A SINGLE-NUCLEUS RNA SEQUENCING ANALYSIS OF HERVS AND LINE1S IN ALZHEIMER'S AND PARKINSON'S DISEASE

By Stein Acker

June 2022
Co-Supervisors: Raquel Garza and Johan Jakobsson
BINP52 (60 credits)

LUND
UNIVERSITY

Abstract

Transposable elements (TEs) are mobile genetic elements that make up roughly 50% of the human genome, with retrotransposons (or transposons which retrotranspose via an RNA intermediate) making up the vast majority of these elements. Recent studies have suggested that TEs could play a role in many neurological conditions, including ALS and dementia. The specific reasons why TEs are associated with these conditions remain somewhat unclear, with potential explanations including mutagenic insertion, immune responses to the presence of transcripts, and cytotoxic peptides resulting from the translation of these transcripts. We performed single-nuclei RNA sequencing (snRNA-seq) to profile the impact that TEs have on the development of Alzheimer's disease (AD) and Parkinson's disease (PD). Our results show clear evidence for activated microglia and reactive astrocytes in PD, suggesting neuroinflammation. Additionally, using trusTEr, a new bioinformatics pipeline for the quantification of TE expression from single nuclei sequencing datasets, we were able to identify cell type-specific expression patterns of LINE1s and HERVs between diseased and control brains, which raise questions about the functional consequences of an aberrant expression of TEs in the human brain and their role in neuroinflammation.

Popular summary

Protein-coding genes – the "important" bits of DNA that contain instructions for our cellular machinery – make up just 2% of our genome. The other 98%, about half of which is a group of ancient viral genes and sequences which have mutated to replicate themselves, have historically been considered irrelevant to study. However, recent research has shown that much, though not all, of this so-called "junk DNA" actually gets copied into RNA fairly frequently and has some influence on nearby areas of the genome. Might it impact human health?

Modern bioinformatics research suggests this may be the case. Retrotransposons, which are DNA sequences that have (or had) the ability to retrotranspose (or create copies of themselves throughout our genetic code using RNA intermediates) make up nearly half our genome but have not been researched so much until recently due to historical constraints in technology. However, what research does exist suggests that RNA intermediates created by retrotransposons can impact brain health in some important ways and are associated with diseases like dementia. For this reason, we chose to research two types of retrotransposons – human endogenous retroviruses (HERVs), which are insertions of viral genetic code which still sometimes create RNA intermediates but never retrotranspose anymore, and long interspersed nuclear element 1s (LINE1s, L1s), which are the only retrotransposons that can still retrotranspose in humans – and their impacts on two common neurodegenerative diseases, Alzheimer's disease (AD) and Parkinson's disease (PD). We wanted to see if the expression of these retrotransposons differed between cell types as well, so we used a method called single-nucleus RNA sequencing (snRNA-seq), which let us examine all the RNA produced by each individual cell and identify cell types based on that.

We found that while LINE1s did not show particularly different expression patterns in brains with AD or PD and healthy brains, HERVs showed some interesting patterns. Perhaps most interesting was the fact that a HERV associated with inflammation was much more expressed in PD microglia than in control microglia. This is especially interesting because microglia are a type of brain cell that forms part of the immune system and causes inflammation, and neuroinflammation is an important factor in the development of many neurodegenerative disorders, including PD. We also found that many genes associated with inflammation were highly expressed in these microglia, providing compelling evidence that HERVs in microglia may be involved in the inflammation seen in PD.

Despite these findings, there is much that we still cannot explain. For instance, we are not certain whether the inflammatory response seen is a result of the RNA intermediates from these HERVs or from the viral proteins these intermediates code for – or even whether this difference in expression is a cause or an effect of neuroinflammation. That said, these findings provide exciting opportunities for future research in the field of neuroinflammation and retrotransposon activity.

Introduction

Transposable elements

Transposable elements (TEs) are mobile genetic elements that can relocate throughout the genome. TEs comprise roughly 50% of the human genome, as opposed to protein-coding genes, which make up just 2% of the human genome (1). They are divided into two classes: class I, or retrotransposons, and class II, or DNA transposons (1–3). Retrotransposons function through a copy-and-paste mechanism, wherein an RNA intermediate is created through the cell's RNA transcription machinery, and has its information written into the genome through reverse transcription at a different locus (1–3). Conversely, DNA transposons function through a cut-and-paste mechanism in which a region of DNA uses cellular machinery to cleave itself from its current locus and insert itself into a different locus (1–3).

Between the two classes, retrotransposons make up a far greater share of the human genome than DNA transposons, which are no longer active in humans but remain active in some other organisms (1,3). Within the retrotransposon class, elements are classified into those containing long-terminal repeat (LTR) elements and those which do not (1,4). Among non-LTRs, elements are further classified into long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and other families (Figure 1) (1,4).

LINEs

Of all retrotransposons, LINE elements are the most plentiful and are further divided into more-common LINE1 (L1) and less-common LINE2 (L2) elements. L1s comprise roughly 17% of the modern human genome (5,6). Despite the incredible amount of the human genome made up of

L1s, only 80-100 L1s have retained the ability to retrotranspose, with 84% of all retrotransposition activity in the modern human genome coming from a handful of L1HS elements (7,8). The fact that L1s have the ability to retrotranspose makes them important to understand, as their retrotransposition has the potential to cause mutagenic insertions in protein coding genes, thus causing a wide variety of health problems including neurodegenerative diseases (1).

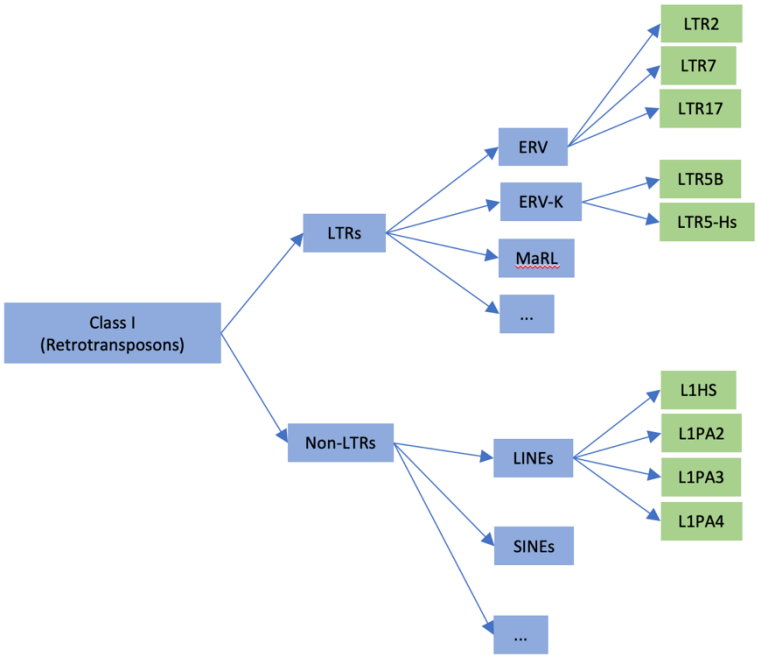

Figure 1: Classification of different retrotransposons. The retrotransposons included in this study are highlighted in green.

L1s are comprised of five regions: a 5' untranslated region (UTR), three open reading frames (ORFs) called ORF0, ORF1, and ORF2, a 3' UTR, and a poly-A tail on the 3' end (6). ORF1 codes for an RNA binding protein, while ORF2 encodes for a reverse transcriptase (6). The role of ORF0 remains somewhat unclear, as it was only discovered quite recently, but it is suspected

to enhance the expression of its respective L1 (6). It is known to contain an antisense promoter, which could potentially initiate the expression of downstream genes (9).

