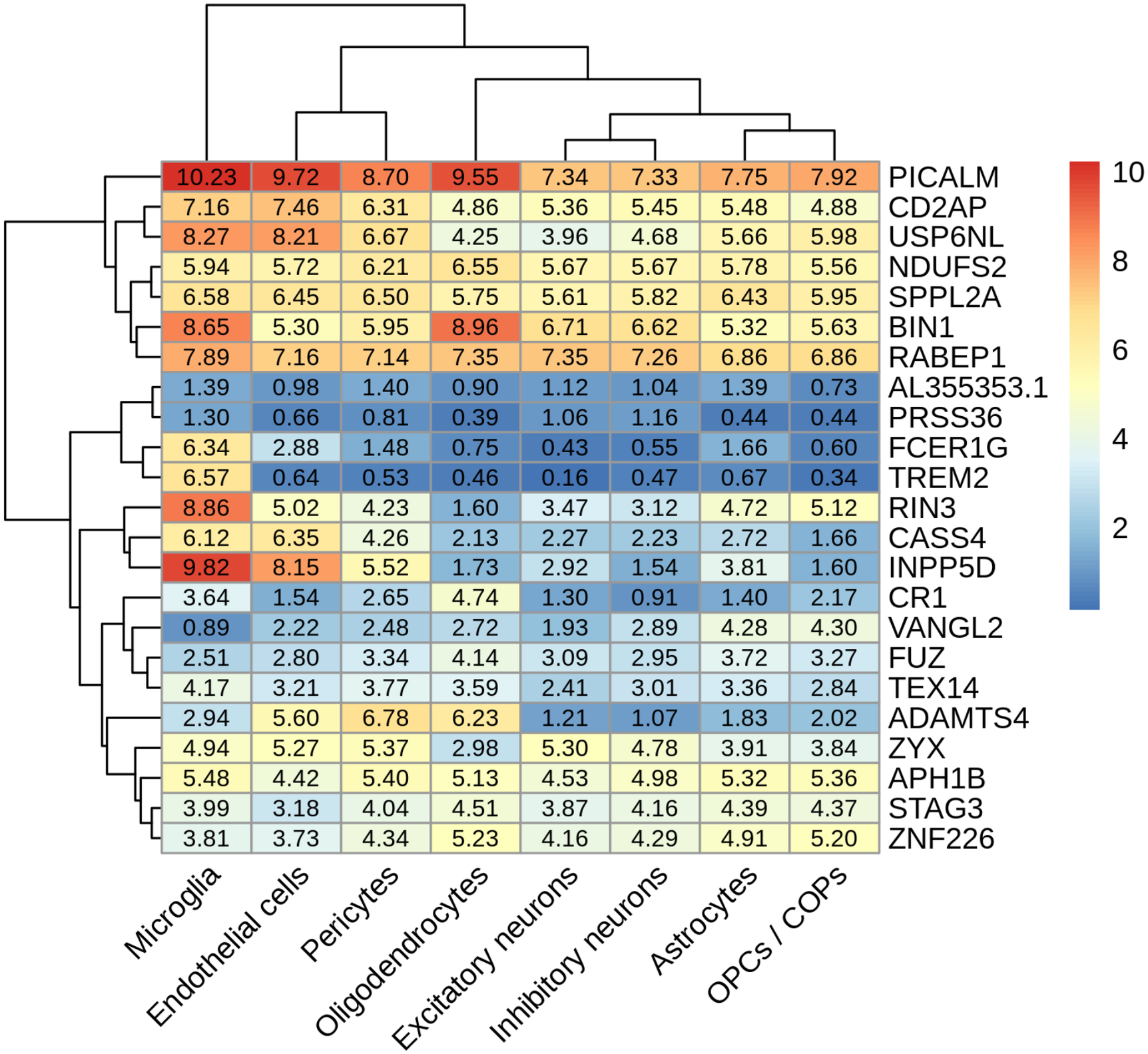


Figure S19: Expression of Alzheimer colocalized genes
Median log2(CPM+1) for each colocalized gene in each cell type across the 192 individuals.

