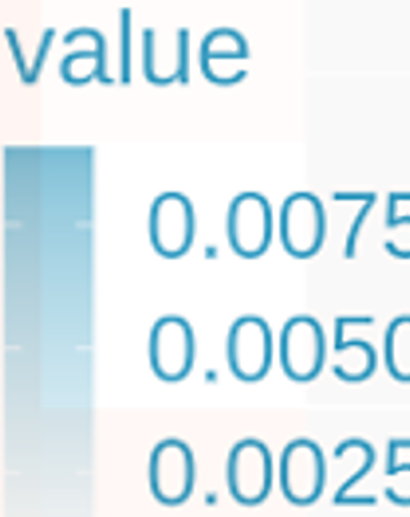
# The eQTL and beyond

A cell type specific view on genetic variant-associated changes in gene (co-)expression in the context of Alzheimer's Disease

Madelon Stol

Delft University of Technology




TU Delft

# The eQTL and beyond

A cell type specific view on genetic variant-associated changes in gene (co-)expression in the context of Alzheimer's Disease

by

Madelon Stol

to obtain the degree of Master of Science
at Delft University of Technology,
to be defended publicly on
October 30th, 2023 at 15:00.

Student number: 4237323
Project duration: February 2023 – October 2023

| Thesis committee: | | |
|---|---|---|
| | Prof. dr. ir. M.J.T. Reinders | TU Delft, Responsible advisor |
| | G.A. Bouland MSc. | TU Delft, Daily supervisor |
| | Dr. N. Tesi | TU Delft, Daily supervisor |
| | Dr. C. Lofi | TU Delft |



TU Delft

*TU Delft - Computer Science Master Thesis*

# The eQTL and beyond: a cell type specific view on genetic variant-associated changes in gene (co-)expression in the context of Alzheimer's Disease

Madelon Stol[1]

[1]Delft Bioinformatics Lab, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands.

## Abstract

**Understanding the role of genes and genetic variants is a key challenge in unraveling the driving mechanisms of Alzheimer's disease (AD). Single-cell RNA sequencing is a technique that quantifies gene expression at the cell (type) level enabling investigation of the roles of different cell types in disease. We analyzed changes in gene (co-)expression associated with genetic variants using single-cell RNA sequencing data (>1.3 million cells) from the dorsolateral prefrontal cortex (DLPFC) of 379 individuals of the ROSMAP cohort. Our single cell expression quantitative trait loci (sc-eQTL) analysis determined 3,337,065 sc-eQTLs, linking 1,882,645 SNPs to changes in expression of 8,057 genes in 7 major cell types. Next, we investigated the association of genetic variants with changes in co-expression for gene pairs (co-eQTLs), focusing on a set of variants and genes relevant to AD. Our novel non-parametric method for co-eQTL analysis compares gene co-expression distributions between SNP genotypes. We found 6,878 cell type specific co-eQTLs (variant-gene-gene combinations) relating to 18 AD variants. Although a substantial proportion of the findings is driven by eQTL effects, our method identified co-eQTLs that would not have been discovered in a correlation-based analysis. Most notable, we found variant rs13237518 (located in the TMEM106B gene) to associate with expression changes in a subset of 25 genes in excitatory neurons which is possibly indicative of higher-level disruptions related to the variant. Overall, we show that exploring genetic variant-associated changes in gene (co-)expression is a promising approach in finding cell type specific mechanisms that may be altered in AD.**

## Introduction

Alzheimer's disease (AD), the most common form of dementia, is known for its hallmark symptoms like memory impairment and decline in executive functioning (DeTure and Dickson 2019). AD is a major healthcare burden on society and a leading cause of death. At present, neither curative nor preventative treatment is available for AD (van der Flier *et al.* 2023). Even though neuropathological features of AD are known, such as the presence of amyloid-beta plaques in the brain and neurofibrillary tau tangles in neurons (typically quantified with Braak staging and CERAD scores), the exact disease mechanism is not fully understood (Bai *et al.* 2021). AD is highly heritable (Bellenguez *et al.* 2022): in fact, the heritability of the most common form of the disease (Late Onset Alzheimer's Disease) ranges between 60-80% based on twin studies (Gatz *et al.* 2006). Therefore, understanding the role of genes and genetic variants is a key challenge in unraveling the mechanisms driving the disease process, and finding possible therapeutic targets.

In order to identify genetic modifiers of diseases, typically genome-wide association studies (GWAS) and differential expression analyses (DEA) are performed (Westra and Franke 2014). GWAS compare the frequency of genetic variants (single nucleotide polymorphisms (SNPs), insertions and deletions (indels)) between individuals with different phenotypes (for example: AD patients and controls), to identify relevant disease loci. In DEA, gene expression is compared between cases and controls, to identify genes with significantly higher or lower gene expression between the respective groups. While these approaches find genetic variants and genes of possible relevance, they do not directly reveal the functional consequences of genetic variants and/or the effects on differential gene expression in disease.

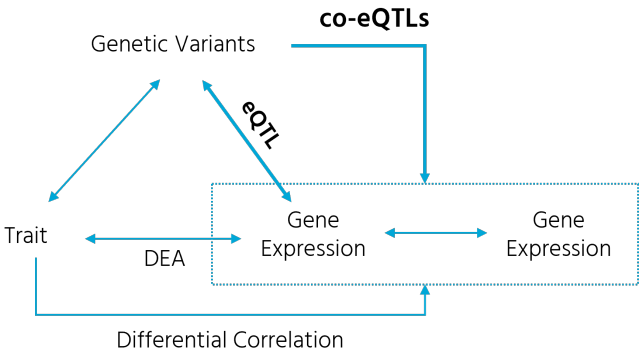



**Figure 1** Studies relating genetic variants and gene expression

To do so, expression quantitative trait loci (eQTL) are often used, which directly relate genetic variants to gene expression (Westra and Franke 2014). The loci correspond to SNP genotypes which can be obtained from genotyping arrays or whole genome sequencing (WGS) data. The quantitative trait is the gene expression which can be obtained from

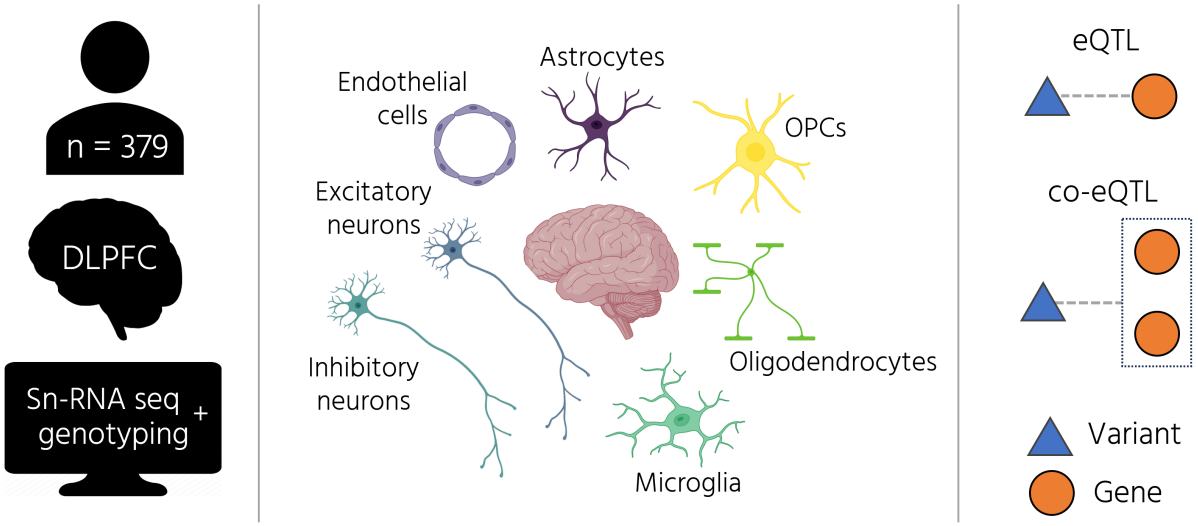



**Figure 2** Data and analysis overview of the (co-)expression QTL analyses

RNA sequencing data by quantification of RNA transcripts mapped to a gene.

Bulk RNA sequencing data provides the number of RNA transcripts on sample level, which reflects the average expression of a gene in a tissue and is strongly driven by the RNA transcripts present in the most prevalent cell type(s) (Zhang and Zhao 2023; Mathys *et al.* 2019). In fact, different cell types execute different functions and as such may have different roles in disease. These functions of a cell are also reflected in their transcriptome, the set of RNA transcripts present in a cell (Jovic *et al.* 2022a). In the past decade, single cell RNA sequencing (scRNA-seq) techniques have been developed that allow for measuring transcripts at the level of individual cells (Jovic *et al.* 2022b). This uncovered a new layer in assessing roles of genes and their relationship with disease. ScRNA-seq data, combined with genetic variant information, enables the examination of eQTLs on cell type specific level (sc-eQTL), which also aids the discovery of eQTLs in less abundant cell types which were previously obscured in bulk eQTL analyses (Zhang and Zhao 2023).

Most eQTL analyses assess variant-gene pairs independently and DEAs assess expression changes in individual genes, while biological networks are characterized by genes affecting and regulating the expression of other genes (Li *et al.* 2023a). For example, a genetic variant relating to a transcription factor may change its binding affinity and regulate expression of another gene (Flynn *et al.* 2022). Therefore, a co-expression analysis that considers genetic variants may highlight these possible interactions.

Recent work by Bouland *et al.* (2023b) investigated gene-gene relations in the context of AD. Their analysis determines differential correlation between genes in a cell type-specific context using scRNA-seq data of brain tissue from AD patients and controls. Significant differences in expression correlation between groups (AD versus controls) are used to build a differential co-expression network, a graph with nodes for genes and edges determined by the differential co-expression relation. Network-level assessment of differential correlation points towards hub genes that are involved in altered associations between genes and can be used for gene prioritization. As this study has a case-control design, the influence of genetic variants on gene-gene relations has not directly been taken into account, the findings do suggest however that exploring gene-gene relations is a promising direction in investigating the role of variants and genes in AD.

Few studies used single cell RNA sequencing data obtained from blood to determine the correlation between gene expression, and further relate these with SNP genotypes. For example, in Oelen *et al.* (2022), co-expression QTLs, defined as *'SNP genotypes affecting the co-expression relationship of a gene pair'*, were assessed for peripheral blood mononuclear cells (PBMCs) considering their response to pathogen stimulation. The relationship was assessed by determining the Spearman correlation coefficient between genes using single-cell gene expression values for each individual and relating the correlation to the SNP genotype in a weighted linear model. Li *et al.* (2023a) extends on this by integrating multiple PBMC datasets into a meta-analysis. Spearman's rank-based correlation was found to be the best fit for their data analysis. The SNP-gene-gene set tested was limited by considering only significant eQTLs from Võsa *et al.* (2021) and testing those SNP-gene combinations for all genes that were expressed in at least 50% of cells.

Here we present our eQTL and co-expression QTL (co-eQTL) analysis performed with single nucleus RNA sequencing (sn-RNA seq) data from the dorsolateral prefrontal cortex (DLPFC) of 379 participants of the ROSMAP cohort of which 141 are diagnosed with AD. We show cell type specific eQTLs for 7 major cell types in the human brain: astro-

| Cogdx | n | Coding |
|---|---|---|
| 1 | 122 | NCI: No cognitive impairment (No impaired domains) |
| 2 | 98 | MCI: Mild cognitive impairment (One impaired domain) and NO other cause of CI |
| 3 | 5 | MCI: Mild cognitive impairment (One impaired domain) AND another cause of CI |
| 4 | 121 | AD: Alzheimer's dementia and NO other cause of CI (NINCDS PROB AD) |
| 5 | 20 | AD: Alzheimer's dementia AND another cause of CI (NINCDS POSS AD) |
| 6 | 13 | Other dementia: Other primary cause of dementia |

**Table 1** Cognitive characteristics: number of individuals per cognitive diagnosis along with diagnosis definition

cytes, endothelial cells, inhibitory neurons, excitatory neurons, microglia, oligodendrocytes and oligodendrocyte progenitor cells (OPCs) (Green *et al.* 2023).

We analyzed the influence of genetic variants on gene co-expression (co-eQTL) focusing on a subset of variants and genes relevant to Alzheimer's disease from previous studies. Unlike previous co-eQTL studies that link the correlation between expression of gene-gene pairs to genetic variants, we used a non-parametric method to assess changes in expression of gene-gene pairs between SNP genotypes. This was done because correlation measures assess linearity (Pearson) or monotonicity (Spearman) of the gene-gene relation, while these assumptions might not hold as genetic regulation mechanisms are known to be subject to non-linear dynamics (Kontio *et al.* 2020; Hou *et al.* 2022).

## Demographics

We used data from participants of the Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP) (Bennett *et al.* 2018). Both studies are longitudinal epidemiological studies of elderly people who were recruited and followed up over time to investigate the development of AD and other diseases. We included only individuals that had both single-cell RNA sequencing of brain tissue obtained from the dorsolateral prefrontal cortex (DLPFC) and genotyping information (based on SNP array) available. In total, 379 individuals ($n_M$ = 126 / $n_F$ = 253) were included, all of European descent. Of these, 141 (72.3% female) were AD cases , 13 (53.8% female) were non-AD dementia cases, 103 (64.1% female) were diagnosed with mild cognitive impairment (MCI) and 122 (63.9% female) were controls with no cognitive impairment (NCI). A summary of neuropathological status (Braak stage for presence of neurofibrillary tangles and CERAD score for neuritic plaques) and cognitive diagnosis of the selected individuals is presented in Table 1 and Table 2 respectively.

| CERAD ↓ Braak → | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 4 | 4 | 15 | 15 | 46 | 19 | | |
| 3 | | 5 | 8 | 14 | 10 | | |
| 2 | 1 | 1 | 6 | 40 | 57 | 24 | |
| 1 | | | 2 | 15 | 38 | 56 | 3 |

**Table 2** Neuropathological characteristics - Number of individuals per Braak stage and CERAD score, severity of pathology increases from the top left to the bottom right corner - Braak stage is a measure for the presence of neurofibrillary tangles ranging from 0 to 6 by increasing severity, where severity increases with increased number of affected areas in the brain. CERAD score is a measure for the presence of neuritic plaques ranging from 1 (definite AD) to 4 (no AD).

## Analysis overview

First, we performed a cell type-specific eQTL (sc-eQTL) analysis to identify SNPs[1] that associate with variation in gene expression. To do so, we used linear regression with SNP genotypes as predictors for gene expression, testing variants in the cis-region (± 1Mbp) of the transcription start site (TSS) for genes passing expression QC (see *Methods*). Our analysis determined sc-eQTLs (cell type-specific SNP-gene pairs with a significant association), resulting in sc-eSNPs (SNPs that have a significant effect on gene expression in a specific cell type) and sc-eGenes (cell type-specific genes whose expression is significantly affected by at least one SNP).

Knowing that genetic variants can be driving factors in changes in co-expression relations between genes (Yang *et al.* 2020), we extended our perspective from investigating variance in gene expression explained by genetic variants (sc-eQTL) to variant-associated change in the expression of gene pairs (co-eQTL). We define a co-eQTL to be SNP-gene-gene combination for which there is a significant change in the joint expression of a gene pair across SNP genotypes.

Our method (Figure 3, see *Methods*) is *hypothesis-free*, meaning that **1)** the input set of variants and genes is not constrained to hypotheses on how the variant may affect the co-expression relation, such as that variants in regulatory domains may affect transcription factor binding with consequences for the co-expression relation, and **2)** the method is non-parametric and therefore does not impose any constraints, such as linearity or monotonicity, on the co-expression relation of a gene-pair. This is different from common methods applied in gene co-expression network analyses, which for simplicity assume linear relationships and assess co-expression using Pearson's correlation coefficient while genetic regulation mechanisms are known to be subject to non-linear dynamics (Kontio *et al.* 2020; Hou *et al.* 2022). We do not limit ourselves to assessing whether there is correlation of expression between genes, which is common in network-based gene co-expression analysis (van Dam *et al.* 2018), but investigate any SNP-associated changes in co-expression.

[1] SNP is used to refer to both single nucleotide polymorphisms and small structural variants (indels) unless explicitly defined otherwise

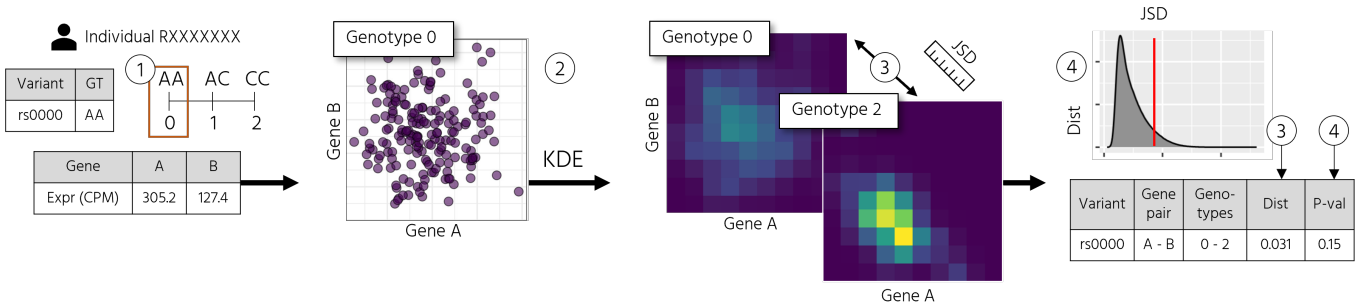



**Figure 3** Co-expression QTL step-wise workflow: **1)** per variant, an individual is assigned to a group (0-1-2) based on its genotype, **2)** scatter plot representation of the joint expression distribution of gene A and gene B in genotype group 0. The individual data points in the scatter plot show the expression for gene A and gene B for an individual assigned to genotype group 0, for every genotype the probability density of the expression of the gene pair is obtained by kernel density estimation (KDE) with a Gaussian kernel placed on each data point, **3)** the resulting co-expression probability density is represented in a heatmap (high local density: yellow - low local density: purple) is compared between 2 genotype groups by calculating a distance metric, the Jensen-Shannon divergence (JSD) per grid-point and summing the local JSD values over all grid-points, and **4)** the nominal significance of the distance between genotype groups for a gene-pair is evaluated using an empirical null distribution of distances created using a permutation strategy per variant per combination of genotype groups - the red vertical line represents a distance for which the nominal p-value is determined by the fraction of larger distances in the null distribution

## sc-eQTL analysis reveals cell type specific gene regulation associated with SNPs

Our analysis found a total of 3,337,065 sc-eQTLs (Table 3), linking 1,882,645 SNPs to changes in expression of 8,057 genes in 7 major cell types that were significant after multiple testing correction (see *Methods*). The largest number of sc-eQTLs (2,064,362) is found in excitatory neurons which is the most abundant cell type present. The number of sc-eQTLs discovered per cell type (Figure 4) correlates with the number of cells per cell type in the input data (Pearson correlation: 0.82, p = 0.0244), which can be explained by the fact that the inclusion of more cells decreases noise in the expression data, in turn leading to more confident results using a linear model. Fujita *et al.* (2022) performed a similar eQTL analysis on the same dataset, a comparison of results is presented in *Supplementary material* where we conclude the main differences in findings stem from differences in multiple testing correction strategy, a lower number of individuals in our analysis, and a different set of tested variants.