LTRs

In humans, nearly all LTR elements are endogenous retroviruses (ERVs), a family of genetic elements thought to have originated with genomic insertions from retroviruses far back in the human lineage (10). ERVs which are found in humans are referred to as human ERVs (HERVs). Structurally, HERVs are fairly similar to the genetic materials of other retroviruses, such as HIV. They are comprised of four coding regions – *gag*, *pro*, *pol*, and *env* – which code for structural viral proteins, such as viral capsids and reverse transcriptase (11). These regions are flanked on the 5' and 3' ends by long terminal repeats (11). HERVs make up a significant proportion of the human genome – roughly 8% – but none are still able to retrotranspose (12). However, they may still generate transcripts in the cell (11–13). Increased expression of HERVs has been shown to be associated with a variety of neurodegenerative diseases, such as Alzheimer's disease (AD) (13,14).

Neurodegenerative diseases

Age-related neurodegenerative diseases are an increasingly urgent public health issue as general health improves throughout the world, causing lifespans grow. In 2010, 35.6 million people in the world had dementia, the most common age-related neurodegenerative disorder; this number is expected to grow to 115.4 million by 2050 (15). Parkinson's disease (PD) is the second most common neurodegenerative disease after dementia – affecting an estimated 3% of individuals older than 90 – making its study extremely relevant in the context of an aging population (16).

Studying both dementia and PD is crucial to help better understand and improve geriatric health in the future.

Alzheimer's disease

The most common form of dementia is AD (17). Current research suggests that it is caused by insoluble clumps of hyperphosphorylated tau protein, which are called "neurofibrillary tangles" (17,18). These neurofibrillary tangles lead to cell death in the prefrontal cortex, which then results in degradation of cognition and memory (17).

Although it is well understood that the presence of tau plaques and neurofibrillary tangles in the brain lead to the development of AD symptoms, the etiology of the disease remains unclear. One hypothesis postulates that vascular dysregulation in the cerebral cortex is one of the primary contributors to the disease (19–21). In this hypothesis, degradation of the blood-brain barrier is the driving biological mechanic behind the disease (21). This degradation results in neuroinflammation and hypoxia, resulting in neuronal cell death (21). Recent research has found that a lack of endothelial progenitor cells may be the cause of this vascular stress in the first place (19).

Parkinson's disease

PD is an age-related neurological condition in which the primary symptoms are related to motion control (17). Its proximate cause is the death of dopaminergic excitatory neurons in the substantia nigra (22). It may occur with or without dementia symptoms, resulting from the

buildup of alpha-synuclein (a-Syn) in the prefrontal cortex called “Lewy bodies,” meaning the substantia nigra and prefrontal cortex are both very relevant areas of the brain to study in PD (16,23).

Inflammation in neurodegenerative diseases

Despite their prevalence, the causes of age-related neurodegenerative diseases remain poorly understood. However, several hypotheses have emerged, one of the foremost being that neuroinflammation is an important cause (24). Research has shown, for instance, that individuals with PD have significantly increased levels of inflammatory cytokines in their blood (25). Additionally, AD progression is well-known to be associated with neuroinflammation, and anti-inflammatory treatments may be used to slow the progression of the disease (26,27). Transposable elements (TEs) are known to cause inflammatory responses in hosts and are known to be associated with a wide range of neurodegenerative disorders, which suggests a possible link between TEs and the diseases studied here (1).

The role of TEs in neurodegenerative diseases

The “transposition theory of aging” postulates that the regulation of retrotransposons becomes more dysfunctional as an individual ages, a pattern which has been found in organisms such as *Drosophila* and mice (28–31). In humans, DNA methylation patterns are known to change with aging – a process that could cause the upregulation of formerly silenced TEs as the genomic regions containing the TE code become demethylated (32,33). Additionally, TAU protein resulting from AD is suspected to impact the expression of heterochromatin protein 1 (HP1), an epigenetic regulator, potentially affecting the regulation of TE expression (34). Moreover, in a

prior study on the relationship between TAU buildup and TE expression, researchers found a positive correlation between the amount of tau in the brains of Alzheimer's patients and the expression of TEs (31). This is relevant to our study, as retrotransposons have the potential to cause myriad deleterious health effects.

One of the most obvious traits of TEs that can impact human health is the mutagenic insertion of TE copies in other genes (1). Although most TEs have developed enough mutations over time to become transcriptionally inactive, some remain active (1,35). However, other health impacts may occur from the transcription of TEs. One such impact is the creation of cytotoxic peptides from TE transcripts (1,14). Additionally, the presence of TE transcripts in the nucleus as noncoding RNA has the potential to impact regulatory processes within the cell (1,36).

As TEs have a long heritage in our genomes, our cells have evolved methods to control their transcription. One example of this defense mechanisms is TRIM28, an epigenetic co-repressor (12,37). In the TRIM28 repressive complex, Krüppel-associated box-zinc finger proteins (KRAB-ZFPs) bind to the TE sequence and recruit Trim28, which then acts as a scaffold for other proteins such as SETDB1 and HP1 to epigenetically repress TEs through DNA methylation (12,37–39). Prior studies have suggested that this mechanism may cause TRIM28 to epigenetically silence nearby genes as well (37).

In contrast, some studies have also shown that TEs may act as alternative promoters or enhancers for nearby genes and thus result in greater transcriptional activity in their immediate surroundings, showing how varied the roles of TEs are in the human genome (1,40).

One crucial factor about TEs to understand when studying their role in neuroinflammation is that TE transcripts and peptides can often be mistaken for viral fragments by the cell. As a result of this similarity, a leading hypothesis on why TEs appear so correlated to neuroinflammatory diseases is the accumulation of RNA transcripts, RNA:DNA hybrids, and extra-chromosomal DNA, which can result from retrotransposon transcription (2). The presence of these TE-related factors alerts the immune pathways within the cell, leading to inflammation (4).

Bioinformatics considerations

From a bioinformatics standpoint, TEs can be rather difficult to study. Perhaps the most prominent reason for this difficulty is the amount of repetition in TE-infiltrated regions (41). Sequence repetition means that mapping TEs to unique loci in the genome is very challenging. Additionally, the fact that TE families tend to be quite homogeneous means that separating two related subfamilies (for instance, L1HS and L1PA2) can be difficult, particularly when performing short-read sequencing. This difficulty is further compounded when performing single-cell analysis, as the number of reads for each cell is quite small when compared to bulk sequencing. Considering these challenges, we chose to pool similar cell transcriptomes together when quantifying TEs, analyze TE expression on a per-subfamily level rather than a per-element level, and use a software specifically developed to quantify TE reads rather than traditional gene-quantification software, ensuring that the reads quantified are not limited to those that can be mapped to a specific locus (42).

Project aims

Based on prior research, we believe it would be relevant to further investigate the relationship between TEs and neurodegenerative disorders (1,14,30,31). Many studies have identified a relationship between neurodegenerative diseases and TE expression; however, few have profiled the roles that different cell types play in this relationship (14,30,31). Therefore, we have chosen to use single-nuclei RNA sequencing (snRNA-seq) rather than bulk sequencing. This approach will allow us to characterize TE expression in diseased and control samples on a per-cell type basis. The aim of this thesis is to profile the expression of four L1 subfamilies (L1HS, L1PA2, L1PA3, and L1PA4) and five HERV subfamilies (LTR2, LTR5B, LTR5-Hs, LTR7, and LTR17) which have been previously shown to activate the immune response or been implicated in inflammation in the substantia nigra and prefrontal cortex of PD and prefrontal cortex of AD (43,44). Additionally, we will assess astrocytes for eight biomarkers of astrocyte reactivity (CHI3L1, C3, S100B, CRYAB, MAOB, NFAT5, HSPB1, and MT2A) and microglia for eight biomarkers of microglial activation (FTL, SPP1, APOE, CD74, FCGR3A, CST3, CSF1R, and PTPRC) (45–49) in each region. This will allow us to assess the inflammatory statuses as well as the TE expression.