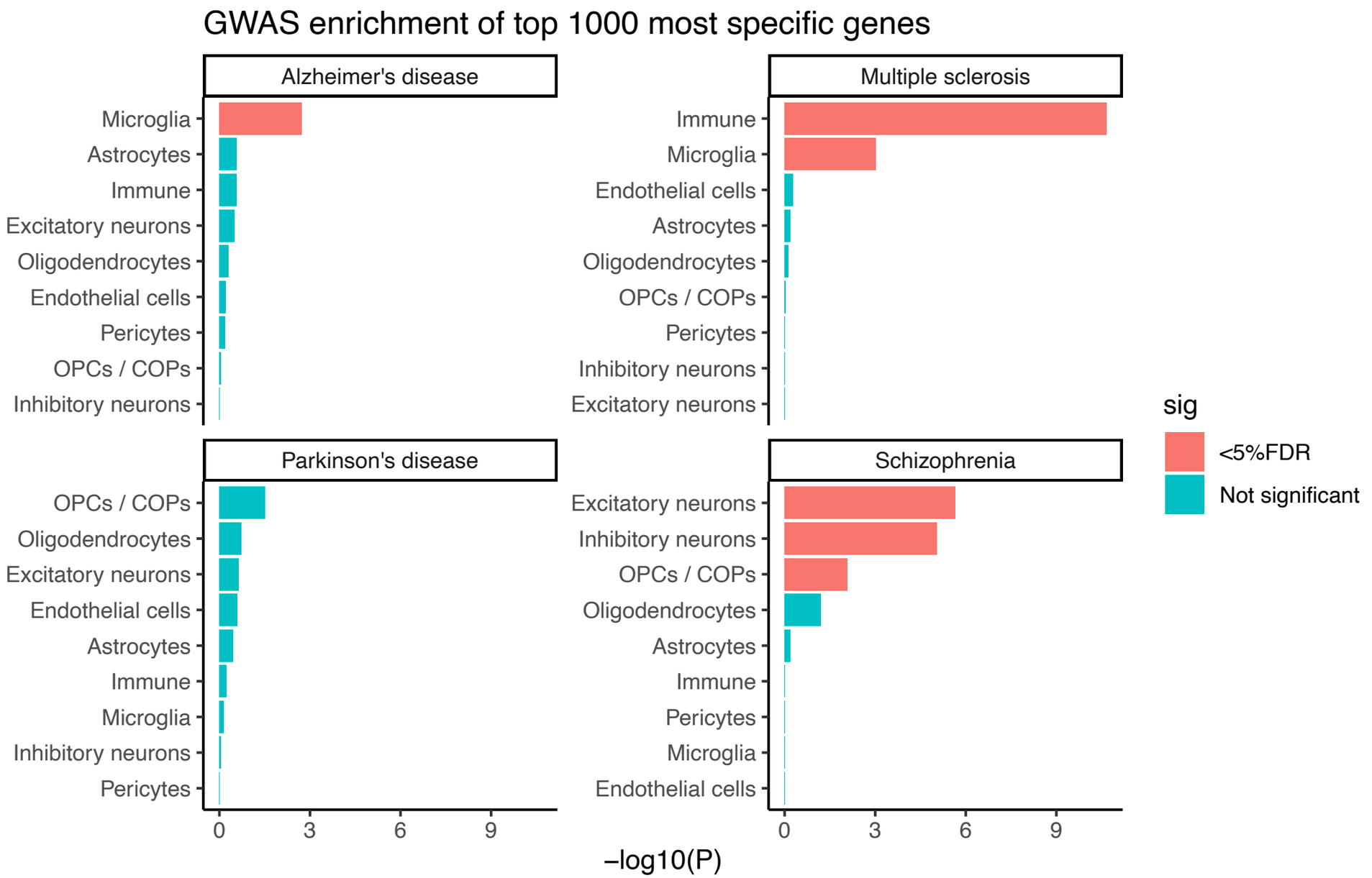


Figure S20: Genetic enrichment of cell type specific genes
Enrichment strength ($-\log_{10}P$) of the top 1000 most specific genes in each cell type in genetic associations from diverse GWAS. The one-sided pvalues were obtained using MAGMA.