Of the 8,057 sc-eGenes, 4,024 (50%) were found to be an eGene in a single cell type only, whereas other sc-eGenes were found in multiple cell types, which emphasizes the importance of analysis of sc-eQTLs on cell type level. Eight genes were identified as eGene in all 7 cell types (*HLA-A*, *HLA-B*, *HLA-C*, *KANSL1*, *LRMDA*, *NUTM2B-AS1*, *RPS26*, *ZDHHC21*). Of note, the first 3 genes are part of the major histocompatibility complex (MHC) class I and encode for cell surface proteins involved in the regulation of the immune system (Janeway *et al.* 2001).

*XRRA1* (X-ray radiation resistance associated 1) repeatedly appears in the top-ranked results (*Supplementary material Table 9*) as an sc-eQTL with similar effect size and direction for all cell types (mean effect size: -1.30 (-1.25 - max -1.36, mean p-value: $4.07\times10^{-80}$ ($9.37\times10^{115}$-$2.44\times10^{-79}$)), this is a demonstration of an eQTL effect that is shared across cell types (Figure 5). Different SNPs are related to the most significant eQTL for *XRRA1* in different cell types, however as these SNPs (*rs10751241*, *rs4944963*, *rs2298746* and *rs7102619*) are linkage with each other ($R^2$ 0.9776, p <0.0001 for all possible SNP-SNP combinations) these sc-eQTLs all relate to the same effect.

| Cell type | # sc-eQTLs | # sc-eSNPs | # sc-eGenes |
|---|---|---|---|
| Astrocyte | 799,682 | 600,552 | 2,511 |
| Endothelial cells | 16,095 | 15,161 | 51 |
| Excitatory neurons | 2,064,362 | 1,305,685 | 5,722 |
| Inhibitory neurons | 1,071,052 | 787,279 | 3,089 |
| Microglia | 207,966 | 181,847 | 666 |
| Oligodendrocytes | 776,375 | 581,892 | 2,416 |
| OPCs | 276,405 | 239,148 | 910 |

**Table 3** Number of significant sc-eQTLs, sc-eSNPs and sc-eGenes per cell type

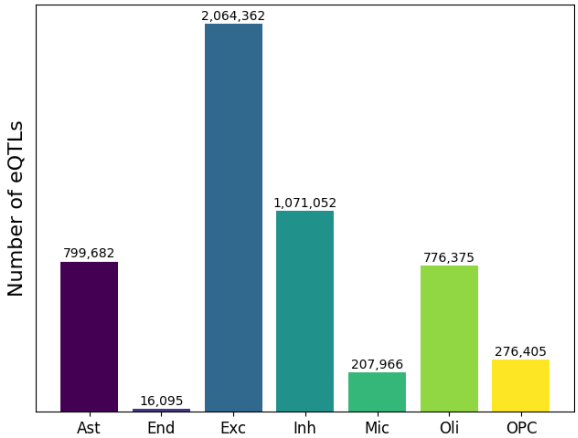



**Figure 4** Number of sc-eQTLs per cell type showing that sc-eQTLs is higher in the most abundant cell type: excitatory neurons

## sc-eQTLs for AD-associated SNPs point to cell type specific gene regulation in AD

To investigate possible sc-eQTLs of relevance to AD, we focused on the 86 SNPs identified in the most recent GWAS of

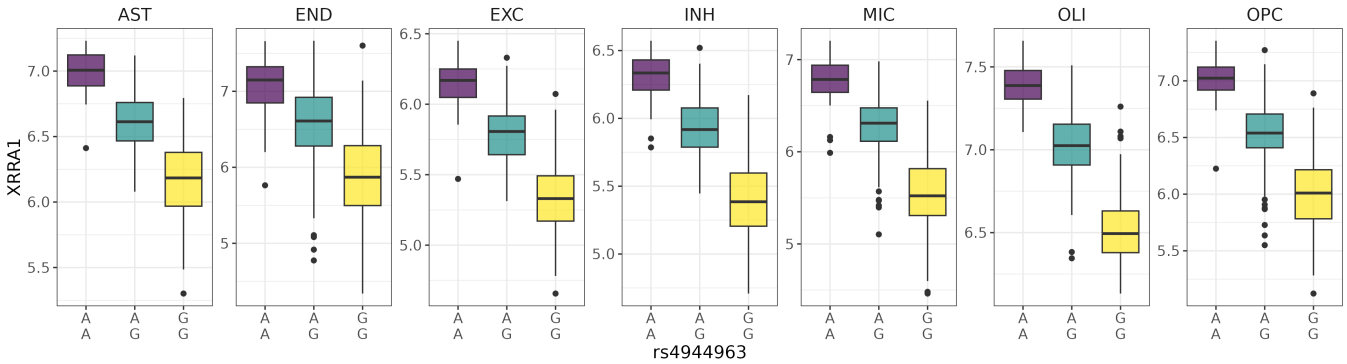



**Figure 5** eQTL effect of rs4944963 on XRRA1 expression ($log_2$ CPM) for all cell types demonstrating which shows the sc-eQTL effect size is similar for all cell types considered

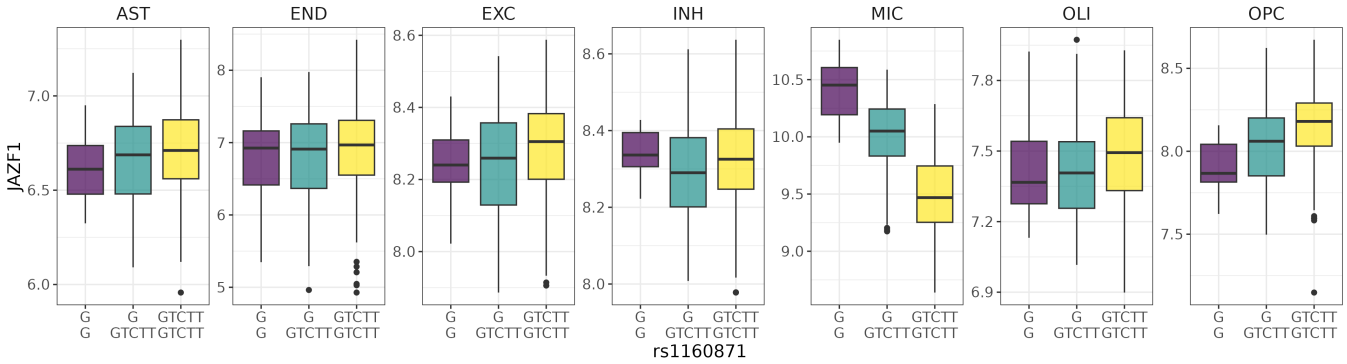



**Figure 6** eQTL effect of rs1160871 on JAZF1 expression ($log_2$ CPM) for all cell types demonstrating opposite effect size in Microglia versus OPCs

AD (Bellenguez *et al.* (2022)) of which 57 SNPs were included in our analysis based on their minor allele frequency (MAF $\geq$ 0.5). Of these, 41 (71.9%) were present in the significant sc-eQTLs (FDR $\leq$ 0.05) for at least one gene in at least one cell type. In total, 132 (of which 85 unique) sc-eQTLs (SNP-gene pairs) were related to AD SNPs, of which 31 were present in more than one cell type. The majority of sc-eQTLs were found in excitatory neurons (Figure 7), this does however not imply that AD effects are more prominent in excitatory neurons, but can be explained by the fact that excitatory neurons have a higher relative abundance as well as higher RNA content, resulting in higher (pseudobulk) expression levels leading to increased likelihood of sc-eQTL discovery (Fujita *et al.* 2022). Looking at the relation between the number of sc-eQTLs found per cell type and the sc-eQTLs identified relevant to the AD GWAS loci, we see a relative increase in sc-eQTLs identified in microglia, the immune cells of the nervous system, suggestive of their differential involvement in AD (Hansen *et al.* 2018).

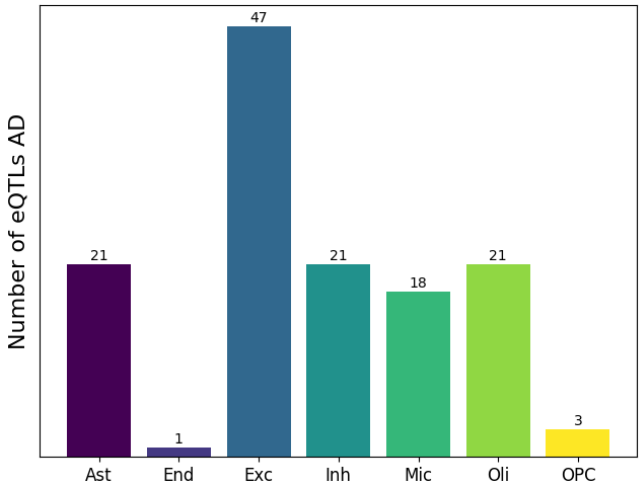



**Figure 7** Number of sc-eQTLs for AD GWAS variants per cell type

On investigation of AD-associated sc-eQTLs that were shared across multiple cell types, we found that most of the shared sc-eQTLs were shared between inhibitory- and excitatory neurons (n = 16, Figure 8). For 16 variants, significant sc-eQTLs were only found in a single cell type. In astrocytes, these are *rs76928645-EGFR* (effect size: 0.24, p=$3.79\times10^{-25}$) and *rs785129-HS3ST5* (effect size: −0.13, p=$2.13\times10^{-7}$), both *EGFR* and *HS3ST5* are known to be over-expressed in AD (Ferreira *et al.* 2022; Romano **and** Bucci 2020). In microglia, this holds for *rs10933431-INPP5D* (effect size -0.05, p= $2.01\times10^{-4}$), *rs12590654-RIN3* (effect size -0.13, p= $5.01\times10^{-10}$), *rs6014724-CASS4* (effect size -0.39 p= $5.37\times10^{-5}$), *rs6584063-BLNK* (effect size -0.28, p =$6.80\times10^{-14}$), *rs6733839-BIN1* (effect size-0.13, p= $6.28\times10^{-7}$ ) and *rs73223431-PTK2B* (effect size: -0.16, p= $1.85\times10^{-23}$), all genes involved have been associated with AD pathology and are also the nearest protein-coding of the sc-eSNP (Tsai *et al.* 2021; Shen *et al.* 2020; Beck *et al.* 2014; Holler *et al.* 2014; Giralt *et al.* 2018). In oligodendro-

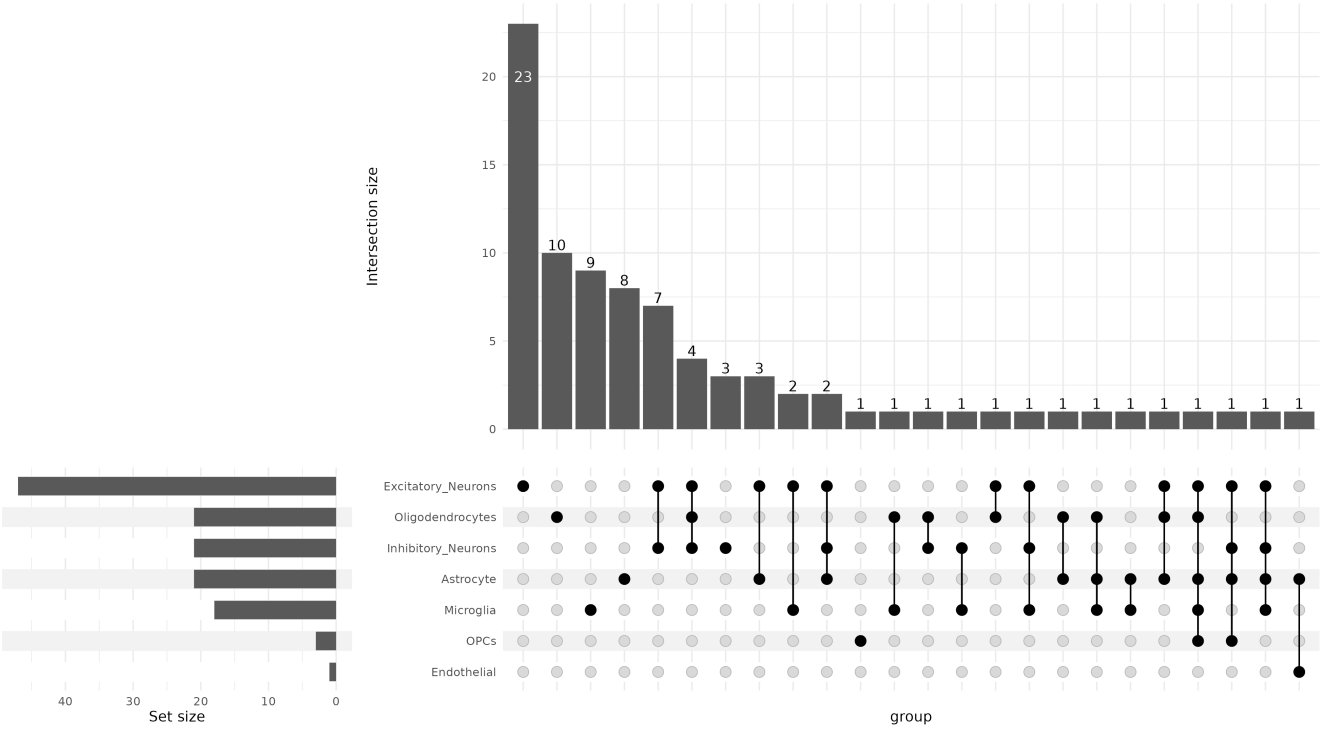



**Figure 8** Upset plot showing sc-eQTLs relating to AD GWAS variants are shared across cell types showing most sc-eQTL shared are shared between excitatory and inhibitory neurons (n=16)

cytes, the only AD-associated cell type specific sc-eQTL identified is *rs456074-MAF* (effect size -0.27, p= $1.41\times10^{-20}$, no evidence was found for *MAF* (MAF bZIP transcription factor) involvement in AD. In excitatory neurons, these sc-eQTLs are *rs12592898-CTSH* (effect size 0.14, p=$1.05\times10^{-5}$), *rs1800978-FSD1L* (effect size -0.02, p=$7.83\times10^{-4}$), *rs3848143-CIAO2A* (effect size -0.03, p=$7.90\times10^{-4}$) and *rs6489896-RITA1* (effect size 0.10, p=$1.70\times10^{-6}$) which are mostly the nearest-protein coding genes of the variants, only for *CTSH* we found evidence of its involvement in AD beyond the GWAS association (Li *et al.* 2023b). In inhibitory neurons, this holds for rs6846529-AC006230.1 (HS3ST1 locus) (effect size 0.06, p=$1.46\times10^{-20}$), which is associated with increased expression in AD (Wang *et al.* 2023b).

***Effect size and direction differences of eQTLs in different cell types*** It should be noted that when an sc-eQTL is shared between multiple cell types, most often it does have the same effect direction, however effect size and direction are not necessarily the same pointing towards different effect of variance on gene expression between different cell types. For example, *rs1160871* (chr7:28129126:G:GTCTT), a regulatory variant located in one of the larger introns of *JAZF1*, has opposite effect size in microglia and OPCs (Figure 6) pointing towards differential regulation of *JAZF1* expression. The difference in effect direction can be possibly explained differences in intron-mediated enhancement, which is a mechanism that is poorly understood (Dwyer *et al.* 2021) or could possibly be the result of differences in epigenetic regulation in microglia compared to OPCs (Xiong *et al.* 2023).

## Validation with eQTLs obtained from bulk RNA-sequencing data

We compared the results of our sc-eQTL analysis to Metabrain de Klein *et al.* (2023), a meta-analysis of eQTL studies that used bulk RNA sequencing data obtained from tissue of different regions in the brain. We assessed the presence of Metabrain cortex (EUR) eQTLs, significant at Q-value <0.05 in our results. Overall, 8,687 out of the 18,396 MetaBrain eQTLs were an sc-eQTL for at least one of the cell types analyzed (Table 4). The percentage of eQTLs shared between MetaBrain and our analysis is higher for the *larger* cell types, which can be explained by the fact that these cells are also more prevalent in bulk samples and are therefore more represented in the results. The number of overlapping findings correlates with the number of cells present per cell type in our analysis (Table 5), confirming this effect (Pearson correlation: 0.8099, p=0.02724). Further, we note that assessing correspondence on eQTL (SNP-gene) level is sensitive to the presence of SNPs in both analyses, where a comparison on eGene level might have shown a stronger overlap between results. Further, only single nucleotide polymorphisms were considered in the Metabrain meta-analysis, where we also consider small structural variants which could possibly explain the lower replication rate. Finally, differences could result from the fact that our sc-eQTL analysis was performed with DLPFC tissue which is a subregion of the cortex considered in Metabrain.

| Cell type | Overlap | Percentage |
|---|---|---|
| Astrocyte | 1,404 | 7.63% |
| Endothelial | 37 | 0.20% |
| Excitatory neurons | 3,334 | 18.12% |
| Inhibitory neurons | 1,802 | 9.80% |
| Microglia | 348 | 1.89% |
| Oligodendrocytes | 1,300 | 7.07% |
| OPCs | 462 | 2.51% |

**Table 4** Number of sc-eQTLs from the replication study that are present in MetaBrain Cortex CEUR significant eQTLs and the percentage of Metabrain eQTLs that were also present as an sc-eQTL in our analysis

## AD-associated cell type specific co-eQTLs

We found 6,878 cell type-specific co-eQTLs relating to 18 variants associated with cell type specific co-expression changes. Of these, 104 co-eQTLs were shared between cell types, 18 were shared between microglia and OPCs, all involving *rs1160871-JAZF1*, and 86 were shared between excitatory and inhibitory neurons, all involving *rs13237518* and mostly involving *CTNNB1* (54/86). First, we zoom in on the co-eQTLs identified in excitatory neurons, and afterwards extend to co-eQTLs identified in the remaining 6 major cell types. We analyzed the resulting co-eQTLs in two ways: by assessing what changes in co-expression are captured by the method and by looking into functional and biological annotations of the SNPs and genes involved.