Results

Experimental setup

To investigate the difference in expression of different TE subfamilies, the analysis was performed using 17 postmortem samples with PD (9 prefrontal cortex, 8 substantia nigra) and 14 normal control (NC) postmortem samples (6 prefrontal cortex, 8 substantia nigra) which were sequenced using snRNA-seq.

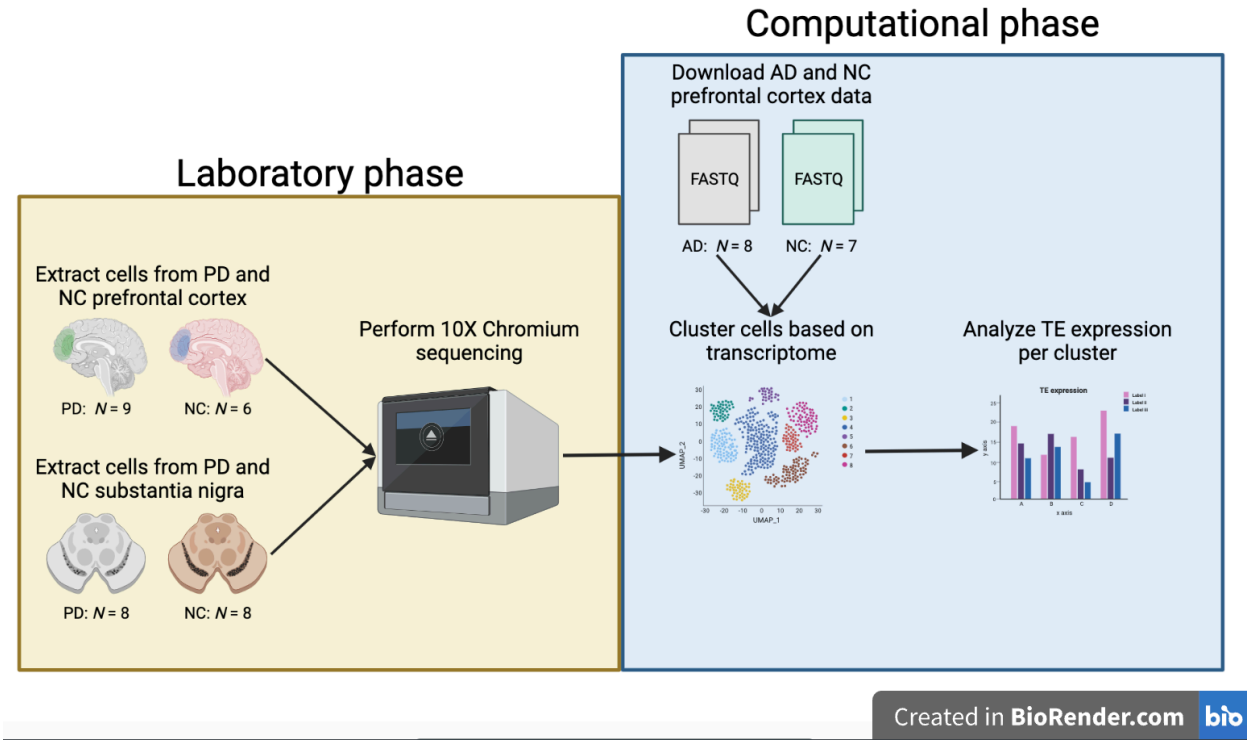

Figure 2: Graphical representation of the PD and AD experiments. For the PD experiments, PD and NC cell samples were sequenced in our laboratory; for the AD experiment, the AD and NC data were downloaded from an online repository. Created in biorender.com.

To study the same question in AD, the analysis was performed using publicly available prefrontal cortex transcriptome data from eight postmortem samples with Alzheimer's disease and seven control postmortem samples (50).

PD experiment – substantia nigra

Quality control and cell type assignment

After filtering out potential doublets (technical artifacts that contain genetic information for two cells despite appearing as one) and low-quality cells, 30,117 cells passed quality control, of which 17,031 belonged to patients with PD and 13,086 of which belonged to the control group (Table 1, Appendix 1).

Cell types were assigned based on the expression of various biomarkers. In the substantia nigra, the detected cell types were neurons (RBFOX3, MAP2, DCX), astrocytes (GFAP, AQP4, GJA1, SLC1A3), microglia (FYB1, P2RY12, CD74), oligodendrocytes (PLP1, MOG, MBP), and oligodendrocyte progenitor cells (VCAN, COL9A1, PDGFRA) (Figure 3b, 3c).

Table 1: Cell type counts in PD and NC substantia nigra samples

	Astrocytes	Microglia	Neurons	Oligodendrocytes	OPCs	Total
PD	1892	1697	127	12079	1236	17031
NC	1428	1004	363	9143	1148	13086
Total	3320	2701	490	21222	2384	30117