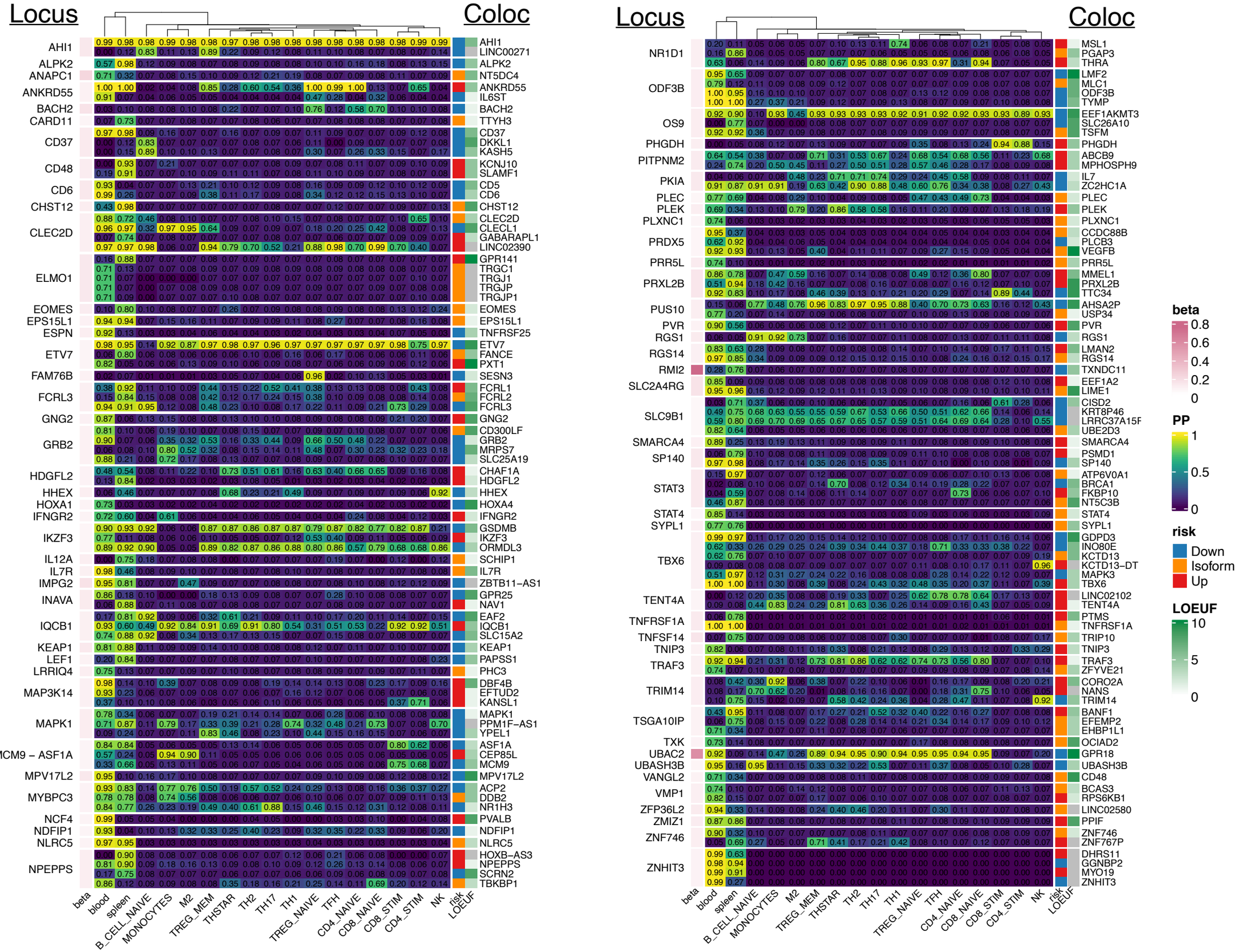


Figure S21: Colocalization results for multiple sclerosis in immune tissues and cell types