### AD associated co-eQTLs in excitatory neurons

In excitatory neurons, 6,587 significant co-eQTLs (variant-gene-gene combinations) were found after correcting for multiple testing (see *Methods*). The top 10 co-eQTLs (*Supplementary Material* Table 10) are all related to *rs13237518* (located in gene *TMEM106B*) and are all associated with significant sc-eQTL or differential correlation effect.

***Top co-eQTL PLEKHA1 - SLC25A4 (rs13237518)*** The most significant result was observed for the co-expression of *PLEKHA1* (Pleckstrin Homology Domain Containing A1) and *SLC25A4* (Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4) for variant *rs13237518* between genotype groups AA and CC. We observed a distance (JSD) between the groups of 0.172 ($p=7.76\times10^{-8}$). The result is driven by sc-eQTL effect of *rs13237518* on the expression of *SLC25A4* with increased expression in the *AA* versus the *CC* genotype as can be seen in the co-expression distribution obtained by kernel density estimation in Figure 9. Further, there is a change in the variance of expression of *PLEKHA1* between the genotypes. There is no differential correlation between the genotype, Pearson correlation of the expression of *PLEKHA1* and *SLC25A4* is -0.162 in *AA* and -0.059 in *CC* respectively, while the overall correlation of expression of the gene pair (considering all genotypes) is 0.059. The gene pair has no significant GO term enrichment for involvement in a common biological process and neither gene is a known transcription factor.

***Differential correlation of CTNNB1 - IRS2 (rs13237518)*** We observed a significant differential correlation, ranked 3 in the top co-eQTLs, that was not driven by a significant eQTL effect, between genes *CTNNB1* (Catenin beta-1) and *IRS2* (Insulin receptor substrate 2) in relation to *rs13237518* (between genotypes AA and CC). The co-expression probability densities obtained through kernel density estimation per genotype are given in Figure 10. The distance (JSD) observed between the groups is 0.158 ($p=7.76\times10^{-8}$). Even though there is no significant eQTL effect for both genes, a slight increase in gene expression is notable for both genes in genotype AA with respect to genotype CC. This result is among the top 5% differentially correlated results, meaning that there is a significant change in correlation between the expression of genes across the SNP genotypes. The Pearson correlation of the expression of *CTNNB1* and *IRS2* is 0.4538 for the AA genotype and -0.0696 for the CC genotype. Both genes are related to the GO terms *phosphatase binding* (GO:0019902) and *protein phosphatase binding* (GO:0019903), but there is no known protein-protein interaction according to STRING and the gene *TMEM106B* in which the variant is located is not a known transcription factor that could explain the co-expression changes observed (Szklarczyk *et al.* 2022; Lambert *et al.* 2018).

### TMEM106B related variant rs13237518 involved in widespread expression changes in excitatory neurons

At variant level, we note that *rs13237518* (chr7:12229967:C:A) is involved in most co-eQTLs (6,352/6,587), of which 3,012 involve a significant eQTL effect and 601 involve a differential correlation (*Supplementary material* Table 16). Variant *rs13237518* is located in the intronic region of *TMEM106B*, but *rs13237518-TMEM106B* is not a significant eQTL. For *rs13237518*, the A allele is associated with protection against AD (Bellenguez *et al.* 2022) and which was also found to be enriched in cognitively healthy centenarians (Tesi *et al.* 2023). *TMEM106B* (transmembrane protein 106B) is located in membranes of endosomes and lysosomes, which are involved in sorting and degradation of cellular waste respectively (Lang *et al.* 2012). *TMEM106B* is known for association with multiple types of neurodegenerative disease (AD, Frontotemporal lobar degeneration (FTLD) and limbic-predominant age-related TAR DNA binding protein 43 (TDP-43) encephalopathy) (Jiao *et al.* 2023).

Notably, 25 of the considered genes are a significant eQTL for *rs13237518* (*KRAS, SLC25A4, CYCS, AKT2, RTN4, PSMD2, EIF2AK2, RTN3, PSMD12, ATP2A2, NDUFA5, UQCRB, CALM2, WIPI2, VDAC1, CALM1, CDK5R1, NDUFS1, UQCR10, EPDR1, MAPK9, TUBA1A, UQCRH, CTSB, PIK3R1*), all of which are *trans*-eQTLs, as none of the genes are located in the *cis*-region of *rs13237518*, most of the genes are not even located on the

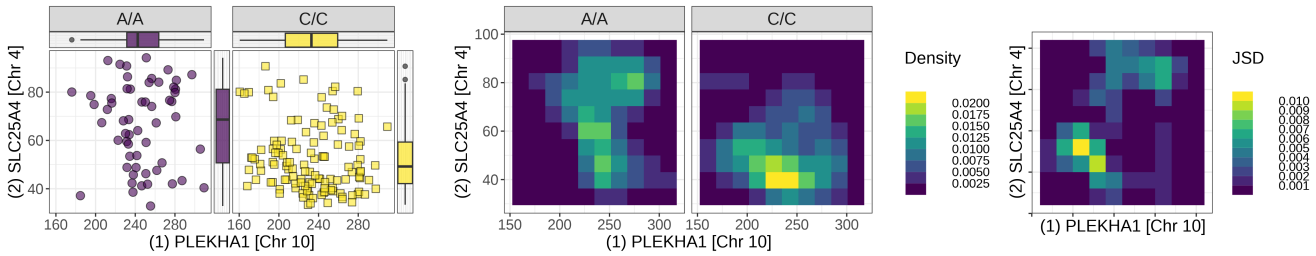



**Figure 9** Top co-eQTL PLEKHA1 in excitatory neurons: PLEKHA1 and SLC25A4 for rs13237518 (chr7)
**Left:** co-expression distribution, every point represents the expression (CPM) of PLEKHA1 and SLC25A4 for an individual which is assigned to genotype AA (left) or CC (right) **Middle:** Probability density estimate obtained through kernel density estimation per genotype **Right:** Local JSD showing large local contribution introduced by SLC25A4 eQTL effect

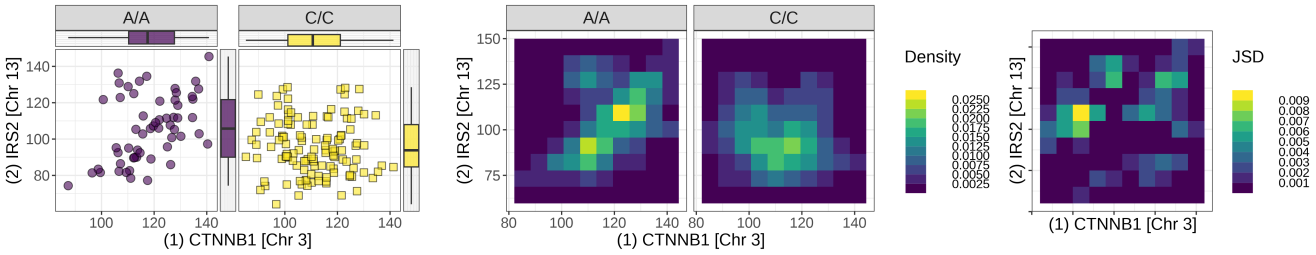



**Figure 10** CTNNB1 and IRS2 for rs13237518 (chr7) in excitatory neurons demonstrate a co-eQTL driven by differential correlation
**Left:** co-expression distribution, every point represents the expression (CPM) of CTNNB1 and IRS2 for an individual assigned to genotype AA (left) or CC (right) **Middle:** Probability density estimate obtained through kernel density estimation per genotype **Right:** Local JSD showing large local contribution by lack of expression local expression density in genotype AA

same chromosome as *rs13237518*. No significant GO term enrichment for these genes was found. For all but AKT2, the eQTL effect is an increase in expression in genotype AA compared to CC. A similar finding is reported in the brain eQTL meta-analysis by de Klein *et al.* (2023), where 85% of the trans-eQTLs found relate to the 7p21.3 locus (TMEM106B), however none of these trans-eQTLs were significant after including 100 expression PCs as a covariate. These 25 genes are also strongly represented in the significant co-eQTLs, the average number of co-eQTLs per gene is 129.08 (75-160).

The following 6 genes are a significant co-eQTL in every gene pair these were tested in for *rs13237518*: *AKT2*, *EIF2AK2*, *KIF5B*, *KRAS*, *SLC25A4*, and *UQCRB*, and all, but KIF5B, have significant eQTL effect for *rs13237518* (Figure 11). *SLC25A4* and *KRAS* are leading with co-eQTLs with large JSD. We looked into the gene expression distribution between the genotypes (Figure 11) to find possible pointers that explain this effect. For *SLC25A4* and *KRAS*, we see a strong peak (lower expression variance) for genotype CC, this effect is also present for the majority of 25 eQTL genes mentioned earlier. This peak is reflected in co-expression space as well and contributes strongly to the distance between co-expression distributions. The top co-eQTL (*PLEKHA1* - *SLC25A4*) (Figure 9) demonstrates the consequences of this effect on the distance between co-expression distributions, the density peak for *SLC25A4* in genotype CC gives rise to the JSD in two ways: the peak density causes a high local JSD and the lack of density increase the JSD as well.

We found the first principal component (PC1) of gene expression in excitatory neurons to be correlated with *rs13237518* (Pearson correlation: -0.2475, $p=1.07\times10^{-6}$), the loading of PC1 is higher for most genes marked as an eQTL, which also hints towards possible widespread variant-driven expression disruption. PC1 is also correlated with age of death (Pearson correlation: -0.1984, $p=10^{-4}$) and cognitive diagnosis (Pearson correlation: -0.1554, $p=2.42\times10^{-3}$), but less strongly than with the *rs13237518* genotype. As we look at a limited subset of genes that show mostly upregulation with respect to rs13237518 (genotype AA), it is likely worthwhile to extend the analysis to the full set of genes for a complete picture of the rs13237518 associated expression changes.

The fact that a SNP was identified in a GWAS as a significant disease-related locus, does not necessarily imply that the disease mechanism involves direct functional consequences of the SNP. It is also possible that the GWAS signal is related to another variant in linkage with the SNP, which could be another SNP, but also longer structural variant (SV) such as a variable number tandem repeat (VNTR) or a transposable element. Structural variants (≥ 50bp) are known to be associated with changes in gene expression (Scott *et al.* 2021) and can be more confidently called with the recent development of long-read sequencing techniques (Balachandran and Beck 2020). Therefore, we explored known associations for variants in linkage with *rs13237518*. Variant *rs13237518* was found to be in strong linkage ($R^2$ = 0.90) with a structural variant, a 323bp Alu deletion (chr7:12242077-12242399) in exon 8 of *TMEM106B* (Wang *et al.* 2023a). Looking at other consequences of the variant in literature, we found

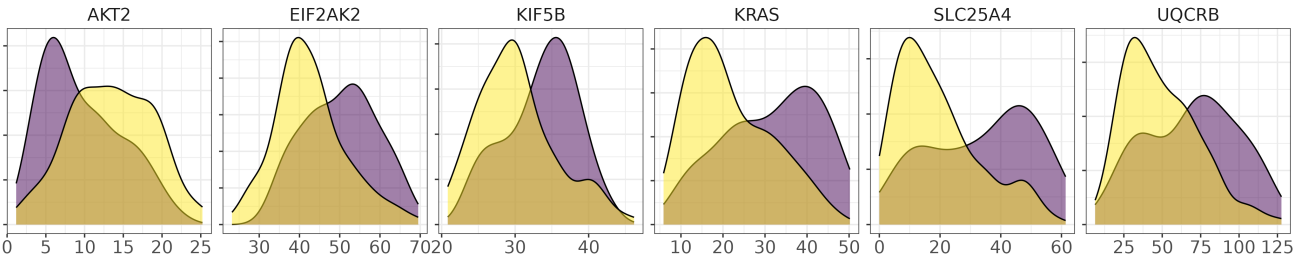



**Figure 11** Distribution of gene expression (CPM) relating to rs13237518 genotype AA and CC of all genes that are a co-eQTL in excitatory neurons for all other genes these were tested with. All genes but KIF5B are an eQTL for rs13237518, the distribution have high kurtosis and in combination with a slight change in expression may drive these effects. KRAS and SLC25A4 are most prominent in the top result, this is driven by the concentrated density peak in genotype CC

.

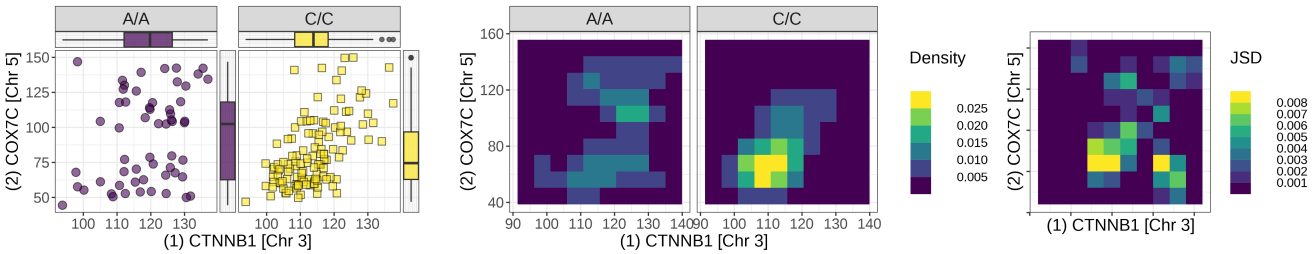



**Figure 12** Top co-eQTL result CTNNB1 and COX7C for rs13237518 (chr7) in inhibitory neurons
**Left:** co-expression distribution, every point represents the expression (CPM) of CTNNB1 and COX7C for an individual which is assigned to genotype AA (left) or CC (right) **Middle:** Probability density estimate obtained through kernel density estimation per genotype **Right:** Point-wise contribution of Jensen-Shannon divergence between genotype AA and CC for CTNNB1 and COX7C for rs13237518 (chr7) for inhibitory neurons

that Fujita *et al.* (2022) reports that *rs5011436* (chr7:12229132), which is in linkage ($R^2$ 0.9799 - p=<0.0001) with *rs13237518*, is a fraction eQTL for sub-cell type *Exc.2* meaning the variant is linked to higher relative abundance of an identified sub-cluster of excitatory neurons. The variant is also in linkage with *rs1990621* ($R^2$= 0.9601, p=<0.0001), a *TMEM106B* variant protective against Frontotemporal lobar degeneration (FTLD) which is also associated with an increased proportion of neurons in a study using bulk RNA-sequencing (Li *et al.* 2020). These associations of variants in linkage with *rs13237518* are also suggestive of higher-order disruptions in gene expression.

## AD-associated co-eQTLs in other major brain cell types

We identified 195 co-eQTLs (*Supplementary Material* Table 16) in the remaining cell types, of which 117 co-eQTLs in microglia, 106 co-eQTLs in inhibitory neurons, 41 co-eQTLs in OPCs, 17 co-eQTLs in endothelial cells and 10 co-eQTLs in oligodendrocytes. No significant co-eQTLs were identified in astrocytes. The variants present in significant co-eQTLs per cell type are provided in *Supplementary Material* (Table 16) and the top 10 results per cell type are provided in *Supplementary Material* (Inhibitory neurons: Table 11, Microglia: Table 12, Endothelial cells: Table 13, Oligodendrocytes: Table 14, OPCs: Table 15).

***In inhibitory neurons, rs13237518 is also key in co-eQTLs with lower eQTL driven effects*** Similar to excitatory neurons, *rs13237518* is a key locus in the significant co-eQTLs for inhibitory neurons with 95/106 co-eQTLs being related to the variant. A striking difference is observed when comparing the co-eQTLs between the neuronal sub-types: in excitatory neurons, the co-eQTLs are strongly driven by eQTL effect on either of the genes involved, whereas this is not the case for inhibitory neurons with only 5 out of 106 co-eQTLs being subject to significant eQTL effect. The most prominent co-eQTL genes are *CTNNB1* (catenin beta 1) (59/106 co-eQTLs), *PPP3CC* (protein phosphatase 3 catalytic subunit gamma) (13/106 co-eQTLs) and *COX7C* (cytochrome c oxidase subunit 7C) (10/106 co-eQTLs), the other genes were involved in 5 or less co-eQTLs. The 3 genes are not significantly enriched for a GO term that points towards involvement in a similar process. *CTNNB1* is a transcription factor, however it was not identified as a transcription factor that targets genes in the subset of genes tested (Han *et al.* 2018). The only gene that qualified as an eQTL for *rs13237518* is *RTN4* (reticulon 4) which is involved in only 5 co-eQTLs, looking at the expression distribution of RTN4 (Figure 13: Bottom right) we see that the expression for genotype AA is far less peaked compared to genotype CC and this more homogeneous expression distribution leads to less extreme distances as the distributions overlap at the point where there is a peak in genotype AA resulting in less pronounced changes in co-expression.

The top co-eQTL involves two genes that are highly prevalent in the significant co-eQTLs: *CTNNB1* and *COX7C* relating to *rs13237518* (Figure 12), with JSD of 0.141 ($p=2.69\times10^{-6}$). Looking into the changes in co-expression distribution, we observe that for *COX7C*, there is a notable increase in variance

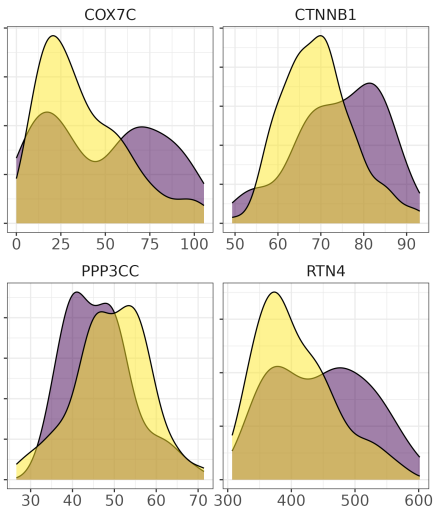



**Figure 13** Gene expression (CPM) distribution in inhibitory neurons relating to rs13237518 genotype AA and CC for top 3 genes in co-eQTLs: COX7C, CTNNB1 and PPP3CC and the only significant eQTL in for rs13237518: RTN4

in expression in genotype *AA* (var 26.29) with respect to *CC* (var 598.02) where the lower variance in CC creates a local peak in the co-expression distribution which is driving the increase in JSD. The difference in expression correlation between the genotypes is also significant (Pearson correlation: 0.285 (AA)- 0.595 (CC) - 0.410 (all genotypes) - $|\Delta|$ correlation (AA-CC): 0.311).