256

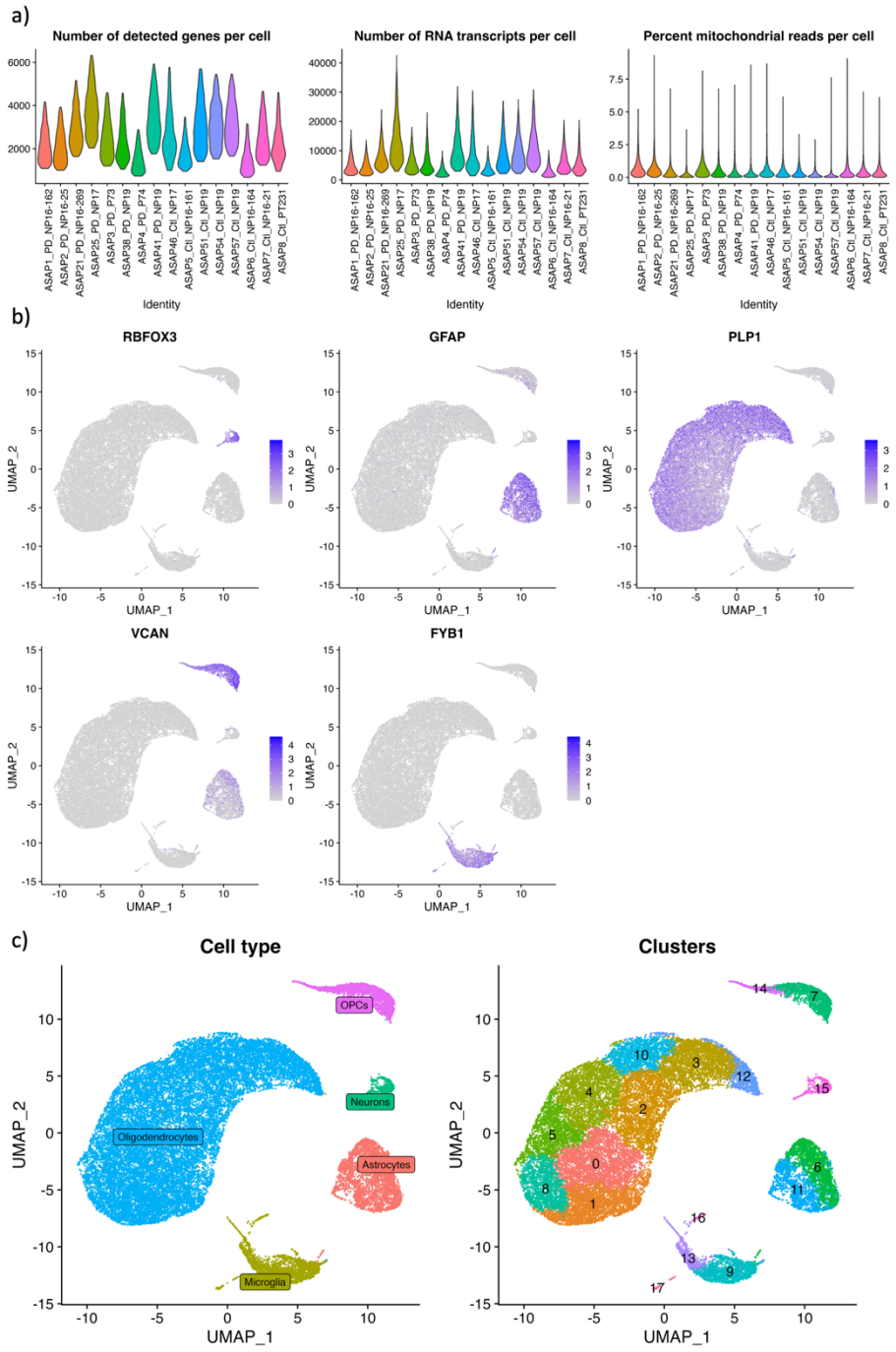

Figure 3: (a) Violin plots showing quality control variables in PD substantia nigra samples (from left: number of genes detected in each cell, total number of RNA reads detected in each cell, and % of reads coming from mitochondria in each cell) (b) Uniform manifold approximation and projection (UMAP) plots showing cell type specific biomarkers (RBFOX3 = neurons, GFAP = astrocytes, PLP1 = oligodendrocytes, VCAN = OPCs, FYB1 = microglia, FLT1 = endothelial cells). (c) UMAPs showing assigned cell type (left) and Seurat cluster (right).

257 *Activated microglia and reactive astrocytes detected in PD*

Some previously annotated biomarkers for activated microglia (FTL, SPP1, APOE, CD74, FDGR3A, PTPRC, CST3, and CSF1R) were found significantly upregulated (Wilcoxon, $p < 0.05$) in PD microglia (Figure 4a), as well as known biomarkers of reactivity in PD astrocytes (CHI3L1, C3, S100B, CRYAB, MAOB, NFAT5, HSPB1, MT2A) (Wilcoxon, $p < 0.05$) (Figure 4b) (45–49).

Following gene set enrichment analysis (GSEA) we corroborated the presence of activated microglia in the substantia nigra of patients with PD, with multiple immune-related gene ontology terms showing upregulation. Such terms included “immune effector process,” “cell activation,” and “cytokine production” in microglia (Figure 4c). Astrocyte terms seemed to be largely focused around extracellular interactions and movement, like “cell motility” and “external encapsulating structure” (Figure 4c).

Interestingly, many gene ontology terms in oligodendrocytes were also upregulated. Most of the upregulated terms were related to the 3D structures of proteins, such as “unfolded protein binding,” “protein folding,” and “response to topologically incorrect protein”.

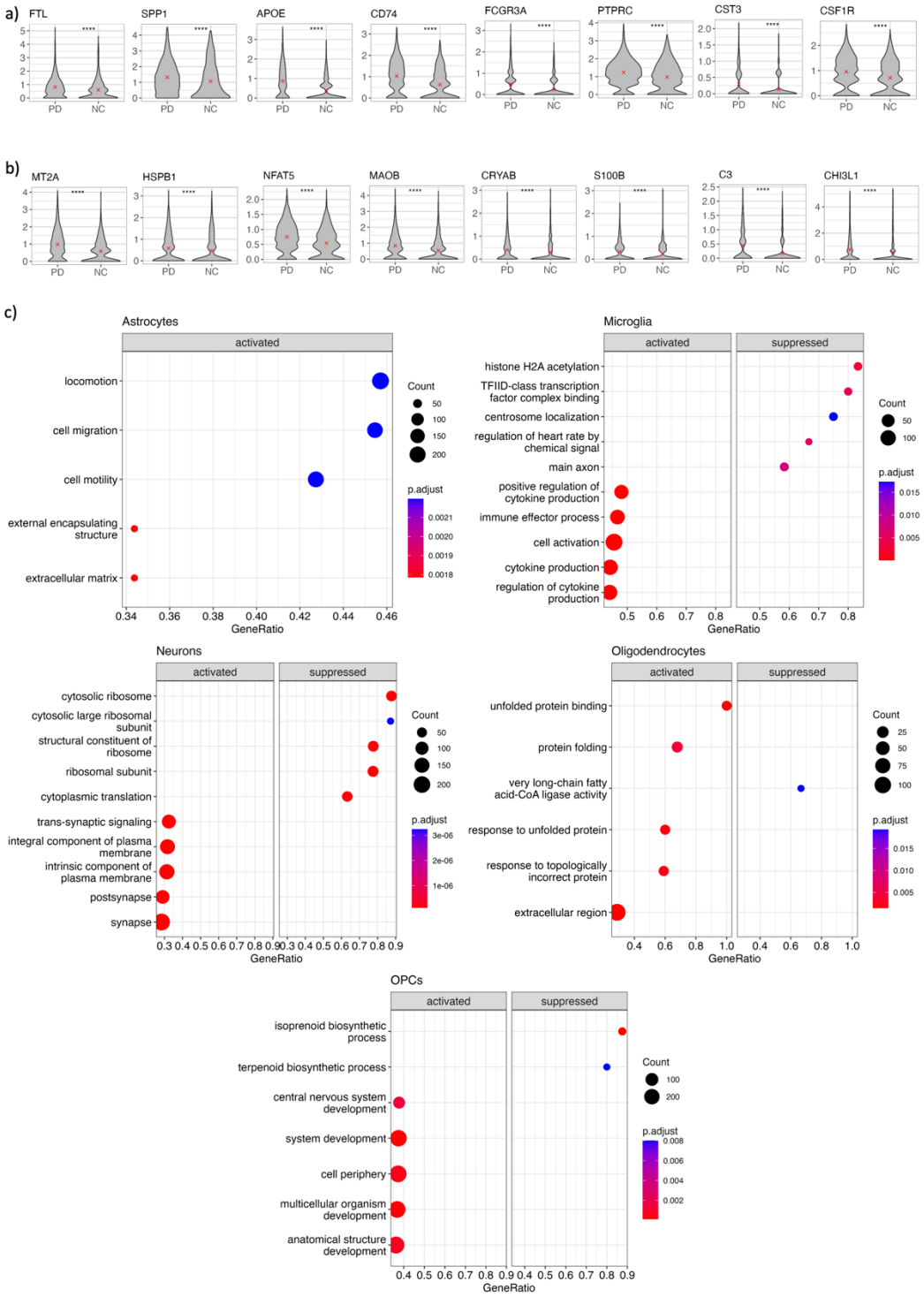

*Figure 4: (a) GSEA results for each cell type, showing the top five most significantly activated and repressed terms in PD (if more than five terms were significant; otherwise, all terms are shown). (b) Violin plots showing expression of microglial activation biomarkers in PD (left) and normal control (right) individuals. (c) Violin plots showing expression of astrocyte activation biomarkers in PD (left) and control (right) individuals. (For (b) and (c): Wilcoxon rank-sum test, ns = ($p > 0.05$); * = ($p < 0.05$); ** = ($p < 0.01$); *** = ($p < 0.001$); **** = ($p < 0.0001$).*

275 OPCs in PD substantia nigra samples showed a large amount of upregulation of GO terms

associated with extracellular interactions. Such terms included “cell periphery” (Figure 4c). Many other activated terms were associated with tissue development, such as “central nervous system development,” “multicellular organism development,” and “system development” (Figure 4c).

Neurons also showed a large degree of both suppression and activation of various terms. The most suppressed terms were broadly related to ribosomes and protein translation, such as “ribosomal subunit” and “cytoplasmic translation” (Figure 4c). The most activated terms were largely related to synaptic activity, such as “postsynapse,” “trans-synaptic signaling,” and simply “synapse” (Figure 4c).