Posterior probabilities of shared genetic signal between the multiple sclerosis GWAS and eQTL discovered in immune tissues from GTEx (blood, spleen) and immune cell types. The OR column indicates the odds ratio of the top GWAS SNP at the locus. The closest gene to the top GWAS signal is indicated on the left, the colocalized genes are indicated on the right. The risk column indicates whether an increase in gene expression leads to an increase in disease risk (red), a decrease in disease risk (blue) or whether the colocalization signal is due to a splicing QTL (orange). The LOEUF column indicates whether the gene is constrained (low score) or not (high score).

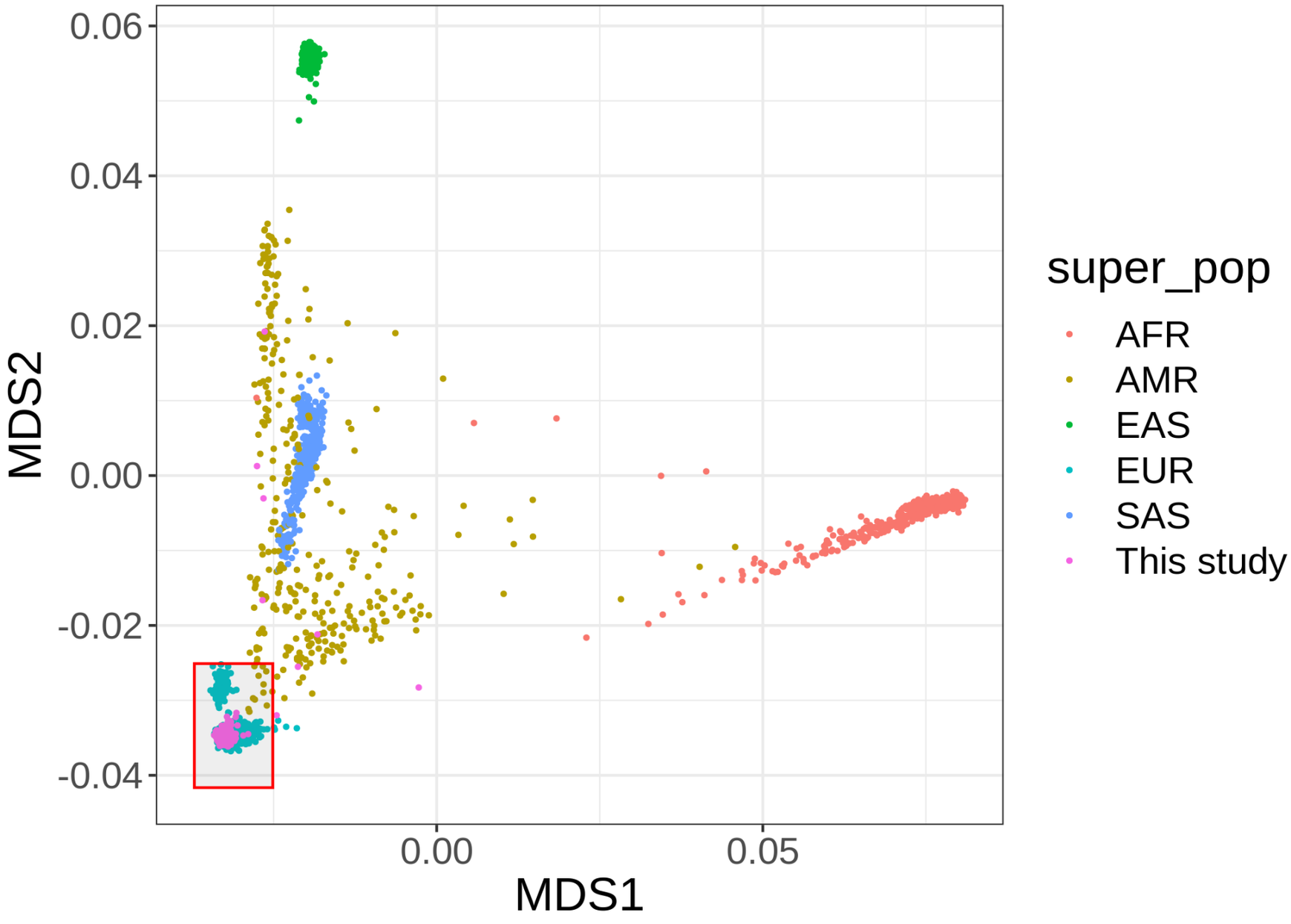


Figure S22: Multidimensional scaling of genotyping data