***JAZF1* related co-eQTLs in microglia and OPCs point to method sensitivity to strong eQTL effect** Three SNPs, *rs1160871* (38/117 co-eQTLs), rs12590654 (35/117 co-eQTLs) and *rs10437655* (33/117 co-eQTLs), are involved in the majority of significant co-eQTLs in microglia (see *Supplementary material* Table 16). For each variant, most co-eQTLs involve a single gene and the SNP-gene combinations are a co-eQTL for every other gene it is tested with.

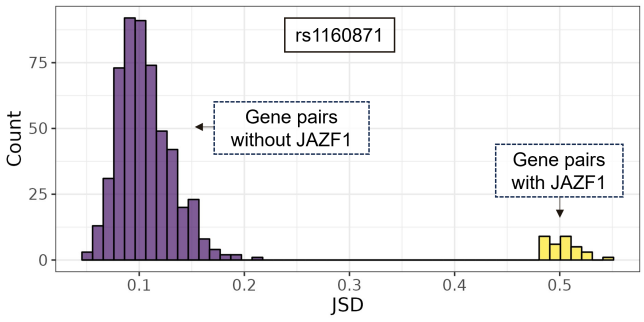



**Figure 14** Histogram of JSD distribution for gene pairs tested for rs1160871 in microglia, showing the difference in JSD for gene pairs without JAZF1 compared to gene pairs with JAZF1

For *rs1160871* (chr7:28129126:GTCTT:G), a regulatory variant located in one of the larger introns of *JAZF1* (JAZF zinc finger 1), most co-eQTLS (33/38) directly involved *JAZF1*. The G allele is associated with lower odds of developing AD. (Bellenguez *et al.* 2022). *JAZF1* is a glucose-production-promoting transcriptional repressor, a known transcription factor and was recently identified as a locus for epigenetic alterations in microglia (Zhou *et al.* 2020; Xiong *et al.* 2023). The variant-gene combination is a significant eQTL and a co-eQTL for every gene it is tested with. *JAZF1* is prevalent in the top-ranked co-eQTLs (see *Supplementary material* Table 12) and stands out by co-eQTLs with large JSD (mean JSD 0.5019 (co-eQTLs with *JAZF1*) mean JSD 0.1286 (co-eQTLs without JAZF1) - Figure 14).

Looking at the expression of *JAZF1* for *rs1160871* between genotypes G/G and GTCTT/GTCTT (Figure 15 - Left), we see there is hardly any overlap in expression density between the genotypes. Taking the top co-eQTL *rs1160871-JAZF1-NCK2* as an example, we see the impact of the strong eQTL-effect of *rs1160871-JAZF1* on the co-eQTL: the separation of JAZF1 expression between the genotypes causes large local contributions to the JSD on both sides of the co-expression distribution (Figure 16). These local distances exist regardless of the expression of the second gene involved in the co-eQTL and therefore do not effectively capture changes in co-expression. This demonstrates a limitation of our methodology: unwanted inflation of results happens for co-eQTLs where the expression distributions for a SNP-gene combination have little overlap. We conclude that the large distances between co-expression distributions are driven by a methodological limitation and do not necessarily reveal a differential co-expression effect, even though this does not rule out the possibility of existence of true co-eQTLs for SNP-gene-gene combinations involving *rs1160871-JAZF1*.

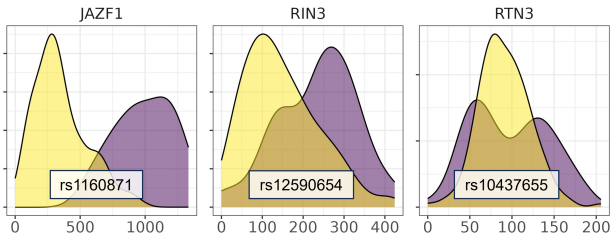



**Figure 15** Density plot of gene expression (CPM - x-axis) in microglia for 3 variant-gene combinations that are highly prevalent in the top co-eQTLs. **Left:** rs1160871-JAZF1 between genotypes GTCTT/GTCTT and GG showing strong eQTL effect and little overlap in expression distribution between the genotypes **Middle:** rs12590654-RIN3 between genotypes AA and GG which has significant eQTL effect **Right:** rs10437655-RTN3 between genotypes AA and AG demonstrating there is no eQTL effect

Previously, we identified *rs1160871-JAZF1* as an eQTL with opposite effect in microglia compared to OPCs. This effect is also showing in the co-eQTL results for OPCs: all co-eQTLs involve *rs1160871-JAZF1* which is a co-eQTL for 41/44 genes it is tested with.

*RIN3* (Ras and Rab interactor 3 - chr 14) is present in all co-eQTLs relating to *rs12590654* (chr14:92472511:G:A), which is marked as a significant eQTL. The variant is located in the intronic region of SLC24A4 (solute carrier family 24 member 4

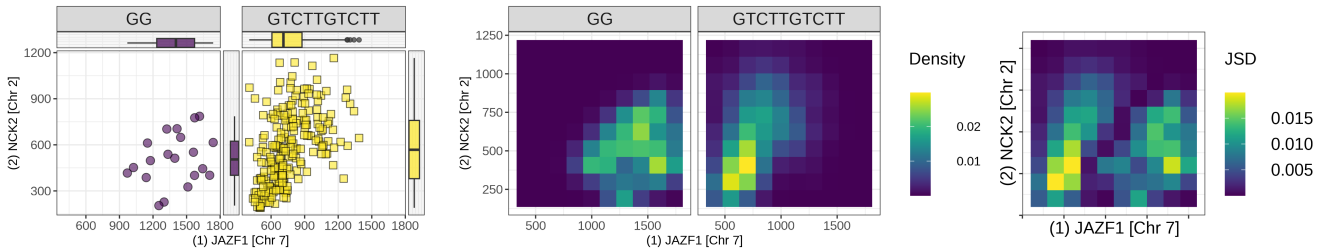



**Figure 16** Microglia: paired expression of JAZF1 and NCK2 with respect to rs13237518
**Left:** co-expression distribution, every point represents the expression (CPM) of JAZF1 and NCK2 for an individual which is assigned to genotype GTCTT/GTCTT (left) or G/G (right) **Middle:** Probability density estimate obtained through kernel density estimation per genotype **Right:** Pointwise JSD between genotype G/G and GTCTT/GTCTT demonstrating the large distances created locally

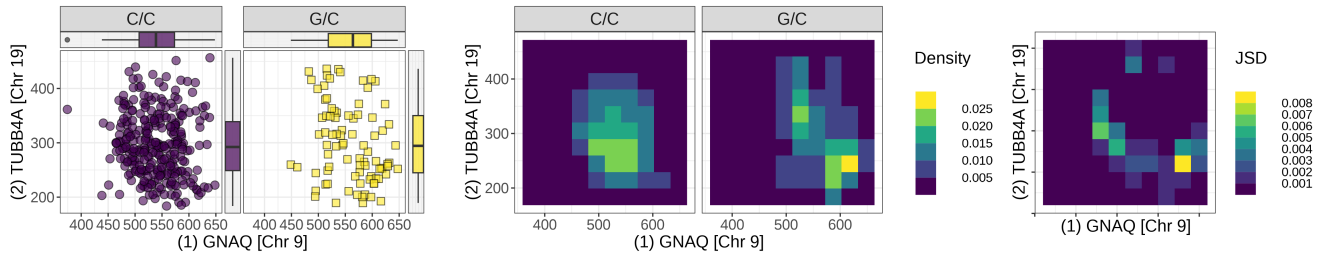



**Figure 17** Oligodendrocytes: paired expression of GNAQ and TUBB4A with respect to rs1800978
**Left:** co-expression distribution, every point represents the expression (CPM) of GNAQ and TUBB4A for an individual which is assigned to genotype C/C(left) or C/G (right) **Middle:** Probability density estimate obtained through kernel density estimation per genotype **Right:** Pointwise JSD between genotype C/C (left) or C/G demonstrating the large local distance created by a local peak in co-expression in density in genotype C/C where

(sodium/potassium/calcium exchanger)). RIN3 and SLC24A4 are adjacent genes in the genome with existing evidence for interaction in STRING (Szklarczyk *et al.* 2022). The A allele of rs12590654 is associated with lower odds of developing AD (Bellenguez *et al.* 2022). Expression of RIN3 is specific to microglia and upregulation of RIN3 was shown to compromise endosomal function in a mouse study (Wang 2021; Shen *et al.* 2020). This corresponds to the increased RIN3 expression we observe for genotype GG with respect to AA (Figure 15: middle). Even though the eQTL effect is less extreme compared to rs1160871-JAZF1, we conclude the co-eQTLs are discovered because of the eQTL effect of rs12590654-RIN3.

For *rs10437655* (chr11:47370397:G:A), an intronic variant in known transcription factor *SPI1* (Spi-1 proto-oncogene) with elevated expression in microglia, SPI1 did not meet the threshold for inclusion in the analysis. All rs10437655-related co-eQTLs found involve *RTN3* (reticulon 3), reticulons are proteins that localize in the endoplasmatic reticulum known for involvement in neural regeneration and are associated with amyloid deposition and neurodegenerative disease (Pradhan and Das 2021; Kulczyńska-Przybik *et al.* 2021). Unlike the other key SNPs in microglia co-eQTLs, *rs10437655-RTN3* is not an eQTL. A quick glance at the expression of RTN3 between genotype AA and AG (Figure 15 : right) shows why the SNP-gene is involved in many co-eQTLs: the bimodal-like shape distribution in genotype AA has a local lower density where the high peak of genotype AG is situated, this difference gives rise to larger distances between distributions.

***Co-eQTLs in oligodendrocytes revolve around rs1800978-GNAQ*** In oligodendrocytes, *rs1800978* (chr9:104903697:C:G) is the main locus present in 6/10 significant co-eQTLs, which is a SNP located in the 5' untranslated region (UTR) of ABCA1 (ATP binding cassette subfamily A member 1), 5' UTRs are known from their involvement in the regulation of translation of RNA to proteins. *GNAQ* (Guanine nucleotide-binding protein G(q) subunit alpha) is present in 5/6 co-eQTLs relating to *rs1800978*, which is not an eQTL. We note that the JSD for co-eQTLs involving *GNAQ* is low (mean 0.0709 (0.0661-0.0838)) signalling less extreme changes in the co-expression distribution. Looking at the co-expression distribution between genotype CC and GC for the top result relating to rs1800978 (Figure 17) for gene pair *GNAQ-TUBB4A* with JSD 0.0838 ($p = 2.46 \times 10^{-6}$), we note the distributions are shifted towards the right because of an apparent outlier with low expression for *GNAQ* (genotype CC). Further, wed see that the largest contribution to the JSD comes from a region with relatively low *TUBB4A* expression and high *GNAQ* expression for genotype GC for which there is no local density in genotype CC. Recalculation after removal of the outlier (Figure 18) resulted in a JSD of 0.0842 ($\Delta$-0.0004), showing an example of possible robustness of the method when outliers are present, even though robustness to outliers of the method was not assessed on full scale and is recommended to be done in future work.

***Most co-eQTLs in endothelial cells involve CLU*** For endothelial cells, the core co-eQTL genetic variant is *rs11787077*

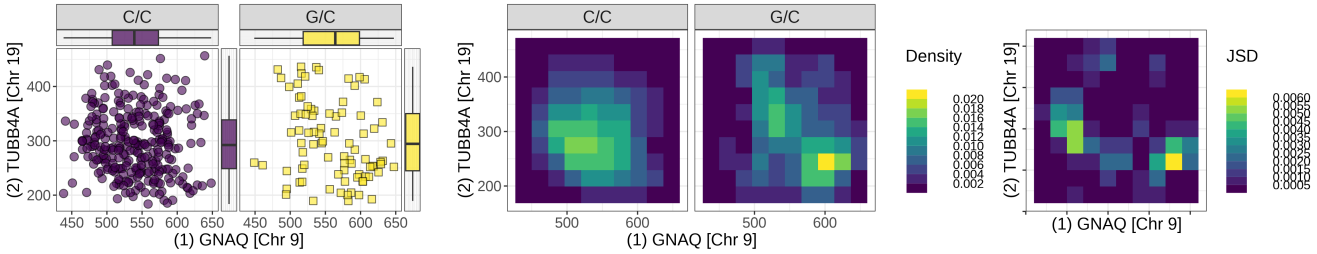



**Figure 18** Oligodendrocytes: paired expression of GNAQ and TUBB4A with respect to rs1800978 after removal of outlier

(chr8:27607795:T:C) which is present in 14 out of 17 significant results. The T allele of *rs11787077* is associated with decreased odds of developing AD (Bellenguez *et al.* 2022). The results are all driven by the strong eQTL effect of *rs11787077* on the expression of CLU (clusterin - apolipoprotein J), CLU is involved in amyloid metabolism and upregulated in AD patients (Liu *et al.* 2022).

## Discussion

We presented a cell type specific eQTL and co-eQTL analysis using genotyping and sn-RNA seq data of the dorsolateral prefrontal cortex of 379 individuals from the ROSMAP cohort, linking genetic variants to changes in gene (co-)expression.

### sc-eQTL analysis

In our sc-eQTL analysis, we found 3,337,065 sc-eQTLs, linking 1,882,645 sc-eSNPs to changes in the expression of 8,057 sc-eGenes in 7 major cell types and replicated our results in an independent dataset of eQTLs obtained from bulk RNA-sequencing data which shows partial overlap per cell type correlating to the relative abundance of cell types included in our analysis. We see that the number of eQTL discoveries per cell type is correlated with the number of cells included, which can be attributed to the fact that inclusion of more cells increases the confidence of the analysis and therefore results in higher significance of results.

Our results support that cell-type specific analysis allows for the discovery of eQTL effects that are obscured in bulk-based analyses. Of the 8,057 sc-eGenes, 4,024 (50%) were found to be an sc-eGene in a single cell type only. For 16 AD GWAS SNPs (Bellenguez *et al.* 2022), we found the SNP to be an sc-eQTL in a single cell type only: astrocytes (2), excitatory neurons (6), inhibitory neurons (1), microglia (6) and oligodendrocytes (1).

***Strengths and limitations*** A major strength of our analysis is the number of individuals included in the dataset that was used, which gives our analysis increased eQTL discovery power compared to earlier efforts. Further, the inclusion of small SVs in the analysis allowed for the discovery of eQTL effects that are obscured in analyses that focus solely on SNPs. A limitation of our analysis is that we focused on eQTL discovery in major cell types, while Fujita *et al.* (2022) shows that assessment of eQTLs on cell sub-type level reveals sc-eGenes that were not found on major cell type level.

### co-eQTL analysis

In our co-eQTL analysis, we found 6,878 cell type-specific co-eQTLs (SNP-gene-gene combinations) relating to 18 variants associated with cell type-specific co-expression changes, relating to a subset of AD variants and genes.

Most strongly, we found many co-eQTLs in excitatory neurons in relation to *rs13237518*, located in *TMEM106B*, a transmembrane protein involved in lysosomal function known for its association with AD and neurodegenerative processes (Salazar *et al.* 2023; Jiao *et al.* 2023). These co-eQTLs involve a subset of 25 genes, of which the majority is upregulated in the AA genotype which is associated with protection against AD (Bellenguez *et al.* 2022). The subset of genes is not enriched for involvement in a certain biological process. As these results are suggestive of higher-level disruptions of gene expression in relation to the variant, we recommend further analysis of expression changes that considers the full transcriptome to gain further insight in the role of the variant in neurodegenerative disease.

***Strength and limitations*** To our knowledge, this work is the first that assesses differences in co-expression associated with genetic variants using a non-parametric method that looks beyond differences in correlation between expression of genes in relation to genetic variants. Our method is effective in doing so as we find the majority of significant co-eQTLs driven by changes other than a difference in expression correlation between genotypes (Figure 20).

The results have to be interpreted with consideration when it comes to concluding that the variants involved in significant co-eQTLs are associated with a change in co-expression. We see that a vast proportion of our results is related to eQTL effect of a SNP-gene involved in the co-eQTL, for example, *rs1160871-JAZF1* in microglia, rather than changes in co-expression of the gene pair (Figure 19). Particularly in the case where a SNP-gene pair is a co-eQTL for the majority of combinations tested, this is often found to be driven by changes in the expression of a single gene between the genotypes rather than co-expression changes.

A limitation of our study is that we chose to constrain the

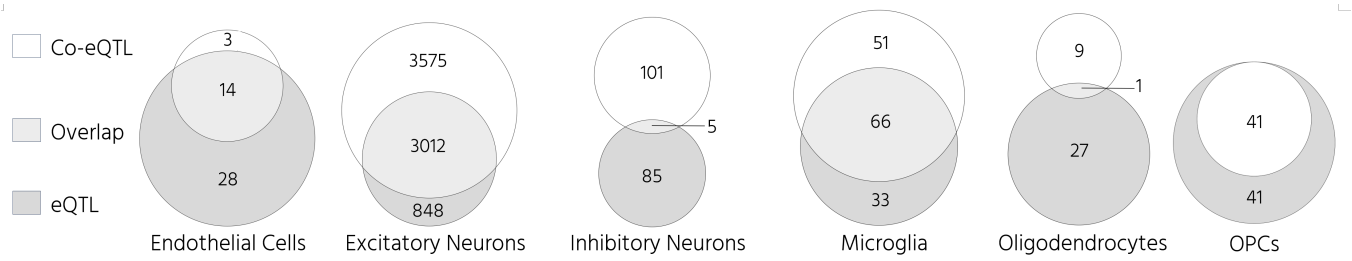



**Figure 19** Co-eQTL: number of significant co-eQTLs per cell type, eQTL: number of SNP-gene-gene combinations for which at least one gene has significant eQTL effect for the SNP. Overlap: Intersection illustrates the proportion of significant co-eQTLs subject to eQTL effect which is strong for endothelial cells, excitatory neurons and microglia.