TE expression – decreases in L1 and HERV expression seen in PD oligodendrocytes

Results are shown for four young L1 subfamilies: L1HS, L1PA2, L1PA3, and L1PA4 (Figure 5a, 5b). In each case, TE expression is somewhat lower in PD samples than in control samples

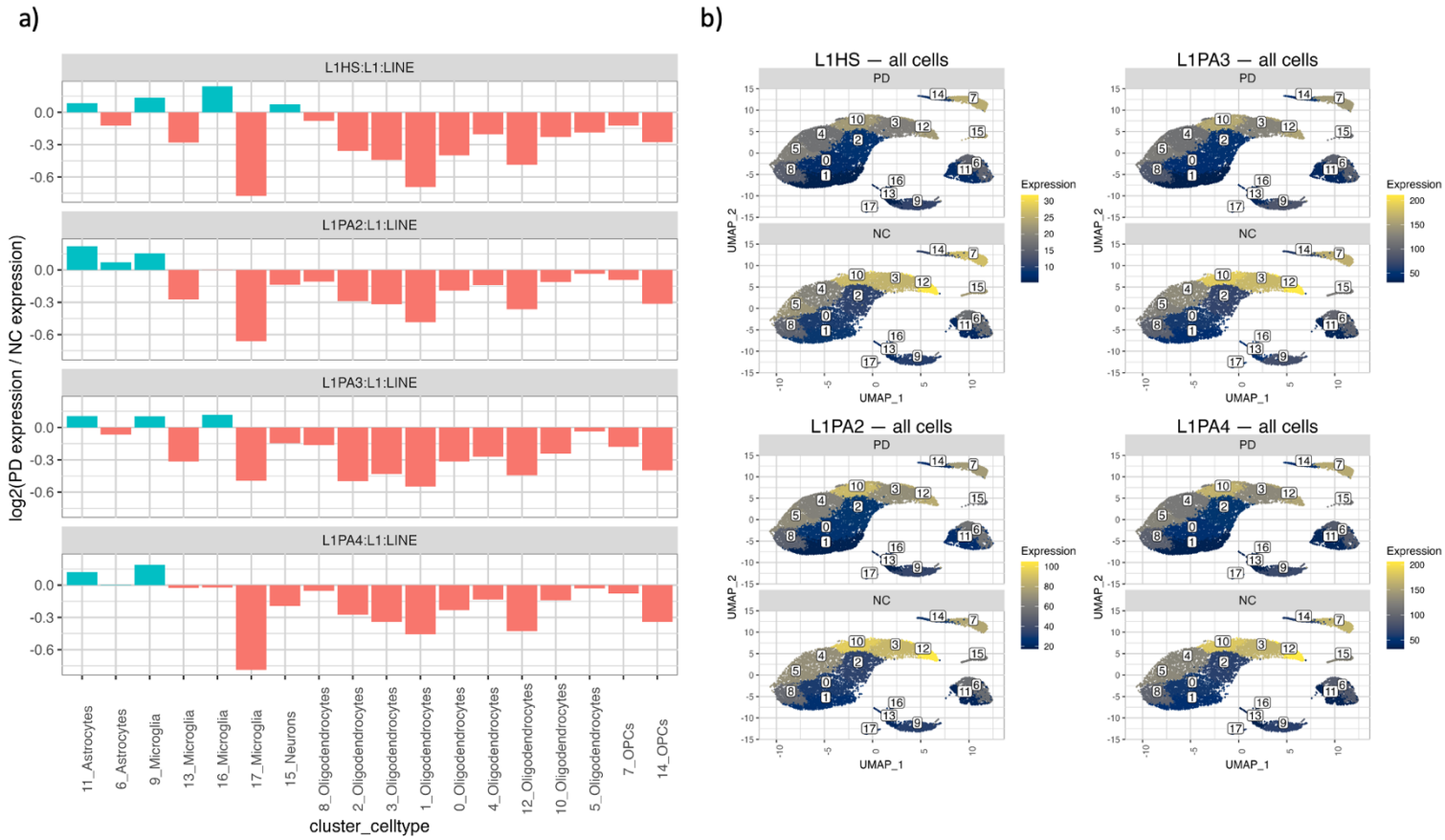

Figure 5: (a) Bar charts showing the LFD of the expression of various TE subfamilies in PD vs. control samples. (b) UMAPs showing the normalized expression (total number of reads / cell count per cluster) of those same TEs.

across many cell types, most notably in the oligodendrocytes; however, notable differences are also present in OPCs (Figure 5a, 5b). That said, no cluster shows a log fold difference (LFD) in expression of below -1 in PD (Figure 5a, 5b). No differences were seen in L1 expression in microglia, astrocytes, or neurons (Figure 5a, 5b).

295 The data in five ERV subfamilies – LTR2, LTR5-Hs, LTR5B, LTR7, and LTR17 – show that

296 LTRs appear to have a similar expression profile in oligodendrocytes to L1s. All five subfamilies

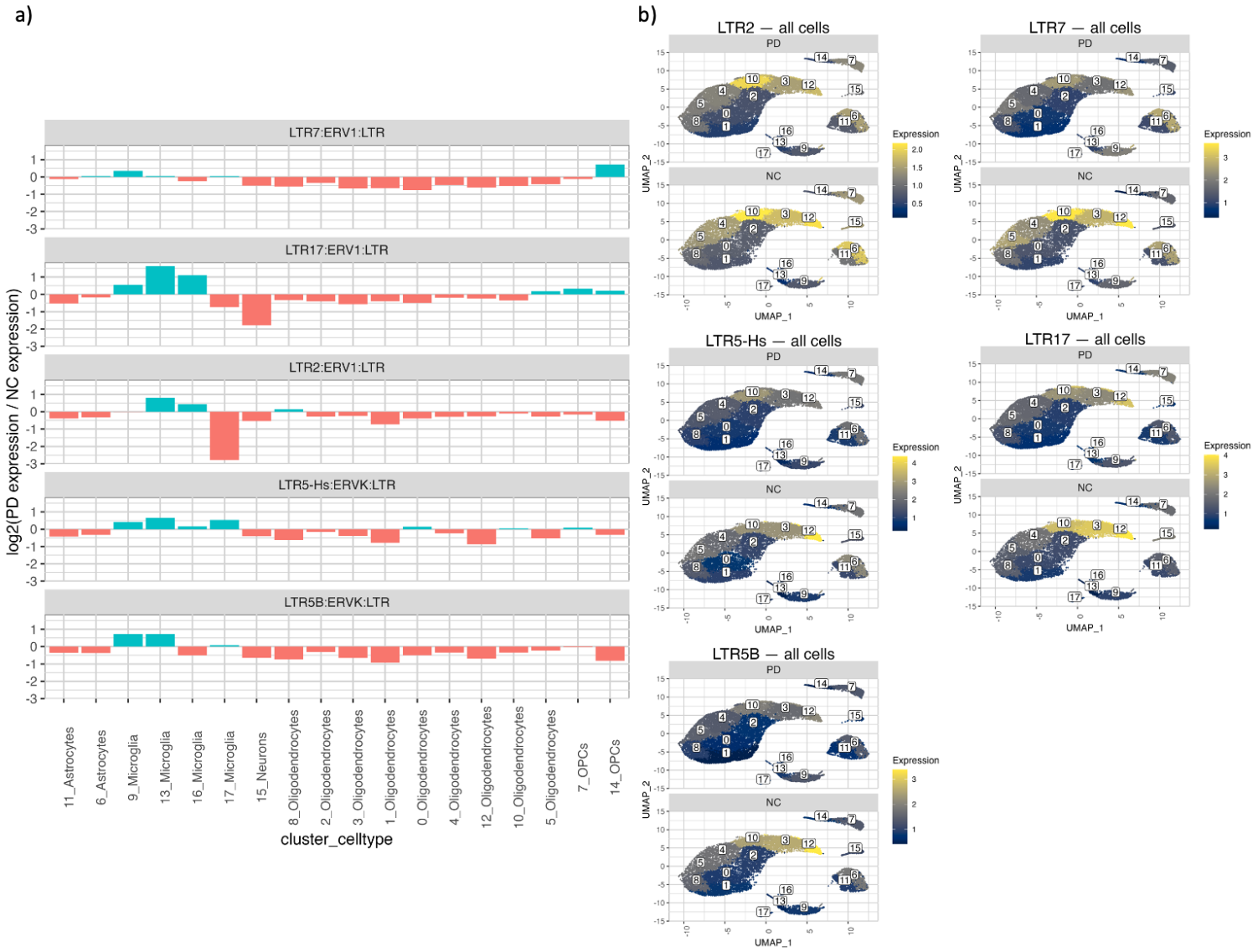

Figure 6: (a) Bar charts showing the LFD of the expression of various TE subfamilies in PD vs. control samples. (b) UMAPs showing the normalized expression (total number of reads / cell count per cluster) of those same TEs.