MDS1 and MDS2 of our genotyped samples (in pink) with individuals from the 1000 genomes project. The red squares indicate 3 standard deviations on MDS1 and MDS2 derived from EUR individuals from the 1000 genomes project.

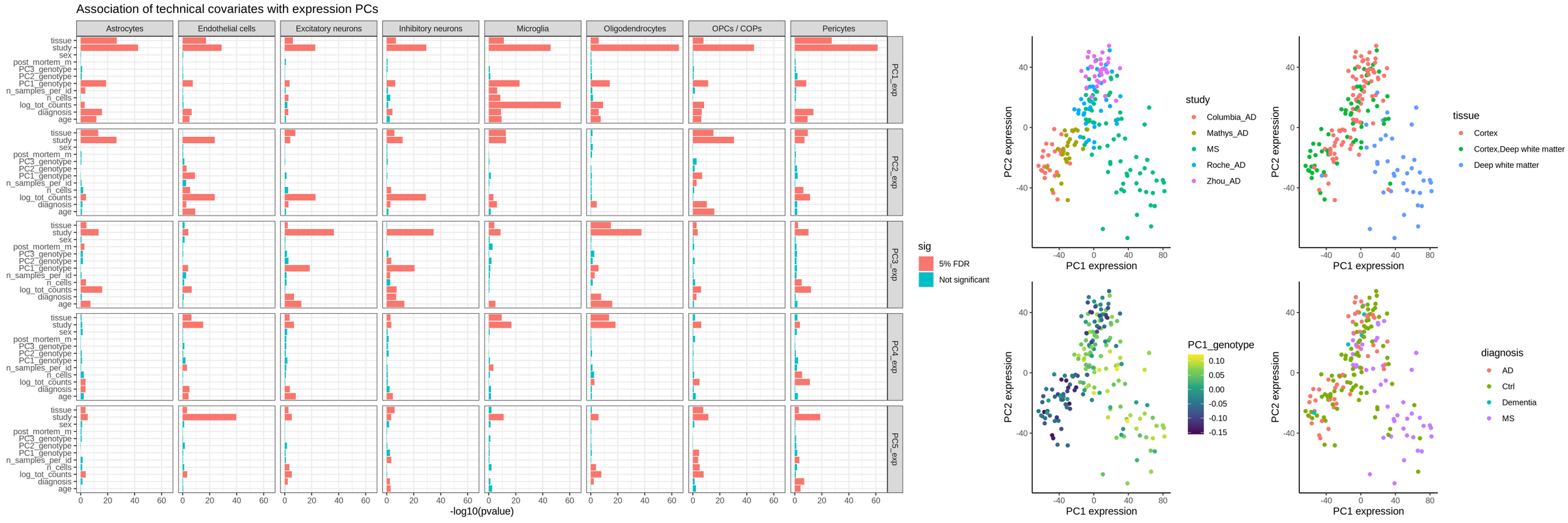


Figure S23: Association between technical covariates and expression PCs

A) Association between technical covariates and the first 5 expression PCs in each cell type (-log10(pvalue) of a pearson correlation for numerical variable and of a F-test between a linear regression model with the categorical variable and a reduced regression model with an intercept only). **B**) Sample projections on PC1 and PC2 for astrocytes, color coded by the most associated technical covariates. Two-sided pvalues were computed using Pearson correlation for numerical variables and using linear regression for categorical variables (F-statistic). Pvalues were not corrected for multiple testing.