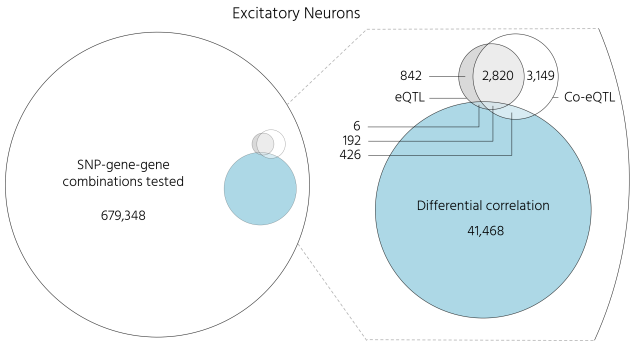



**Figure 20** Test space for co-eQTLs in Excitatory Neurons, 721,280 SNP-gene-gene combinations were tested of which 6,587 co-eQTLs are significant. Of the tested co-eQTLs, 42,092 (of which 618 significant) were differentially correlated and 3,860 (of which 3,012 significant) had eQTL effect on at least one of the genes involved looking at the intersection of eQTL and co-eQTL in the diagram.

variants and genes assessed to known relevant AD loci variants and genes, as analyzing the impact of genetic variants on gene co-expression relations is subject to a tremendous multiple testing burden. The limited set of genes considered also hindered biological interpretation of the significant co-eQTLs in enrichment analysis, as the limited set of tested genes has to be considered as background. Our method does not allow for assessing co-expression effects for alleles with low prevalence as a minimum number of individuals needs to be present for confident co-expression probability density estimate.

We derived recommendations for future co-eQTL studies based on observations in our analysis. As our method is shown to be sensitive to eQTL effects, we recommend further analysis on how the method can be tailored to allow for true co-eQTLs to be detected that involve changes in co-expression rather than expression changes driven by a SNP-gene effect. On re-evaluation, we consider the expression threshold used for inclusion of genes in our analysis to be overly conservative, as we see that selection of genes that are expressed in the majority of cells only might lead to missing changes in co-expression which are driven by a suppressive effect of a variant on transcription.

Even though we intentionally did not constrain the search space in a hypothesis-driven way, we feel that for future analysis this may be a useful way of limiting the multiple-testing burden while assessing the effectiveness of the method in discovery of true co-eQTLs. For future work, we suggest constraining the search space to 1) known transcription factors and/or RNA binding proteins, and variants in their proximity and test for changes in co-expression with other genes and, 2) testing all other genes to for a known disease-associated variant-gene combination to identify downstream consequences of the variant. The input set of genes can be further limited by determining genes with high expression correlation, as genes with high expression correlation are likely to result in similar co-eQTLs when their co-expression is analyzed with a third gene.

As an alternative to pair-wise assessment of co-expression relations between genes, methods that assess co-expression changes on module level related to genetic variants might be a better fit that is less sensitive to eQTL effects of single genes while limiting the overall number of tests. Modules in this case could be sets of genes known to be involved in a biological function for example.

Overall, we show that exploring genetic variant-associated changes in gene (co-)expression using sc-RNA seq data is a promising approach in finding cell type-specific mechanisms that may be altered in AD.

## References


Bai B, Vanderwall D, Li Y, Wang X, Poudel S, Wang H, Dey KK, Chen PC, Yang K, Peng J. 2021. Proteomic landscape of Alzheimer's Disease: novel insights into pathogenesis and biomarker discovery. Molecular Neurodegeneration. 16:55.

Balachandran P, Beck CR. 2020. Structural variant identification and characterization. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology. 28:31–47.

Beck TN, Nicolas E, Kopp MC, Golemis EA. 2014. Adaptors for disorders of the brain? The cancer signaling proteins NEDD9, CASS4, and PTK2B in Alzheimer's disease. Oncoscience. 1:486–503.

Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, Naj AC, Campos-Martin R, Grenier-Boley B, Andrade V *et al*. 2022. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nature Genetics. 54:412–436. Number: 4 Publisher: Nature Publishing Group.

Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. 2018. Religious Orders Study and Rush Memory and Aging Project. Journal of Alzheimer's disease : JAD. 64:S161–S189.

Bioconductor Package Maintainer. 2023. liftOver: Changing genomic coordinate systems with rtracklayer::liftOver. R package. https://www.bioconductor.org/help/workflows/liftOver/.

Bouland GA, Mahfouz A, Reinders MJT. 2023a. Consequences and opportunities arising due to sparser single-cell RNA-seq datasets. Genome Biology. 24:86.

Bouland GA, Marinus KI, van Kesteren RE, Smit AB, Mahfouz A, Reinders MJT. 2023b. Single-cell RNA sequencing data reveals rewiring of transcriptional relationships in Alzheimer's Disease associated with risk variants. medRxiv: The Preprint Server for Health Sciences. p. 2023.05.15.23289992.

Brown CD, Mangravite LM, Engelhardt BE. 2013. Integrative Modeling of eQTLs and Cis-Regulatory Elements Suggests Mechanisms Underlying Cell Type Specificity of eQTLs. PLOS Genetics. 9:e1003649. Publisher: Public Library of Science.

Cano-Gamez E, Trynka G. 2020. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Frontiers in Genetics. 11.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 4:7.

de Klein N, Tsai EA, Vochteloo M, Baird D, Huang Y, Chen CY, van Dam S, Oelen R, Deelen P, Bakker OB *et al*. 2023. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. Nature Genetics. 55:377–388. Number: 3 Publisher: Nature Publishing Group.

DeTure MA, Dickson DW. 2019. The neuropathological diagnosis of Alzheimer's disease. Molecular Neurodegeneration. 14:32.

Dwyer K, Agarwal N, Pile L, Ansari A. 2021. Gene Architecture Facilitates Intron-Mediated Enhancement of Transcription. Frontiers in Molecular Biosciences. 8:669004.

Ferreira A, Royaux I, Liu J, Wang Z, Su G, Moechars D, Callewaert N, De Muynck L. 2022. The 3-O sulfation of heparan sulfate proteoglycans contributes to the cellular internalization of tau aggregates. BMC Molecular and Cell Biology. 23:61.

Flynn ED, Tsu AL, Kasela S, Kim-Hellmuth S, Aguet F, Ardlie KG, Bussemaker HJ, Mohammadi P, Lappalainen T. 2022. Transcription factor regulation of eQTL activity across individuals and tissues. PLoS Genetics. 18:e1009719.

Fujita M, Gao Z, Zeng L, McCabe C, White CC, Ng B, Green GS, Rozenblatt-Rosen O, Phillips D, Amir-Zilberstein L *et al*. 2022. Cell-subtype specific effects of genetic variation in the aging and Alzheimer cortex. Pages: 2022.11.07.515446 Section: New Results.

Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL. 2006. Role of Genes and Environments for Explaining Alzheimer Disease. Archives of General Psychiatry. 63:168–174.

Giralt A, de Pins B, Cifuentes-Díaz C, López-Molina L, Farah AT, Tible M, Deramecourt V, Arold ST, Ginés S, Hugon J *et al*. 2018. PTK2B/Pyk2 overexpression improves a mouse model of Alzheimer's disease. Experimental Neurology. 307:62–73.

Green GS, Fujita M, Yang HS, Taga M, McCabe C, Cain A, White CC, Schmidtner AK, Zeng L, Wang Y *et al*. 2023. Cellular dynamics across aged human brains uncover a multicellular cascade leading to Alzheimer's disease. bioRxiv: The Preprint Server for Biology. p. 2023.03.07.531493.

Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E *et al*. 2018. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Research. 46:D380–D386.

Hansen DV, Hanson JE, Sheng M. 2018. Microglia in Alzheimer's disease. The Journal of Cell Biology. 217:459–472.

Holler CJ, Davis PR, Beckett TL, Platt TL, Webb RL, Head E, Murphy MP. 2014. Bridging integrator 1 (BIN1) protein expression increases in the Alzheimer's disease brain and correlates with neurofibrillary tangle pathology. Journal of Alzheimer's disease: JAD. 42:1221–1227.

Hou J, Ye X, Feng W, Zhang Q, Han Y, Liu Y, Li Y, Wei Y. 2022.

Distance correlation application to gene co-expression network analysis. BMC Bioinformatics. 23:81.

Huang J, Chen J, Esparza J, Ding J, Elder J, Abecasis GR, Lee YA, Lathrop GM, Moffatt MF, Cookson WOC *et al*. 2015. eQTL mapping identify insertion and deletion specific eQTLs in multiple tissues. Nature communications. 6:6821.

Huang Q, Ritchie SC, Brozynska M, Inouye M. 2018. Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Research. 46:e133.

Janeway C, Travers P, Walport M. 2001. *Immunobiology: The Immune System in Health and Disease*. Garland Science. New York. fifth edition.

Jiao HS, Yuan P, Yu JT. 2023. TMEM106B aggregation in neurodegenerative diseases: linking genetics to function. Molecular Neurodegeneration. 18:54.

Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. 2022a. Single-cell RNA sequencing technologies and applications: A brief overview. Clinical and Translational Medicine. 12:e694.

Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. 2022b. Single-cell RNA sequencing technologies and applications: A brief overview. Clinical and Translational Medicine. 12:e694.

Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. Protein Science: A Publication of the Protein Society. 28:1947–1951.

Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Research. 51:D587–D592.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 28:27–30.

Kang JB, Raveane A, Nathan A, Soranzo N, Raychaudhuri S. 2023. Methods and Insights from Single-Cell Expression Quantitative Trait Loci. Annual Review of Genomics and Human Genetics. 24:277–303. _eprint: https://doi.org/10.1146/annurev-genom-101422-100437.

Kontio JAJ, Rinta-aho MJ, Sillanpää MJ. 2020. Estimating Linear and Nonlinear Gene Coexpression Networks by Semiparametric Neighborhood Selection. Genetics. 215:597–607.

Kulczyńska-Przybik A, Mroczko P, Dulewicz M, Mroczko B. 2021. The Implication of Reticulons (RTNs) in Neurodegenerative Diseases: From Molecular Mechanisms to Potential Diagnostic and Therapeutic Approaches. International Journal of Molecular Sciences. 22:4630.

Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. 2018. The Human Transcription Factors. Cell. 172:650–665.

Lang CM, Fellerer K, Schwenk BM, Kuhn PH, Kremmer E, Edbauer D, Capell A, Haass C. 2012. Membrane Orientation and Subcellular Localization of Transmembrane Protein 106B (TMEM106B), a Major Risk Factor for Frontotemporal Lobar Degeneration?. The Journal of Biological Chemistry. 287:19355–19365.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 28:882–883.

Li S, Schmid KT, de Vries DH, Korshevniuk M, Losert C, Oelen R, van Blokland IV, Groot HE, Swertz MA, van der Harst P *et al*. 2023a. Identification of genetic variants that impact gene co-expression relationships using large-scale single-cell data. Genome Biology. 24:80.

Li Y, Xu M, Xiang BL, Li X, Zhang DF, Zhao H, Bi R, Yao YG. 2023b. Functional genomics identify causal variant underlying the protective CTSH locus for Alzheimer's disease. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology. 48:1555–1566.

Li Z, Farias FG, Dube U, Del-Aguila JL, Mihindukulasuriya KA, Fernandez MV, Ibanez L, Budde JP, Wang F, Lake AM *et al*. 2020. The TMEM106B FTLD-protective variant, rs1990621, is also associated with increased neuronal proportion. Acta neuropathologica. 139:45–61.

Lin J. 1991. Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory. 37:145–151. Conference Name: IEEE Transactions on Information Theory.

Liu X, Che R, Liang W, Zhang Y, Wu L, Han C, Lu H, Song W, Wu Y, Wang Z. 2022. Clusterin transduces Alzheimer-risk signals to amyloidogenesis. Signal Transduction and Targeted Therapy. 7:1–4. Number: 1 Publisher: Nature Publishing Group.

Luo J, Wu X, Cheng Y, Chen G, Wang J, Song X. 2023. Expression quantitative trait locus studies in the era of single-cell omics. Frontiers in Genetics. 14.

Machiela MJ, Chanock SJ. 2015. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics (Oxford, England). 31:3555–3557.

Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X *et al*. 2019. Single-cell transcriptomic analysis of Alzheimer's disease. Nature. 570:332. Publisher: NIH Public Access.

Oelen R, de Vries DH, Brugge H, Gordon MG, Vochteloo M, Ye CJ, Westra HJ, Franke L, van der Wijst MGP. 2022. Single-cell RNA-sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. Nature Communications. 13:3267. Number: 1 Publisher: Nature Publishing Group.

Pradhan LK, Das SK. 2021. The Regulatory Role of Reticulons in Neurodegeneration: Insights Underpinning Therapeutic Potential for Neurodegenerative Diseases. Cellular and Molecular Neurobiology. 41:1157–1174.

Purcell S, Chang CC. 2023a. PLINK 1.9. www.cog-genomics.org/plink/1.9/.

Purcell S, Chang CC. 2023b. PLINK 2.0. www.cog-genomics.org/plink/2.0/.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139–140.

Romano R, Bucci C. 2020. Role of EGFR in the Nervous System. Cells. 9:1887. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Ryu C. 2023. dlookr: Tools for Data Diagnosis, Exploration, Transformation. R package version 0.6.2 https://CRAN.R-project.org/package=dlookr.

Salazar A, Tesi N, Hulsman M, Knoop L, Lee Svd, Wijesekera S, Krizova J, Reinders M, Holstege H. 2023. An AluYb8 mobile element characterises a risk haplotype of TMEM106B associated in neurodegeneration. Pages: 2023.07.16.23292721.

Scott AJ, Chiang C, Hall IM. 2021. Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes. Genome Research. 31:2249–2257.

Scott DW. 1992. *Multivariate density estimation: theory, practice, and visualization*. Wiley series in probability and mathematical statistics. Wiley. New York.

Shabalin AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 28:1353–1358.

Shen R, Zhao X, He L, Ding Y, Xu W, Lin S, Fang S, Yang W, Sung K, Spencer B *et al*. 2020. Upregulation of RIN3 induces endosomal dysfunction in Alzheimer's disease. Translational Neurodegeneration. 9:26.

Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva N, Pyysalo S *et al*. 2022. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research. 51:D638–D646.

Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM *et al*. 2021. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 590:290–299. Number: 7845 Publisher: Nature Publishing Group.

Tesi N, Lee Svd, Hulsman M, Schoor NMv, Huisman M, Pijnenburg Y, Flier WMvd, Reinders M, Holstege H. 2023. Cognitively Healthy Centenarians are genetically protected against Alzheimer's disease specifically in immune and endo-lysosomal systems. Pages: 2023.05.16.23290049.

Tesi N, van der Lee S, Hulsman M, Holstege H, Reinders MJT. 2021. snpXplorer: a web application to explore human SNP-associations and annotate SNP-sets. Nucleic Acids Research. 49:W603–W612.

Tsai AP, Lin PBC, Dong C, Moutinho M, Casali BT, Liu Y, Lamb BT, Landreth GE, Oblak AL, Nho K. 2021. INPP5D expression is associated with risk for Alzheimer's disease and induced by plaque-associated microglia. Neurobiology of Disease. 153:105303.

van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. 2018. Gene co-expression analysis for functional classification and gene–disease predictions. Briefings in Bioinformatics. 19:575–592.

van der Flier WM, de Vugt ME, Smets EMA, Blom M, Teunissen CE. 2023. Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia. Nature Aging. 3:494–505. Number: 5 Publisher: Nature Publishing Group.

Venables WN, Ripley BD, Venables WN. 2002. *Modern applied statistics with S*. Statistics and computing. Springer. New York. fourth edition. OCLC: ocm49312402.

Võsa U, Claringbould A, Westra HJ, Bonder MJ, 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. Nature Genetics. 53:1300–1310. Number: 9 Publisher: Nature Publishing Group.

Wang H. 2021. Microglia Heterogeneity in Alzheimer's Disease: Insights From Single-Cell Technologies. Frontiers in Synaptic Neuroscience. 13.

Wang H, Dombroski BA, Cheng PL, Tucci A, Si YQ, Farrell JJ, Tzeng JY, Leung YY, Malamon JS, Project TADS *et al*. 2023a. Structural Variation Detection and Association Analysis of Whole-Genome-Sequence Data from 16,905 Alzheimer's Diseases Sequencing Project Subjects. Pages: 2023.09.13.23295505.

Wang Z, Patel VN, Song X, Xu Y, Kaminski AM, Doan VU, Su G, Liao Y, Mah D, Zhang F *et al*. 2023b. Increased 3-O-sulfated heparan sulfate in Alzheimer's disease brain is associated with genetic risk gene HS3ST1. Science Advances. 9:eadf6232.

Westra HJ, Franke L. 2014. From genome to function by studying eQTLs. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 1842:1896–1902.

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L *et al*. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2.

Publisher: Elsevier.

Xiong X, James BT, Boix CA, Park YP, Galani K, Victor MB, Sun N, Hou L, Ho LL, Mantero J *et al*. 2023. Epigenomic dissection of Alzheimer's disease pinpoints causal variants and reveals epigenome erosion. Cell. 186:4422–4437.e21. Publisher: Elsevier.

Yang HS, White CC, Klein HU, Yu L, Gaiteri C, Ma Y, Felsky D, Mostafavi S, Petyuk VA, Sperling RA *et al*. 2020. Genetics of Gene Expression in the Aging Human Brain Reveal TDP-43 Proteinopathy Pathophysiology. Neuron. 107:496–508.e6. Publisher: Elsevier.

Yang J, Wang D, Yang Y, Yang W, Jin W, Niu X, Gong J. 2021. A systematic comparison of normalization methods for eQTL analysis. Briefings in Bioinformatics. 22:bbab193.

Zhang J, Zhao H. 2023. eQTL studies: from bulk tissues to single cells. Journal of Genetics and Genomics. .

Zhou M, Xu X, Wang H, Yang G, Yang M, Zhao X, Guo H, Song J, Zheng H, Zhu Z *et al*. 2020. Effect of central JAZF1 on glucose production is regulated by the PI3K-Akt-AMPK pathway. The FASEB Journal. 34:7058–7074. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1096/fj.201901836RR.

## Methods

### Data sources and pre-processing

All data was obtained from the AD knowledge portal at Synapse.org (https://adknowledgeportal.synapse.org/).