297 show decreased expression in oligodendrocytes and neurons from PD samples, with LTR2

298 showing decreased expression in astrocytes and OPCs. LTR2, LTR5-Hs, and LTR17 all show

299 increased expression in PD microglia (Figure 6a, 6b).

300

301 *PD experiment – cortex*

302 *Quality control and cell type assignment*

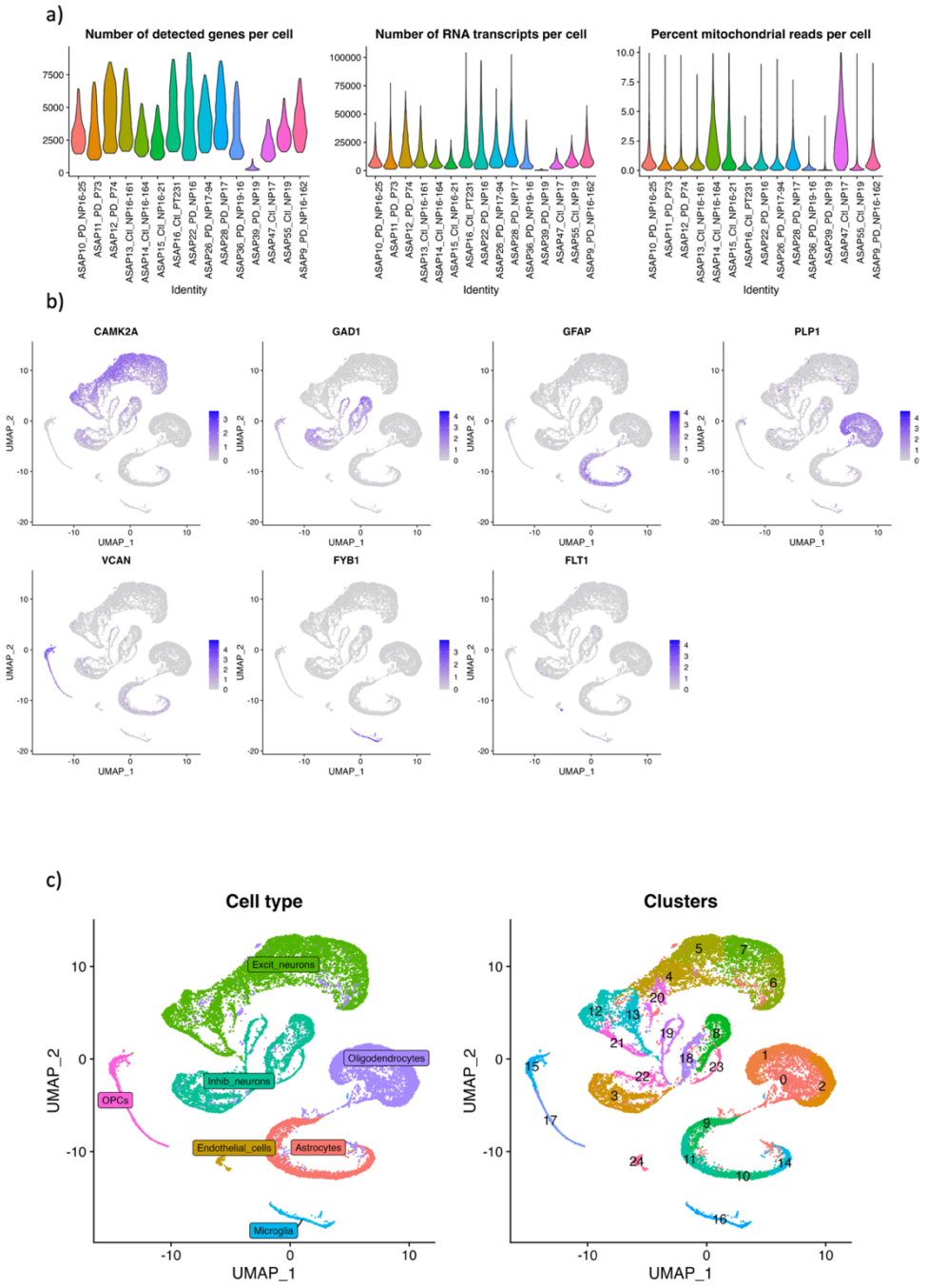

Figure 7: (a) Violin plots showing quality control variables in PD cortex samples (from left: number of genes detected in each cell, total number of RNA reads detected in each cell, and % of reads coming from mitochondria in each cell) (b) UMAPS showing cell type specific biomarkers (CAMK2A = excitatory neurons, GAD1 = inhibitory neurons, GFAP = astrocytes, PLP1 = oligodendrocytes, VCAN = OPCs, FYB1 = microglia, FLT1 = endothelial cells). (c) UMAPs showing assigned cell type (left) and Seurat cluster (right).

Similar exclusion criteria were used in this experiment as in the substantia nigra experiment to ensure that low-quality cells and doublets were excluded from the analysis. In total, 31,348 cells passed quality control, with 13,426 cells belonging to control patients and 17,922 cells belonging to PD samples (Table 2, Appendix 2). We found all of the expected cell types, which we categorized based on the expression of various biomarkers. The detected cell types included astrocytes (GFAP, AQP4, GJA1, SLC1A3), microglia (FYB1, P2RY12, CD74), oligodendrocytes (PLP1, MOG, MBP), excitatory neurons (GRIN1, HS3ST2, CAMK2A), inhibitory neurons (GAD1, GAD2, CALB2, CNR1), OPCs (VCAN, COL9A1, PDGFRA), and a small population of endothelial cells (FLT1, PECAM1) (Figure 7b, 7c).

Table 2: Cell type counts in PD and NC prefrontal cortex samples

	Astrocytes	Endothelial cells	Excitatory neurons	Inhibitory neurons	Microglia	Oligodendrocytes	OPCs	Total
PD	2791	89	5825	3300	587	4428	902	17922
NC	1952	154	4474	2117	301	3557	871	13426
Total	4743	243	10299	5417	888	7985	1773	31348

Microglia and astrocytes activated in PD prefrontal cortex

Some biomarkers of microglial activation – SPP1, CD74, FCGR3A, and CST3 – showed upregulation in PD microglia, likely pointing towards activation of microglia in the prefrontal cortex as well (Figure 8a). This effect was accompanied with the significant upregulation of several biomarkers for astrocyte reactivity in PD astrocytes such as MT2A, HSPB1, MAOB, CRYAB, and CHI3L1 – which suggests the activation of astrocytes in the prefrontal cortex (Figure 8b) (45–49).

GSEA showed a similar astrocyte gene set expression profile as the PD substantia nigra. Many of the gene ontology terms associated with the differentially expressed genes involved interactions with the outside of the cell, such as “external encapsulating structure,” “extracellular matrix,” and “cell adhesion” which were significantly upregulated (Figure 8c). Additionally, many terms related to cell movement, such as “cell migration” and “locomotion,” were activated (Figure 8c).