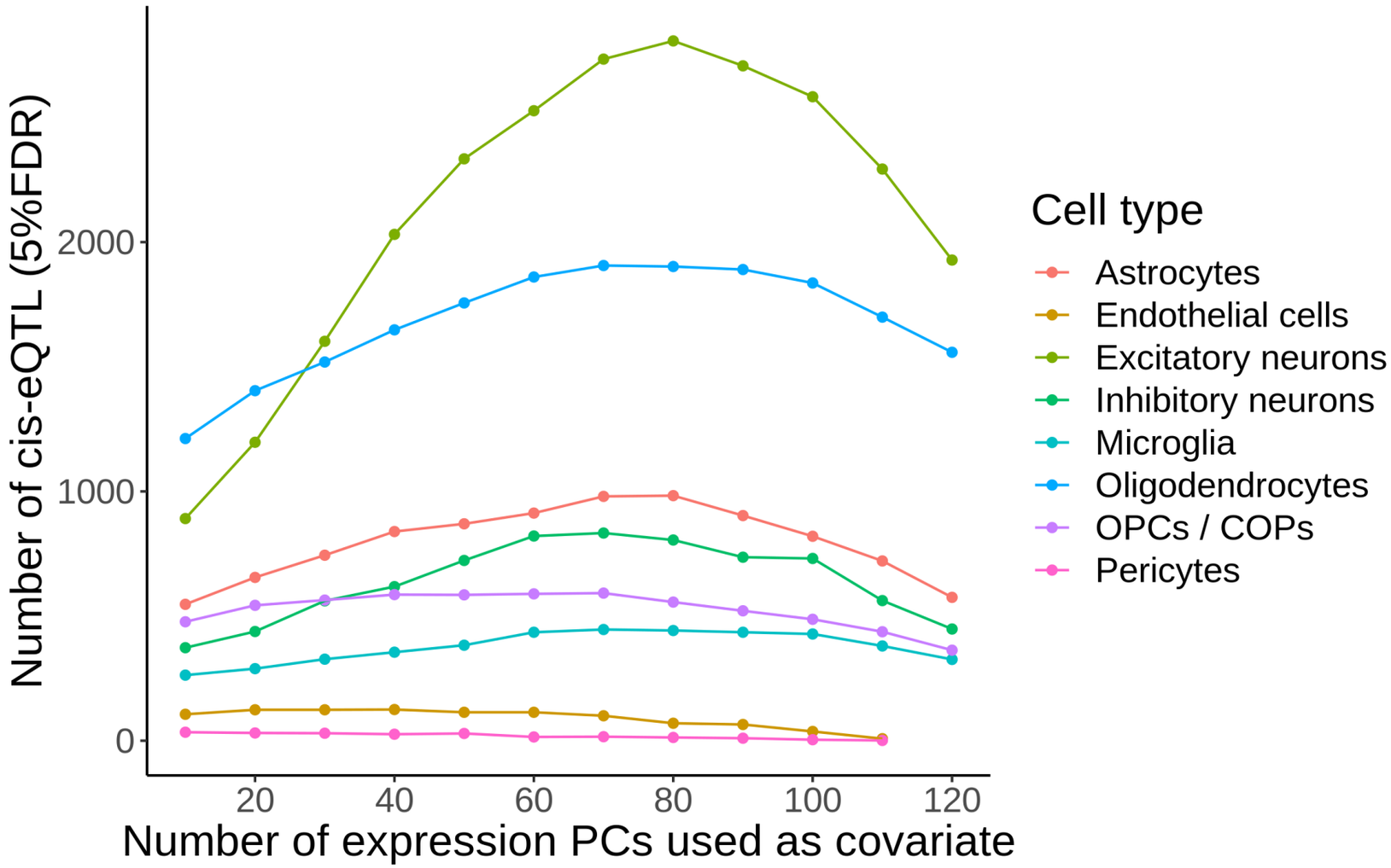


Figure S24: Number of cis-eQTLs discovered in function of number of principal components
Number of cis-eQTLs discovered (5% FDR) in function of the number of gene expression principal components used as covariate for the 8 major brain cell types.