***Genetic variants*** Raw PLINK files (.bed/.bim/.fam) were sourced from the Synapse portal (SynID: syn17008939). The genotyping data consisted of two batches ($n_1$ = 1709 / $n_2$ = 382). The raw files were based on reference genome hg18 for batch 1 and hg19 (GRCh37) for batch 2 respectively. Only variants on autosomal chromosomes were considered.

Individuals with low-quality genotyping data (>4% of missing genotypes) were removed. Variants with low genotyping rate (<98%) were removed. The genotyping data of batch 1 was lifted over to hg19 using the liftOver R package using the hg18ToHg19.over.chain coordinates file from UCSC (http://hgdownload.soe.ucsc.edu/goldenPath/hg18/liftOver/ (Bioconductor Package Maintainer 2023). Duplicate variants present in the data were removed.

Besides the genotyping data from ROSMAP, genotyping data from Seattle Alzheimer's Disease Brain Cell Atlas was sourced from the Synapse portal (SynID: syn28257618) for possible inclusion in further analysis. Only variants were kept that are present in the genotyping data for Seattle AD and both ROSMAP batches.

The genotyping data was prepared for imputation using the Perl script (HRC-1000G-check-bim-v4.2.7) from the McCarthy Group (https://www.well.ox.ac.uk/~wrayner/tools/) and compared to HRC reference panel (ftp://ngs.sanger.ac.uk/production/hrc/HRC.r1-1/HRC.r1-1.GRCh37.wgs.mac5.sites.tab.gz) using a frequency file generated with PLINK v1.90b6.16 (Purcell and Chang 2023a; Chang *et al.* 2015). The resulting PLINK files were converted to VCF using PLINK v2.00a2.3LM (Purcell and Chang 2023b) and were imputed using the TOPMED imputation panel (phasing Eagle v2.4 - rsq filter 0.2) (Taliun *et al.* 2021) (version R2 on GRCh38) on the TOPMED Imputation Server 1.7.3 TOPMED r2 panel (https://imputation.biodatacatalyst.nhlbi.nih.gov), the resulting variants are mapped to reference genome GRCh38.

Post imputation and QC, 5,519,683 variants with a minor allele frequency (MAF) above 0.05 and a Hardy-Weinberg p-value of $10^{-6}$ remained for analysis. Functional annotation of genetic variants was done with snpXplorer (Tesi *et al.* 2021). Linkage of genetic variants was assessed using LDlink (Machiela and Chanock 2015).

Genetic variants are reported in the following format: *chr7:28129126:G:GTCT*, where *chr* stands for the chromosome where the variant is located, followed by the position based on reference genome hg38, the first allele (*G*) is the counted allele and the second (*GTCT*) is the alternative allele. Please note that the counted and alternative allele do not necessarily correspond to the major and minor allele of the variant.

***Metadata*** Clinical data for the ROSMAP cohort was sourced from the Synapse portal (SynID syn3191087). In the clinical data, the exact age at the time of death was hidden for individuals over 90 years old, in the analysis we use age 90 for these individuals. Specimen-level metadata was acquired from the Synapse portal as well (SynID syn21323366 and syn21314550).

***Single-cell RNA expression data*** Single-cell RNA-sequencing data was obtained by preparing a sequencing library for a batch of a maximum of 8 individuals. Per batch, a technical replicate was prepared with a different library, we refer to the 2 replicates as A and B respectively. Every replicate was sequenced at 2 centers. Figure 21 depicts the replicates present in the data schematically. In the data available at the Synapse portal, the results of the sequencing runs at 2 different centers were already merged into a single file per library batch.

Processed count data was taken from the Synapse portal (SynID syn51123521). The count data was obtained by aggregating the results of the 2 sequencing runs, and demultiplexing (assigning the reads to an individual) was done using WGS data. The reads were assigned to genes using the transcriptome model 'GRCh38-2020-A' provided by 10X Genomics. Please refer to Fujita *et al.* (2022) for the complete description of the sequencing workflow.

***Cell type annotation and QC*** Cell types were annotated using the annotation (SynID: syn51218314 V1) provided by the authors of Green *et al.* (2023) which also links cell barcodes to individuals. Cells for the following major brain cell types were retained for analysis: astrocytes, endothelial cells, excitatory neurons, inhibitory neurons, oligodendrocytes, oligodendrocyte progenitor cells (OPCs) and microglia, as these cell types are present in quantities sufficient for analysis with sufficient detection power.

Cells that have a cell type annotation in Green *et al.* (2023) are cells that have passed a thorough cell type-specific quality control which uses thresholds on the number of Unique Molecular Identifiers (UMI) and the number of genes detected on the cell type level. As the cells were already subjected to a thorough quality control process, we refrained from performing additional quality control, leaving a total of 1,337,224 annotated cells for inclusion in the analysis (Table 5).

***Selection of individuals*** The individuals selected for this analysis are the individuals that have both genotyping and single cell RNA-sequencing data available, and are present in the cell type annotation by Green *et al.* (2023). Individuals that were excluded on the individual level by the authors because of quality control issues (such as duplicate samples, samples that could not be demultiplexed because of missing whole genome sequencing data and specimens with median UMI counts less than 1500) were also excluded from the analysis.

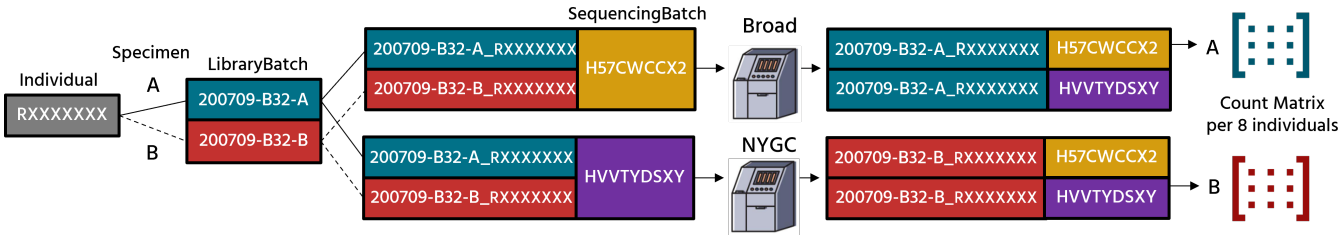



**Figure 21** Single cell RNA-sequencing workflow

| Cell type | # cells | % of total |
|---|---|---|
| Astrocyte | 189,874 | 14.20 % |
| Endothelial | 8,479 | 0.63 % |
| Excitatory Neurons | 533,469 | 39.89 % |
| Inhibitory Neurons | 210,703 | 15.76 % |
| Microglia | 69,370 | 5.19 % |
| OPCs | 48,467 | 3.62 % |
| Oligodendrocytes | 276,862 | 20.70 % |

**Table 5** Number of cells per major cell type included in the sc-eQTL analysis

## sc-eQTL analysis

***Pseudobulk pre-processing and normalization*** ScRNA-seq data is known for its sparsity: for the majority of genes zero transcripts are measured in a single cell and these zero measurements can be the result of biological and technical factors (Kang *et al.* 2023; Luo *et al.* 2023; Bouland *et al.* 2023a). We deal with this by aggregating the scRNA-seq count data in a pseudobulk dataset containing an expression matrix for each of the 7 cell types, with a column per individual and a row per gene. The expression matrices were populated with the sum of the counts assigned to a certain gene for an individual. An individual was only included in the expression matrix for a specific cell type in case it had at least 10 cells for that respective cell type. Per individual, the counts from both technical replicates (sequencing library A and B) were aggregated together in a single pseudobulk matrix. In total 1,336,849 cells qualified for inclusion in the analysis. All of the following steps were executed with R (version 4.2.0). The data was subjected to the following normalization pipeline. Per cell type, gene-level quality control using *filterByExpr* from edgeR (3.40.0) (Robinson *et al.* 2010) was performed with the default thresholds (min.count = 10, min.total.count = 15, large.n = 10, min.prop = 0.7). This function uses these thresholds to calculate library-size (CPM) aware thresholds for retaining genes. The function *calcNormFactors* from edgeR was used to obtain Trimmed mean of M-values (TMM) normalization factors, these are subsequently used with the *voom* function of limma (3.54.0) to obtain quantile normalized $\log_2$ CPM values. TMM normalization is a well-established method for normalization of bulk-RNA sequencing (Yang *et al.* 2021). Genes with a mean expression (considering all individuals) of $\log_2$ CPM <2 were discarded. Batch correction was performed using the *ComBat* function from the sva package (3.46.0) (Leek *et al.* 2012). The batches considered to be present in the data were taken to be the prefix of the specimenID (for example: 190403-B4). Any resulting gene with zero expression variance was removed.

***Linear model with covariates*** The eQTL analysis was performed using the package MatrixeQTL (2.3) (Shabalin 2012). We used an additive linear model (dosage ~ gene expression) with covariates to test for association between the genotype and gene expression. An additive linear model assumes the effect of a variant to be proportional with the number of alleles present. Nominal p-values for the resulting eQTLs were determined using a t-statistic. For every gene, genetic variants were tested that were located in the range from -1Mbp to +1Mbp from the transcription start site (TSS) as variants associated with gene expression are known to cluster in this region (Brown *et al.* 2013). The position of the TSS per gene was determined by accessing the genomic coordinates for the Ensembl gene IDs given in the features file of the scRNA-seq data using Ensembl Biomart (www.ensembl.org/biomart - accessed March 2023). Genes that could not be resolved to a chromosome were excluded from the analysis.

The same covariates were included in our analysis as in the Fujita *et al.* (2022) study: sex, post-mortem interval, age at time of death, the first 3 principal components of the genotyping data, the first 30 principal components of the gene expression data, the number of genes expressed per individual and whether the individual was part of the ROS or MAP study. Genotype principal components were generated using PLINK v2.00a2.3LM. Expression principal components were determined per cell type using the built-in *prcomp* function in R.

***Multiple testing correction*** Multiple testing correction was performed at the sc-eQTL, sc-eSNP and sc-eGene level. On eQTL (eSNP-eGene) level per cell type, sc-eQTLs were considered significant at false discovery rate (FDR) $\leq 0.05$ by considering all tests for that cell type. The FDR was calculated using the adapted Benjamini-Hochberg procedure implemented in MatrixEQTL, the procedure is adapted so that the FDR can be calculated without storing all intermediate results (Shabalin 2012). We considered all unique SNPs involved in sc-eQTLs significant at FDR $\leq 0.05$ as eSNPs. On the sc-eGene level, a Bonferroni corrected p-value per sc-eGene was calculated by taking the lowest nominal p-value from all eQTLs relating to that gene and multiplying it by the number of tests per-

formed for that gene. Significant sc-eGenes were determined using the Benjamini-Hochberg procedure, ranking sc-eGenes based on their Bonferroni corrected p-value and considering sc-eGenes with an FDR below 0.05 as significant.

## co-eQTL analysis

***Pre-processing and normalization*** The scRNA-seq count data was aggregated in pseudobulk matrices for every major cell type as described for the eQTL analysis, but was subjected to a different pre-processing workflow. Batch correction was performed using the *ComBat-seq* function from the sva package (3.46.0) (Leek *et al.* 2012), returning batch-corrected count data. The count data was library-size normalized (CPM) using the *cpm* function from edgeR (3.40.0) (Robinson *et al.* 2010).

***Selection of genetic variants and genes*** As we are interested in co-expression QTLs in the context of AD and unconstrained analysis of gene pairs posed a vast multiple testing burden, we limited this analysis to the set of known AD-associated SNPs (Bellenguez *et al.* (2022)) potentially explaining changes in co-expression of AD-associated genes from the AD KEGG pathway (Kanehisa *et al.* 2023; Kanehisa 2019; Kanehisa and Goto 2000) and the nearest protein-coding genes for the GWAS SNPs. The reason for including the latter is that most variants affect genes in their proximity and we are interested in how this affects the co-expression relation (Cano-Gamez and Trynka 2020). Changes in gene co-expression were investigated for all genes passing the cell-based expression threshold (see *Methods*). We tested all variants for every gene pair as we imposed no assumptions on the driving factors behind co-expression changes. Per variant, only the genotypes were considered that had at least 20 individuals, this indirectly also excludes genetic variants with a low minor allele frequency (MAF), resulting in the inclusion of 56/86 genetic variants in the analysis.

Per cell type, an expression threshold was put used to filter genes for inclusion in the analysis: per individual, the fraction of cells that express a gene (gene count per cell > 0) was determined. Genes were included in further analysis if the mean fraction of expression over all individuals was $\geq 0.5$. This threshold was chosen assuming that co-expression effects of a gene pair can only be meaningful if both genes have an important role in a cell type, which is assumed to be the case when a gene is expressed in a considerable proportion of cells. The number of genes and gene pairs tested per cell type is given in Table 6.

***Gene annotation and enrichment analysis*** Human transcription factors were sourced from http://humantfs.ccbr.utoronto.ca/ (Lambert *et al.* 2018). Gene names were sourced from HGNC (www.genenames.org). Protein-protein interactions of genes were determined using StringDB. (Szklarczyk *et al.* 2022). GO term enrichment for biological processes, cellular components, and molecular function was evaluated using R package *clusterProfiler* (4.6.2) (Wu *et al.* 2021), the subset of genes tested was considered as background where applicable.

| Cell type | # genes | # gene pairs |
|---|---|---|
| Astrocyte | 41 | 820 |
| Endothelial | 43 | 903 |
| Excitatory Neurons | 161 | 12,880 |
| Inhibitory Neurons | 91 | 4,095 |
| Microglia | 34 | 561 |
| Oligodendrocytes | 29 | 406 |
| OPCs | 45 | 990 |

**Table 6** Number of genes and gene pairs tested per cell type in co-eQTL analysis, all gene pairs are tested for every variant considered

***Density estimation of gene-pair co-expression distribution*** A dosage for a genetic variant is a continuous value between 0 and 2, representing the probable number of counts for the allele marked as the counted allele. We round the dosages to the nearest integer and assign individuals to group 0, 1 or 2 based on the number of copies of the counted allele present.

For every gene pair considered, a co-expression distribution per cell type is created for every genotype group with sufficient individuals. This gives a scatter plot representation per genotype group as shown in Figure 22 where every data point represents an individual and the position is based on the expression values for that individual for gene A (x-axis) and gene B (y-axis) respectively. The axis limits are defined by the minimum and maximum expression value of the genes considering all genotypes jointly.

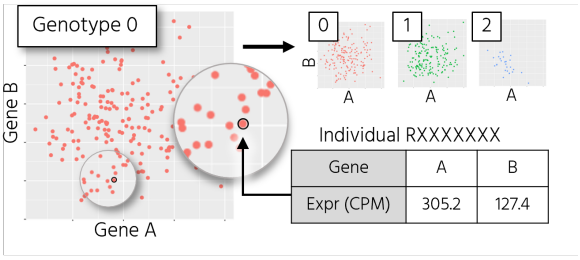



**Figure 22** Co-expression for gene pair (A-B) per genotype group

Per genotype group, the probability density of the 2D co-expression distribution for a gene pair is estimated by performing kernel density estimation. Kernel density estimation is a non-parametric method, that estimates the probability density by placing a kernel on every data point and summing their contributions to the probability density. A kernel density estimate (KDE) was calculated using a Gaussian kernel with the *kde2d* function of MASS (7.3-60) in R (Venables *et al.* 2002). The bandwidth of the kernel was set using *bandwidth.nrd* based on the standard deviation and interquartile range of the expression of the genes (for all genotype groups) (Scott 1992). The contribution of all kernels is summed up, and evaluated on a grid of 10 points in both x and y direction,

resulting in a 100-point local density estimate. The resulting density was normalized to unit probability by dividing the local density at a grid point by the sum of density. A graphic example of the resulting probability density per group for a gene pair is shown in Figure 23.

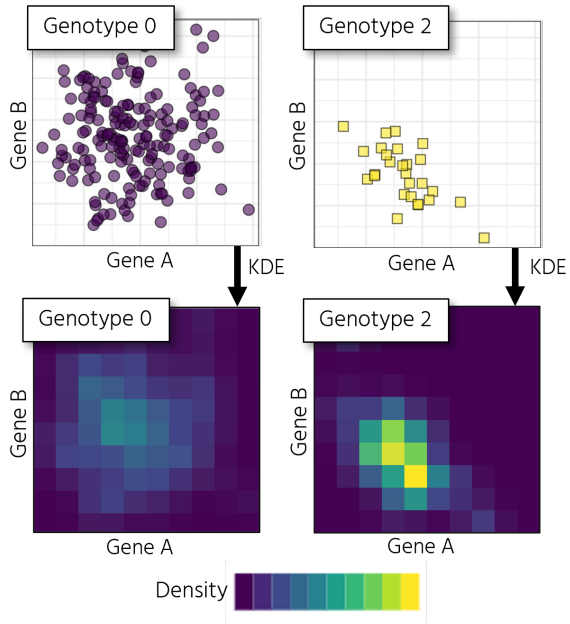



**Figure 23** Estimation of the probability density of gene co-expression per genotype group through kernel density estimation - Color gradient in the kernel density estimate represents the local probability density

***Distance between co-expression distributions*** The distance between the kernel density estimates of 2 genotypes is determined by calculating the Jensen-Shannon Divergence (JSD), which is a symmetric distance metric based on the Kullback-Leibler divergence given in Equation 1, where P and Q are the two probability distribution compared over a discrete grid X with individual points x (Lin 1991). The distance is symmetrized by determining M, the average of distribution P and Q as given in Equation 2, The resulting Jensen-Shannon divergence can be calculated as per Equation 3.

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \quad (1)$$

$$M = \frac{1}{2}(P + Q) \quad (2)$$

$$JSD(P||Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M) \quad (3)$$

The distance of a genotype group for a variant to the other groups was only calculated in case the group contained at least 20 individuals (5% of all individuals), as a minimum number of individuals is needed to create a reliable kernel density estimate. The JSD was calculated using the function *jsd* from dlookr (version 0.6.2) in R (Ryu 2023). The overall JSD is calculated by determining the local JSD for the same gridpoint in the KDE for both genotype groups and summing all local distances over the grid.

The grid size of the KDE was optimized for the convergence of the distance metric, meaning evaluation on a finer grid would increase computational load without considerable changes in distances between distributions.