Likewise, microglia showed similar signs of activation, with the most upregulated terms being “cellular activation” and “immune system process” (Figure 8a). Many of the most downregulated terms were related to ribosomes and protein translation, such as “cytosolic translation” and “ribosome” (Figure 8c).

Endothelial cells showed a very interesting expression pattern in the PD cortex samples. Many GO terms related to vasculature were upregulated, such as “blood vessel development” and “angiogenesis” (Figure 8c). Terms related to the 3D structures of proteins, such as “protein folding,” were suppressed (Figure 8c).

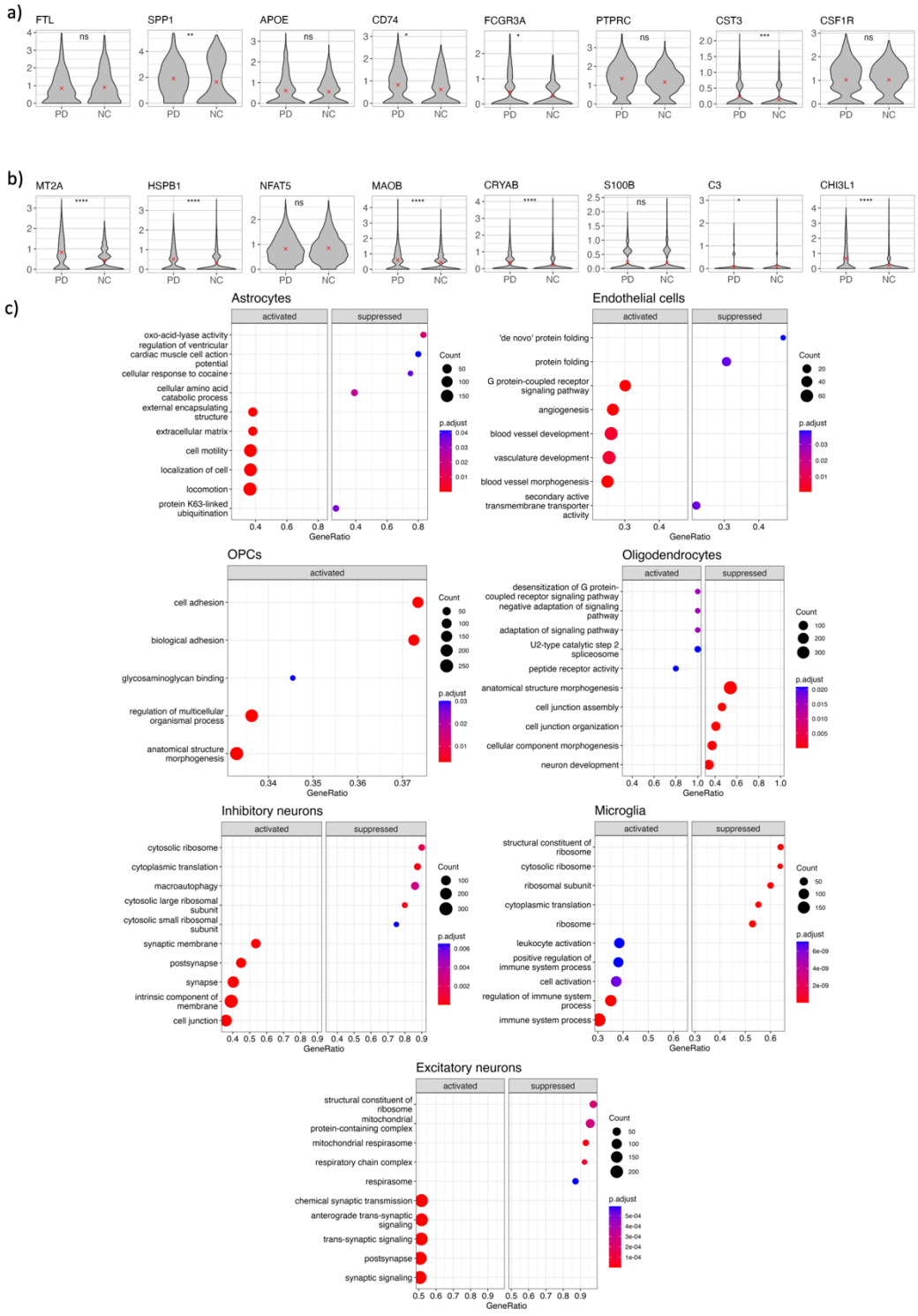

*Figure 8: (a) GSEA results for each cell type, showing the top five most significantly activated and repressed terms in PD (if more than five terms were significant; otherwise, all terms are shown). (b) Violin plots showing expression of microglial activation biomarkers in PD (left) and normal control (right) individuals. (c) Violin plots showing expression of astrocyte activation biomarkers in PD (left) and control (right) individuals. (For (b) and (c): Wilcoxon rank-sum test, ns = ($p > 0.05$); * = ($p < 0.05$); ** = ($p < 0.01$); *** = ($p < 0.001$); **** = ($p < 0.0001$)*

339 Oligodendrocytes in PD prefrontal cortex samples showed a marked decrease in many GO terms

associated with morphogenesis. Such terms included “neuron development,” “cellular component morphogenesis,” and “cell junction assembly” (Figure 8c). There were several upregulated terms as well, but they did not appear to follow a clear pattern (Figure 8c).

In excitatory neurons, many of the most-upregulated terms in PD were related to synaptic activity, with strongly activated terms including “chemical synaptic transmission” and “synaptic signaling” (Figure 8c). Many suppressed terms, like “mitochondrial respirasome,” were related to respiration and other cellular energetic processes.

Inhibitory neurons showed a somewhat similar expression pattern from excitatory neurons. Here, the most upregulated terms tended to be related to cell communication, with terms like “synapse” and “cell junction” showing the most upregulation (Figure 8c). Many of the suppressed terms were related to ribosomes and protein creation, such as “cytosolic ribosome” and “cytoplasmic translation” (Figure 8c).

TE expression – PD astrocytes and microglia show upregulation of TEs in prefrontal cortex

We noticed that astrocytes, microglia, OPCs, and excitatory and inhibitory neurons from PD samples had a marked increase in the expression of the young L1 subfamilies assessed. Of these, astrocytes showed the largest upregulation, with all four clusters showing an LFD of at least 0.75 (Figure 9a, 9b). Oligodendrocytes did not show a strong change in expression, and endothelial cells from PD samples showed a marked *decrease* in expression of all evolutionarily young L1 subfamilies relative to cells from control samples (Figure 9a, 9b).