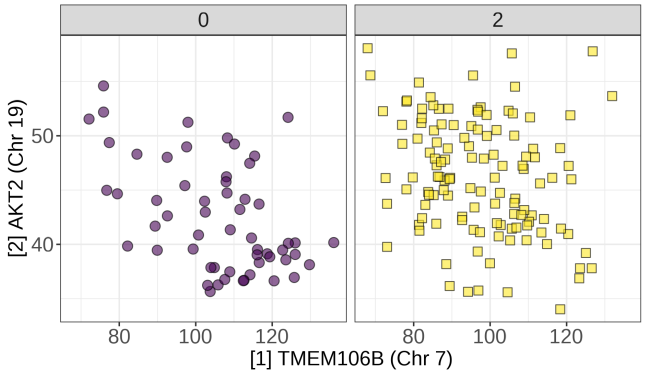
***Statistical significance and multiple testing*** Per variant, an empirical null distribution was created to evaluate the significance of a result (distance between 2 co-expression distributions) in the absence of a true null distribution of distances between gene co-expression distributions.

At most three possible pair of genotype groups exist for a variant (0-1, 0-2, 1-2). For every pair of genotypes, first, the nominal distance between groups is calculated for all gene pairs considered. Per variant, the combination of genotype groups with the highest average nominal distance between distributions is retained for analysis. This is done assuming that if there is a significant effect of a variant on the co-expression relations of gene pairs, this will be reflected most strongly in the pair of groups with the largest distances between distributions. Rationale behind considering the combinations of genotype groups separately is that every group has a different number of individuals, and keeping the imbalance between the number of individuals the same will allow for the identification of significant gene pairs at the variant level as the distance between distributions are determined between distributions with the same imbalance.

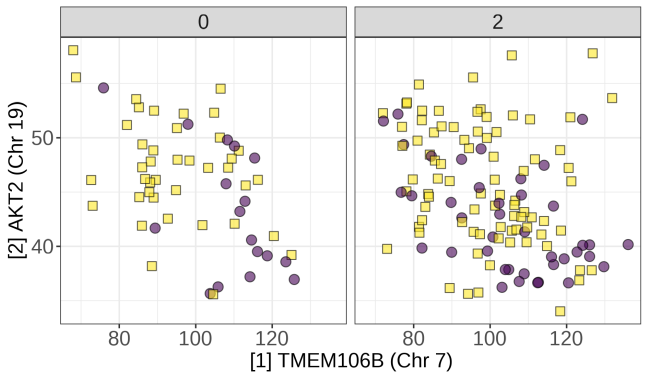
For the selected combination of genotype groups, the empirical null distribution is built out of distances between distributions for all gene pairs considered. For every gene-pair, the genotype labels are permuted (re-distributed based on random sampling, keeping the number of individuals per genotype group the same), which gives a new co-expression distribution per group for which a kernel density estimate is determined.

The distance between the kernel density estimates is calculated and added to the null distribution. This procedure is performed 1000 times per gene pair to create a null distribution with sufficient resolution. The repeated calculation of permuted distributions is computationally intensive which needs to be considered when the analysis is applied for a larger set of variants and genes.

The nominal significance (p-value) for a gene pair (for a variant-genotype combination) is determined using the fraction of distances in the null distribution for that variant that are higher than the (non-permuted) distance result for that gene pair. The Benjamini-Hochberg procedure is applied to the nominal p-value to control for FDR of the results per variant, results with FDR < 0.05 are considered significant. Reason for choosing the control for FDR instead of correcting for family-wise error rate is that Bonferroni correction is based on the assumption that the tests are independent were this is not the case are genes re-occur in different tests for differ-




**(a)** Pre-permutation: co-expression distribution of genotype 0 and 2




**(b)** Post-permutation: co-expression distribution of genotype 0 and 2 - data points are colored by the genotype where they originally belong to

**Figure 24** Random permutation of the genotype labels

ent gene pairs.

***Explaining co-expression variation*** To determine which factors are driving the significant co-eQTLs discovered we assess whether the change in co-expression distribution is driven by eQTL effect of the variant on either of the genes involved and whether there is a change in expression correlation between the genotypes for the gene-pair considered.

We use a simplified version of the eQTL workflow described earlier to assess for presence of eQTL effect of the variant on gene expression using a linear model (gene expression $\sim$ dosage). The nominal p-values from the linear regression were corrected for false discovery rate (FDR) using the Benjamini-Hochberg method, considered all findings with an FDR <0.05 as an eQTL.

The correlation of expression for a gene-pair for a genotype was determined by calculating the Pearson correlation coefficient (using *cor.test* in R) between the expression values of a gene-pair for every genotype for every variant. For a result, the difference in correlation coefficient for a gene pair between the genotypes was calculated. A result was considered to be driven by differential correlation, if the difference in correlation coefficient between the genotypes for the gene pair was in the top 5% of absolute differences in correlation.

## Supplementary material

### Comparison to the eQTL study by Fujita et al. (2022)

Our analysis was performed analogously to the analysis described in Fujita *et al.* (2022) which was performed using the same sn-RNAseq dataset of participants of the ROS and MAP cohort studies. Comparing our analysis to the Fujita *et al.* (2022) study, we identified the following differences: we considered variants (SNPs and small structural variants) on autosomal chromosomes from (imputed) genotyping data (see *Methods*), whereas Fujita uses SNPs from autosomal and sex chromosomes determined by whole genome sequencing (WGS). Our analysis includes less individuals (n=379) compared to the Fujita study (n=424) as we had to exclude individuals with low quality genotyping. We consider inclusion of small structural variants (indels) worthwhile as structural variants are also identified as drivers of eQTL effects (Huang *et al.* 2015). We limited ourselves to eQTLs present in major cell types whereas Fujita also investigates eQTLs on cell subtype level.

In the Fujita study, a total of 5,678,710 variants was tested compared to 5,293,613 variants in our analysis, in this comparison we only consider variants that are located in the *cis*-region for at least one genes considered. In total 4,737,171 variants were present in both analyses (Figure 25). We expect the additional variants in the Fujita analysis to be variants that could be called with the WGS data but not with the genotyping data, and low MAF variants that do not pass the MAF filter for the subset of individuals we consider.

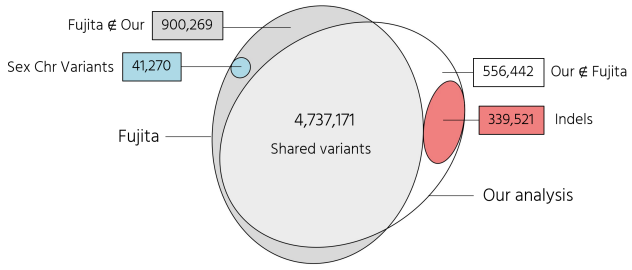



**Figure 25** Comparison of variants tested in our analysis versus Fujita et al. (2022) showed our analysis includes small SVs which were not considered in Fujita and in Fujita variants on sex chromosomes are considered.

We compared our results to the number of eQTLs, eSNPs and eGenes reported by the author (Table 7). As can be seen, in our analysis about twice (mean 1.96 (1.82-2.17)) the number of eQTLs and eSNPs were found compared to Fujita. Considering eGenes, approximately three-quarters (mean 0.74 (0.66-0.78)) of the number of eGenes found by Fujita were found in our replication study. The excess findings in our analysis on sc-eQTL and sc-eSNP level are contrary to what one would expect from an analysis with less individuals and less variants.

Many different approaches exist to multiple testing correction in eQTL studies and no best practice has been established, this unclarity is reflected in the variance observed in approaches between studies (Huang *et al.* 2018). We compare differences in the multiple testing strategy between our analysis and the Fujita study to see whether this explains the difference in results. In Fujita *et al.* (2022), significance of eQTLs, eSNPs and eGenes is determined using a 3-step hierarchical method[2]: **1)** Bonferroni correction of the nominal p-values by considering the number of tests per gene (local p-value), **2)** per gene, the minimum local p-value is corrected for false discovery rate using the Benjamini Hochberg procedure, and **3)** for eGenes that are significant (FDR $\leq$ 0.05) by step 2, eQTLs and eSNPs are determined by applying a threshold to the locally adjusted p-values based on the maximum globally adjusted p-value that has FDR $\leq$ 0.05.

On sc-eQTL and sc-eSNP level, we applied a different approach and controlled for false positives using the Benjamini-Hochberg method and considered eQTLs at FDR $\leq$ 0.05 as significant, taking the SNPs in the significant sc-eQTLs as sc-eSNPs. The method we applied is more lenient, in the sense that overall more eQTLs are considered significant. A major drawback is however that this does not account for the local structure of tests performed per gene and therefore favors results with low nominal p-values, which in turn may lead to missed discoveries.

The summary statistics for the Fujita study were made available on the Synapse portal (Synid: *syn52363777*) and we applied our strategy for multiple testing correction on the eQTL and eSNP level. This resulted in a dramatic change in number of findings (Table 8) with on average an excess 25% of eQTLs compared to our analysis which is to be expected from an analysis that has more statistical power because of the inclusion of more individuals.

On sc-eGene level, we have applied a similar multiple testing procedure to Fujita and attribute the lower number of discoveries in our analysis to the inclusion of a lower number of individuals in our analysis and testing less genes because we do not consider genes on sex chromosomes.

In conclusion, the large difference in findings can be mostly attributed to differences in multiple testing correction strategy between our analysis and the Fujita study. Even though the hierarchical method is very stringent and may falsely discard findings as Bonferroni correction on gene-level assumes independence of tests which is not the case, for future analysis we will consider using an adapted hierarchical approach instead to account for the local structure (linkage) when evaluating sc-eQTLs and sc-eSNPs where we currently only did this for sc-eGenes or consider taking into account linkage of sc-eSNPs through a clumping strategy.

[2] Multiple testing strategy based on email correspondence with the author, October 2023

| | Number of sc-eQTLs | | | Number of sc-eSNPs | | | Number of sc-eGenes | | |
|---|---|---|---|---|---|---|---|---|---|
| Cell type | This study | Fujita | Δ | This study | Fujita | Δ | This study | Fujita | Δ |
| Astrocyte | 799,682 | 406,890 | 392,792 | 600,552 | 321,885 | 278,667 | 2,511 | 3,350 | -839 |
| Endothelial cells | 16,095 | 7,388 | 8,707 | 15,161 | 7,347 | 7,814 | 51 | 77 | -26 |
| Excitatory neurons | 2,064,362 | 1,128,711 | 935,651 | 1,305,685 | 804,223 | 501,462 | 5,722 | 7,331 | -1,609 |
| Inhibitory neurons | 1,071,052 | 559,689 | 511,363 | 787,279 | 437,607 | 349,672 | 3,089 | 4,182 | -1,093 |
| Microglia | 207,966 | 100,578 | 107,388 | 181,847 | 89,080 | 92,767 | 666 | 899 | -233 |
| Oligodendrocytes | 776,375 | 411,615 | 364,760 | 581,892 | 319,215 | 262,677 | 2,416 | 3,115 | -699 |
| OPCs | 276,405 | 142,048 | 134,357 | 239,148 | 122,005 | 117,143 | 910 | 1,272 | -362 |

**Table 7** Number of eQTLs, eSNPs, and eGenes found in our analysis compared to Fujita *et al.* (2022)

| | Number of sc-eQTLs | | | | | Number of sc-eSNPs | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | This study | Fujita | Δ | Fujita* | Δ | This study | Fujita | Δ | Fujita* | Δ |
| Astrocyte | 799,682 | 406,890 | -49.1% | 976,413 | +22.1% | 600,552 | 321,885 | -46.4% | 722,721 | +22.1% |
| Endothelial | 16,095 | 7,388 | -54.1% | 22,224 | +38.1% | 15,161 | 7,347 | -51.5% | 20,584 | +38.1% |
| Excitatory neurons | 2,064,362 | 1,128,711 | -45.3% | 2,520,204 | +22.1% | 1,305,685 | 804,223 | -38.4% | 1,549,766 | +22.1% |
| Inhibitory neurons | 1,071,052 | 559,689 | -47.7% | 1,319,603 | +23.2% | 787,279 | 437,607 | -44.4% | 948,424 | +23.2% |
| Microglia | 207,966 | 100,578 | -51.6% | 263,249 | +26.6% | 181,847 | 89,080 | -51.0% | 225,933 | +26.6% |
| Oligodendrocytes | 776,375 | 411,615 | -47.0% | 920,337 | +18.5% | 581,892 | 319,215 | -45.1% | 680,337 | +18.5% |
| OPCs | 276,405 | 142,048 | -48.6% | 346,244 | +25.3% | 239,148 | 122,005 | -49.0% | 298,436 | +25.3% |

**Table 8** Number of eQTLs and eSNPs and eGenes found in our analysis compared to Fujita *et al.* (2022), Δshows the difference compared to our analysis. Fujita* shows the results after applying of our multiple testing strategy to the summary statistics of the Fujita study, showing that the difference in number of findings is driven by the difference in multiple testing strategy applied.

## Top 5 sc-eQTLs per cell type

| Cell type | Chr | Pos (hg38) | rsid | Allele Counted | Alt | Gene | Beta | P-value (nominal) |
|---|---|---|---|---|---|---|---|---|
| Astrocyte | 11 | 74848863 | rs10751241 | G | T | XRRA1 | -1.28 | $5.43\times10^{105}$ |
| | 1 | 205820154 | rs823080 | G | A | AC119673.2 | -1.24 | $9.56\times10^{-96}$ |
| | 11 | 30311469 | rs7926465 | C | T | AL358944.1 | 1.31 | $4.28\times10^{-86}$ |
| | 8 | 62256176 | rs4738948 | A | G | AC023095.1 | 1.03 | $5.24\times10^{-85}$ |
| | 15 | 45264631 | rs11631021 | A | G | AC051619.5 | -0.93 | $2.17\times10^{-83}$ |
| Endothelial | 7 | 71220553 | rs35853756 | T | C | GALNT17 | 0.92 | $2.32\times10^{-43}$ |
| | 12 | 56042145 | rs1131017 | C | G | RPS26 | 0.80 | $8.14\times10^{-32}$ |
| | 3 | 147317552 | rs7639019 | C | G | AC092957.1 | -0.95 | $2.01\times10^{-30}$ |
| | 6 | 31268690 | rs2524096 | G | T | HLA-C | 0.66 | $7.75\times10^{-30}$ |
| | 6 | 29951244 | rs9260415 | A | C | HLA-A | -0.59 | $2.87\times10^{-24}$ |
| Excitatory neurons | 8 | 70778944 | rs201801376 | C | T | XKR9 | -1.19 | $8.74\times10^{-144}$ |
| | 11 | 74892136 | rs4944963 | G | A | XRRA1 | -1.28 | $9.37\times10^{-115}$ |
| | 8 | 70805871 | rs62508819 | T | C | AC022858.1 | -1.22 | $1.30\times10^{-114}$ |
| | 17 | 2037754 | rs12451788 | A | G | DPH1 | 1.05 | $2.39\times10^{-110}$ |
| | 7 | 26399032 | rs1229663 | C | G | AC004540.1 | 1.16 | $5.87\times10^{-106}$ |
| Inhibitory neurons | 11 | 74888549 | rs59143825 | C | T | XRRA1 | -1.29 | $3.13\times10^{-107}$ |
| | 12 | 75473785 | rs7969930 | T | C | GLIPR1L2 | 1.01 | $1.38\times10^{-96}$ |
| | 8 | 70781284 | rs6993170 | A | T | XKR9 | -1.13 | $1.20\times10^{-94}$ |
| | 7 | 26399032 | rs1229663 | C | G | AC004540.1 | 1.14 | $4.32\times10^{-93}$ |
| | 10 | 58629241 | rs2028205 | A | G | BICC1 | -0.92 | $4.88\times10^{-91}$ |
| Microglia | 13 | 50715893 | rs2796882 | T | G | DLEU1 | 1.27 | $5.63\times10^{-85}$ |
| | 13 | 50718901 | rs2761843 | T | C | DLEU7 | 1.14 | $7.47\times10^{-83}$ |
| | 11 | 74933260 | rs7102619 | C | T | XRRA1 | -1.31 | $2.44\times10^{-79}$ |
| | 7 | 28137682 | rs11495981 | C | T | JAZF1 | -1.09 | $8.16\times10^{-78}$ |
| | 4 | 141208286 | rs13111980 | T | C | RNF150 | 1.25 | $7.46\times10^{-74}$ |
| Oligodendrocytes | 11 | 74848863 | rs10751241 | G | T | XRRA1 | -1.36 | $3.18\times10^{-97}$ |
| | 9 | 12745644 | rs2209271 | G | A | LURAP1L | -1.22 | $1.66\times10^{-96}$ |
| | 9 | 12759782 | rs1326788 | C | T | LURAP1L-AS1 | -1.18 | $5.99\times10^{-91}$ |
| | 1 | 205813590 | rs9438393 | A | G | AC119673.2 | -1.22 | $1.66\times10^{-87}$ |
| | 4 | 92566314 | rs6532377 | C | A | GRID2 | -1.32 | $3.90\times10^{-86}$ |
| OPCs | 11 | 74842789 | rs2298746 | C | T | XRRA1 | -1.25 | $2.27\times10^{-86}$ |
| | 8 | 62256176 | rs4738948 | A | G | AC023095.1 | 1.19 | $3.04\times10^{-80}$ |
| | 8 | 96692367 | rs72682307 | A | G | CPQ | 1.25 | $2.31\times10^{-66}$ |
| | 1 | 72346757 | rs2815752 | G | A | AL513166.1 | -0.99 | $8.15\times10^{-63}$ |
| | 12 | 75472588 | rs67705855 | A | G | GLIPR1L2 | 1.02 | $2.02\times10^{-62}$ |

**Table 9** Top 5 sc-eQTLs per cell type. P-value reported is the nominal p-value from the test-statistic. We kept the sc-eQTL (SNP-gene pair) with lowest nominal p-value per gene for display of the top results to avoid the top results showing multiple eQTLs for the same gene with multiple SNPs in linkage

## Top co-eQTLs per cell type

| # | Chr : pos (hg38) | rsid | Alleles | GT | GeneA | Chr | eQTL | GeneB | Chr | eQTL | DC | △cor | JSD | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 : 12229967 | rs13237518 | C-A | 0-2 | PLEKHA1 | 10 | | SLC25A4 | 4 | x | | 0.103 | 0.172 | 1.67E-04 |
| 2 | 7 : 12229967 | rs13237518 | C-A | 0-2 | KRAS | 12 | x | PLEKHA1 | 10 | | | 0.016 | 0.159 | 1.67E-04 |
| 3 | 7 : 12229967 | rs13237518 | C-A | 0-2 | CTNNB1 | 3 | | IRS2 | 13 | | x | 0.523 | 0.158 | 1.67E-04 |
| 4 | 7 : 12229967 | rs13237518 | C-A | 0-2 | PIK3CA | 3 | | SLC25A4 | 4 | x | | 0.164 | 0.153 | 1.67E-04 |
| 5 | 7 : 12229967 | rs13237518 | C-A | 0-2 | HS3ST5 | 6 | | KRAS | 12 | x | | 0.099 | 0.150 | 1.67E-04 |
| 6 | 7 : 12229967 | rs13237518 | C-A | 0-2 | ATG2B | 14 | | KRAS | 12 | x | x | 0.316 | 0.150 | 1.67E-04 |
| 7 | 7 : 12229967 | rs13237518 | C-A | 0-2 | HS3ST5 | 6 | | SLC25A4 | 4 | x | | 0.080 | 0.148 | 2.67E-04 |
| 8 | 7 : 12229967 | rs13237518 | C-A | 0-2 | ATG2B | 14 | | SLC25A4 | 4 | x | | 0.101 | 0.148 | 2.67E-04 |
| 9 | 7 : 12229967 | rs13237518 | C-A | 0-2 | EIF2AK2 | 2 | x | PSMD1 | 2 | | | 0.069 | 0.147 | 2.67E-04 |
| 10 | 7 : 12229967 | rs13237518 | C-A | 0-2 | ITPR1 | 3 | | SLC25A4 | 4 | x | | 0.105 | 0.147 | 2.67E-04 |

**Table 10** Top 10 co-eQTL for excitatory neurons based on significance. **Alleles:** counted allele-alternative allele **GT:** Genotype groups between which there is a significant change in co-expression for GeneA and GeneB (0: 2x alt allele - 1: 1x alt allele + 1 x counted allele - 2: 2 x counted allele) / **eQTL:** marked when co-eQTL involves significant eQTL effect of the variant on either or both geneA and geneB / **DC:** checked when co-eQTL involves significant differential correlation / **|△|:** absolute difference in Pearson correlation of expression of geneA and geneB between the genotypes / **JSD:** distance (Jensen-Shannon divergence) of co-expression density between the genotype groups / **FDR:** false discovery rate calculated with p-value based on null distribution determined per variant

| # | Chr : pos (hg38) | rsid | Alleles | GT | GeneA | Chr | eQTL | GeneB | Chr | eQTL | DC | \|△\| cor | JSD | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 : 12229967 | rs13237518 | C-A | 0-2 | COX7C | 5 | | CTNNB1 | 3 | | x | 0.311 | 0.141 | 1.10E-02 |
| 2 | 7 : 12229967 | rs13237518 | C-A | 0-2 | CTNNB1 | 3 | | UQCRB | 8 | | | 0.263 | 0.133 | 1.40E-02 |
| 3 | 7 : 12229967 | rs13237518 | C-A | 0-2 | CTNNB1 | 3 | | PPP3CC | 8 | | | 0.261 | 0.131 | 1.40E-02 |
| 4 | 7 : 12229967 | rs13237518 | C-A | 0-2 | CTNNB1 | 3 | | PSMD1 | 2 | | | 0.002 | 0.129 | 1.40E-02 |
| 5 | 7 : 12229967 | rs13237518 | C-A | 0-2 | BIN1 | 2 | | CTNNB1 | 3 | | | 0.079 | 0.127 | 1.40E-02 |
| 6 | 7 : 12229967 | rs13237518 | C-A | 0-2 | ATG2B | 14 | | CTNNB1 | 3 | | | 0.030 | 0.127 | 1.40E-02 |
| 7 | 7 : 12229967 | rs13237518 | C-A | 0-2 | CTNNB1 | 3 | | NDUFS1 | 2 | | | 0.190 | 0.127 | 1.40E-02 |
| 8 | 7 : 12229967 | rs13237518 | C-A | 0-2 | PPP3CC | 8 | | RTN3 | 11 | | x | 0.352 | 0.127 | 1.40E-02 |
| 9 | 7 : 12229967 | rs13237518 | C-A | 0-2 | CTNNB1 | 3 | | PIK3CB | 3 | | | 0.144 | 0.127 | 1.40E-02 |
| 10 | 7 : 12229967 | rs13237518 | C-A | 0-2 | CTNNB1 | 3 | | UQCRH | 1 | | | 0.273 | 0.127 | 1.40E-02 |

**Table 11** Top 10 co-eQTL for inhibitory neurons based on significance. **Alleles:** counted allele-alternative allele **GT:** Genotype groups between which there is a significant change in co-expression for GeneA and GeneB (0: 2x alt allele - 1: 1x alt allele + 1 x counted allele - 2: 2 x counted allele) / **eQTL:** marked when co-eQTL involves significant eQTL effect of the variant on either or both geneA and geneB / **DC:** checked when co-eQTL involves significant differential correlation / **|△|:** absolute difference in Pearson correlation of expression of geneA and geneB between the genotypes / **JSD:** distance (Jensen-Shannon divergence) of co-expression density between the genotype groups / **FDR:** false discovery rate calculated with p-value based on null distribution determined per variant

| # | Chr : pos (hg38) | rsid | Alleles | GT | GeneA | Chr | eQTL | GeneB | Chr | eQTL | DC | \|△\| cor | JSD | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | NCK2 | 2 | | | 0.231 | 0.543 | 3.03E-05 |
| 2 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | ADAM10 | 15 | | JAZF1 | 7 | x | | 0.257 | 0.531 | 3.03E-05 |
| 3 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | RTN3 | 11 | | | 0.342 | 0.526 | 3.03E-05 |
| 4 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | GNAQ | 9 | | JAZF1 | 7 | x | | 0.107 | 0.525 | 3.03E-05 |
| 5 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | MAF | 16 | | | 0.334 | 0.517 | 3.03E-05 |
| 6 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | MAP3K5 | 6 | | x | 0.495 | 0.517 | 3.03E-05 |
| 7 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | CALM2 | 2 | | JAZF1 | 7 | x | | 0.025 | 0.513 | 3.03E-05 |
| 8 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | SORL1 | 11 | | | 0.162 | 0.512 | 3.03E-05 |
| 9 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | AKT3 | 1 | | JAZF1 | 7 | x | | 0.191 | 0.511 | 3.03E-05 |
| 10 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | GSK3B | 3 | | JAZF1 | 7 | x | | 0.044 | 0.507 | 3.03E-05 |

**Table 12** Top 10 co-eQTL for microglia based on significance. **Alleles:** counted allele-alternative allele **GT:** Genotype groups between which there is a significant change in co-expression for GeneA and GeneB (0: 2x alt allele - 1: 1x alt allele + 1 x counted allele - 2: 2 x counted allele) / **eQTL:** marked when co-eQTL involves significant eQTL effect of the variant on either or both geneA and geneB / **DC:** checked when co-eQTL involves significant differential correlation / **|△|:** absolute difference in Pearson correlation of expression of geneA and geneB between the genotypes / **JSD:** distance (Jensen-Shannon divergence) of co-expression density between the genotype groups / **FDR:** false discovery rate calculated with p-value based on null distribution determined per variant

| # | Chr : pos (hg38) | rsid | Alleles | GT | GeneA | Chr | eQTL | GeneB | Chr | eQTL | DC | \|Δ\| cor | JSD | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 17 : 1728046 | rs35048651 | TGAG-T | 1-2 | PPP3R1 | 2 | | RB1CC1 | 8 | | | 0.064 | 0.113 | 1.30E-02 |
| 2 | 2 : 37304796 | rs17020490 | T-C | 1-2 | PPP3R1 | 2 | | RB1CC1 | 8 | | | 0.117 | 0.101 | 2.00E-02 |
| 3 | 8 : 27607795 | rs11787077 | T-C | 0-2 | CLU | 8 | x | RTN3 | 11 | | | 0.208 | 0.136 | 2.68E-02 |
| 4 | 8 : 27607795 | rs11787077 | T-C | 0-2 | BRAF | 7 | | CLU | 8 | x | | 0.207 | 0.134 | 2.68E-02 |
| 5 | 8 : 27607795 | rs11787077 | T-C | 0-2 | CLU | 8 | x | PPP3CC | 8 | | | 0.086 | 0.133 | 2.68E-02 |
| 6 | 8 : 27607795 | rs11787077 | T-C | 0-2 | APP | 21 | | CLU | 8 | x | | 0.018 | 0.132 | 2.68E-02 |
| 7 | 8 : 27607795 | rs11787077 | T-C | 0-2 | CLU | 8 | x | CTNNB1 | 3 | | | 0.244 | 0.130 | 2.88E-02 |
| 8 | 10 : 80494228 | rs6586028 | C-T | 1-2 | MAPK1 | 22 | | RTN4 | 2 | | | 0.035 | 0.252 | 3.10E-02 |
| 9 | 8 : 27607795 | rs11787077 | T-C | 0-2 | ATP2A2 | 12 | | CLU | 8 | x | | 0.299 | 0.127 | 3.77E-02 |
| 10 | 8 : 27607795 | rs11787077 | T-C | 0-2 | CLU | 8 | x | PIK3R3 | 1 | | | 0.246 | 0.127 | 3.77E-02 |

**Table 13** Top 10 co-eQTL for endothelial cells based on significance. **Alleles:** counted allele-alternative allele **GT:** Genotype groups between which there is a significant change in co-expression for GeneA and GeneB (0: 2x alt allele - 1: 1x alt allele + 1 x counted allele - 2: 2 x counted allele) / **eQTL:** marked when co-eQTL involves significant eQTL effect of the variant on either or both geneA and geneB / **DC:** checked when co-eQTL involves significant differential correlation / **|Δ|:** absolute difference in Pearson correlation of expression of geneA and geneB between the genotypes / **JSD:** distance (Jensen-Shannon divergence) of co-expression density between the genotype groups / **FDR:** false discovery rate calculated with p-value based on null distribution determined per variant

| # | Chr : pos (hg38) | rsid | Alleles | GT | GeneA | Chr | eQTL | GeneB | Chr | eQTL | DC | \|Δ\| cor | JSD | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 9 : 104903697 | rs1800978 | C-G | 1-2 | GNAQ | 9 | | TUBB4A | 19 | | | 0.183 | 0.084 | 1.00E-03 |
| 2 | 5 : 14724304 | rs112403360 | T-A | 1-2 | CLU | 8 | | RTN3 | 11 | | x | 0.608 | 0.128 | 1.70E-02 |
| 3 | 15 : 58764824 | rs602602 | T-A | 1-2 | ADAM10 | 15 | x | RTN4 | 2 | | | 0.023 | 0.056 | 2.00E-02 |
| 4 | 9 : 104903697 | rs1800978 | C-G | 1-2 | CALM2 | 2 | | GNAQ | 9 | | | 0.087 | 0.071 | 3.00E-02 |
| 5 | 9 : 104903697 | rs1800978 | C-G | 1-2 | PLEKHA1 | 10 | | RTN3 | 11 | | | 0.059 | 0.069 | 3.33E-02 |
| 6 | 8 : 11844613 | rs1065712 | G-C | 1-2 | AKT3 | 1 | | MAPK10 | 4 | | | 0.028 | 0.128 | 3.70E-02 |
| 7 | 9 : 104903697 | rs1800978 | C-G | 1-2 | BIN1 | 2 | | GNAQ | 9 | | | 0.112 | 0.068 | 4.55E-02 |
| 8 | 9 : 104903697 | rs1800978 | C-G | 1-2 | CSNK1A1 | 5 | | GNAQ | 9 | | | 0.074 | 0.066 | 4.77E-02 |
| 9 | 9 : 104903697 | rs1800978 | C-G | 1-2 | GNAQ | 9 | | PSEN1 | 14 | | x | 0.287 | 0.066 | 4.77E-02 |
| 10 | 14 : 92464917 | rs7401792 | G-A | 0-2 | CSNK1A1 | 5 | | PIK3R1 | 5 | | | 0.057 | 0.123 | 4.90E-02 |

**Table 14** Top 10 co-eQTL for oligodendrocytes based on significance. **Alleles:** counted allele-alternative allele **GT:** Genotype groups between which there is a significant change in co-expression for GeneA and GeneB (0: 2x alt allele - 1: 1x alt allele + 1 x counted allele - 2: 2 x counted allele) / **eQTL:** marked when co-eQTL involves significant eQTL effect of the variant on either or both geneA and geneB / **DC:** checked when co-eQTL involves significant differential correlation / **|Δ|:** absolute difference in Pearson correlation of expression of geneA and geneB between the genotypes / **JSD:** distance (Jensen-Shannon divergence) of co-expression density between the genotype groups / **FDR:** false discovery rate calculated with p-value based on null distribution determined per variant

| # | Chr : pos (hg38) | rsid | Alleles | GT | GeneA | Chr | eQTL | GeneB | Chr | eQTL | DC | \|Δ\| cor | JSD | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | RTN4 | 2 | | | 0.193 | 0.303 | 6.02E-03 |
| 2 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | PIK3CA | 3 | | | 0.129 | 0.266 | 6.02E-03 |
| 3 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | LRP1 | 12 | | | 0.256 | 0.265 | 6.02E-03 |
| 4 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | ANK3 | 10 | | JAZF1 | 7 | x | x | 0.470 | 0.262 | 6.02E-03 |
| 5 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | MAP3K5 | 6 | | | 0.345 | 0.262 | 6.02E-03 |
| 6 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | PIK3R1 | 5 | | | 0.378 | 0.262 | 6.02E-03 |
| 7 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | PPP3CC | 8 | | | 0.059 | 0.260 | 6.02E-03 |
| 8 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | CLU | 8 | | JAZF1 | 7 | x | | 0.185 | 0.258 | 6.02E-03 |
| 9 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | GAPDH | 12 | | JAZF1 | 7 | x | | 0.177 | 0.257 | 6.02E-03 |
| 10 | 7 : 28129126 | rs1160871 | GTCTT-G | 0-2 | JAZF1 | 7 | x | MAPK8 | 10 | | | 0.187 | 0.257 | 6.02E-03 |

**Table 15** Top 10 co-eQTLs for OPCs based on significance. **Alleles:** counted allele-alternative allele **GT:** Genotype groups between which there is a significant change in co-expression for GeneA and GeneB (0: 2x alt allele - 1: 1x alt allele + 1 x counted allele - 2: 2 x counted allele) / **eQTL:** marked when co-eQTL involves significant eQTL effect of the variant on either or both geneA and geneB / **DC:** checked when co-eQTL involves significant differential correlation / **|Δ|:** absolute difference in Pearson correlation of expression of geneA and geneB between the genotypes / **JSD:** distance (Jensen-Shannon divergence) of co-expression density between the genotype groups / **FDR:** false discovery rate calculated with p-value based on null distribution determined per variant

## Variant level summary of co-eQTLs identified per cell type

| CT | Variant | rsid | # | # eQTL | # DC | GT |
|---|---|---|---|---|---|---|
| End | chr8:27607795:T:C | rs11787077 | 14 | 14 | 2 | 0-2 |
| End | chr10:80494228:C:T | rs6586028 | 1 | 0 | 0 | 1-2 |
| End | chr17:1728046:TGAG:T | rs35048651 | 1 | 0 | 0 | 1-2 |
| End | chr2:37304796:T:C | rs17020490 | 1 | 0 | 0 | 1-2 |
| Exc | chr7:12229967:C:A | rs13237518 | 6532 | 3012 | 601 | 0-2 |
| Exc | chr7:7817263:T:C | rs6943429 | 48 | 0 | 13 | 0-2 |
| Exc | chr9:104903697:C:G | rs1800978 | 4 | 0 | 1 | 1-2 |
| Exc | chr15:50701814:A:G | rs8025980 | 1 | 0 | 1 | 0-2 |
| Exc | chr17:1728046:TGAG:T | rs35048651 | 1 | 0 | 1 | 1-2 |
| Exc | chr2:233117202:G:C | rs10933431 | 1 | 0 | 1 | 0-2 |
| Inh | chr7:12229967:C:A | rs13237518 | 95 | 5 | 13 | 0-2 |
| Inh | chr7:7817263:T:C | rs6943429 | 8 | 0 | 2 | 0-2 |
| Inh | chr11:47370397:G:A | rs10437655 | 1 | 0 | 1 | 0-2 |
| Inh | chr17:1728046:TGAG:T | rs35048651 | 1 | 0 | 0 | 1-2 |
| Inh | chr2:127135234:C:T | rs6733839 | 1 | 0 | 0 | 0-2 |
| Mic | chr7:28129126:GTCTT:G | rs1160871 | 38 | 33 | 3 | 0-2 |
| Mic | chr14:92472511:G:A | rs12590654 | 35 | 33 | 5 | 0-2 |
| Mic | chr11:47370397:G:A | rs10437655 | 33 | 0 | 2 | 0-1 |
| Mic | chr2:37304796:T:C | rs17020490 | 8 | 0 | 1 | 1-2 |
| Mic | chr7:8204382:T:C | rs10952097 | 2 | 0 | 0 | 0-1 |
| Mic | chr15:50701814:A:G | rs8025980 | 1 | 0 | 1 | 0-1 |
| Oli | chr9:104903697:C:G | rs1800978 | 6 | 0 | 1 | 1-2 |
| Oli | chr14:92464917:G:A | rs12590654 | 1 | 0 | 0 | 0-2 |
| Oli | chr15:58764824:T:A | rs602602 | 1 | 1 | 0 | 1-2 |
| Oli | chr5:14724304:T:A | rs112403360 | 1 | 0 | 1 | 1-2 |
| Oli | chr8:11844613:G:C | rs1065712 | 1 | 0 | 0 | 1-2 |
| OPC | chr7:28129126:GTCTT:G | rs1160871 | 41 | 41 | 4 | 0-2 |

**Table 16** Summary results of significant co-eQTLS in all cell types, no significant co-eQTLs were found in astrocytes. **CT:** cell type / **Variant:** chromosome:position_hg38:counted_allele:alternative_allele **#** : number of significant co-eQTLs for the variant / **# DC:** number of co-eQTLs subject to differential correlation / **GT:** genotype groups involved in the co-eQTL - 0: 2x alternative allele - 1: 1x alternative allele + 1 x counted allele - 2: 2 x counted allele