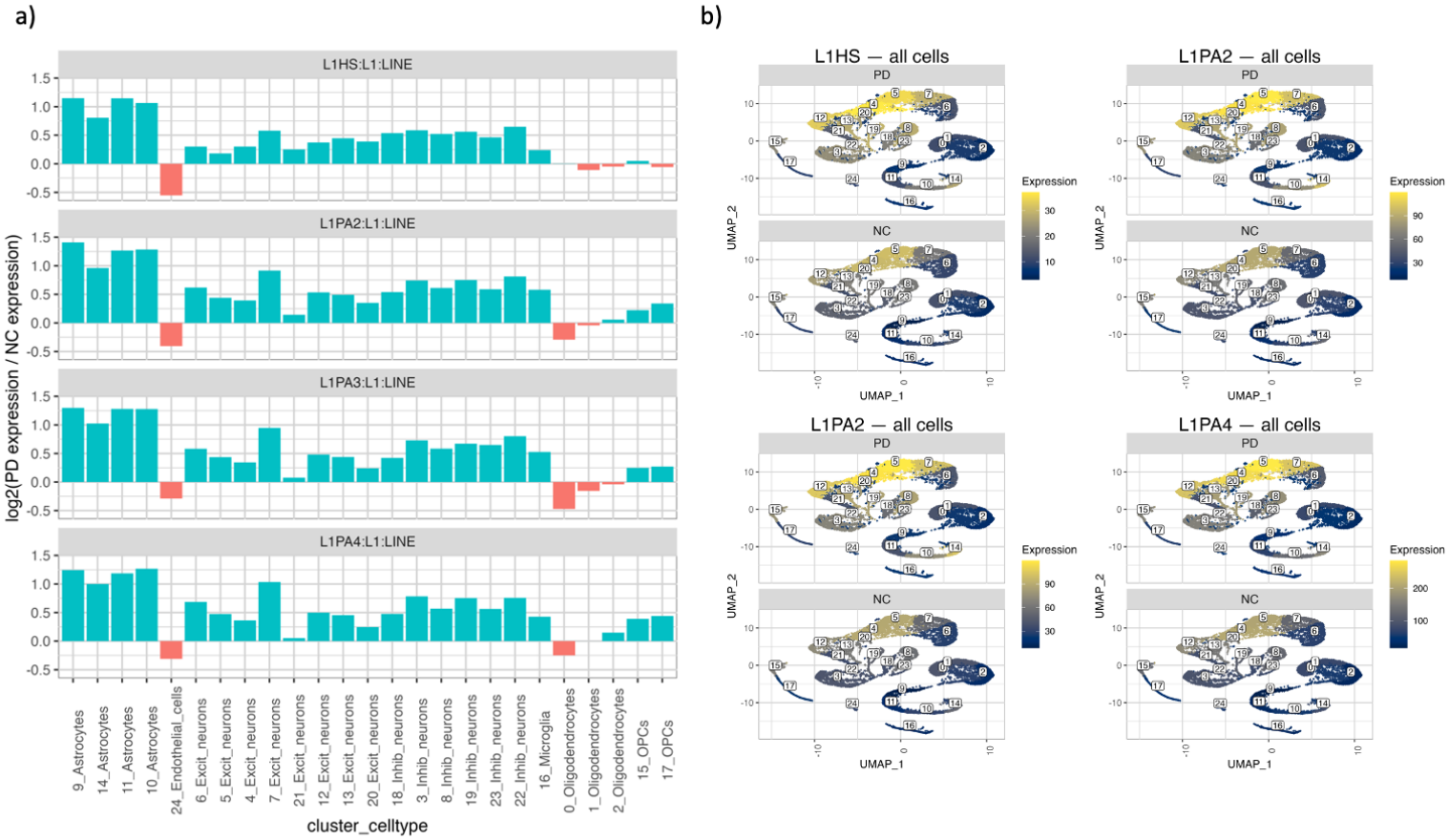

Figure 9: (a) Bar charts showing the LFD of the expression of various TE subfamilies in PD vs. control samples. (b) UMAPs showing the normalized expression (total number of reads / cell count per cluster) of those same TEs.

363

364 LTR2 and LTR5-Hs showed a mild increased expression in excitatory neurons, inhibitory

365 neurons, and astrocytes coming from PD samples (Figure 10a, 10b). The largest differences in

366 LTR5B expression seemed to come from astrocytes, where expression was increased across all

367 clusters in PD samples (Figure 10a, 10b). LTR7 also showed marked increases in expression in

368 astrocytes, as well as in microglia, excitatory neurons, and inhibitory neurons (Figure 10a, 10b).

369 LTR17 showed increased expression in astrocytes and decreased expression in endothelial cells

370 but showed less consistent expression patterns in excitatory and inhibitory neurons (Figure 10a,

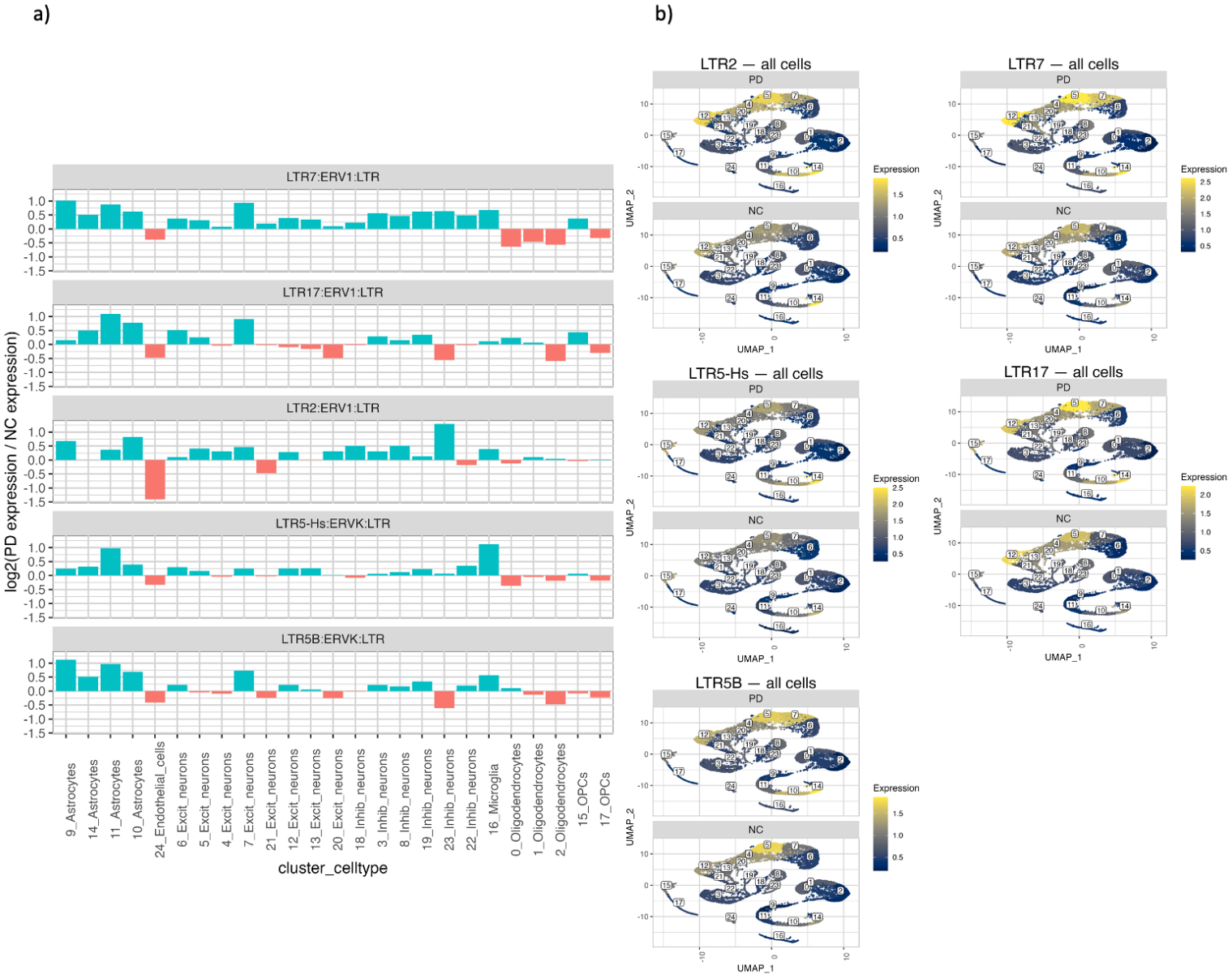

Figure 10: (a) Bar charts showing the LFD of the expression of various TE subfamilies in PD vs. control samples. (b) UMAPs showing the normalized expression (total number of reads / cell count per cluster) of those same TEs.

371 10b). OPC cluster 15 also shows increased expression of LTR17 in PD samples (Figure 10a,

372 10b). Oligodendrocytes did not show any dysregulation of ERV expression.

373

374 *AD experiment*

375 *Quality control and cell type assignment*

376 Overall, 113,535 cells passed quality control. 60,658 of these were from AD samples, while

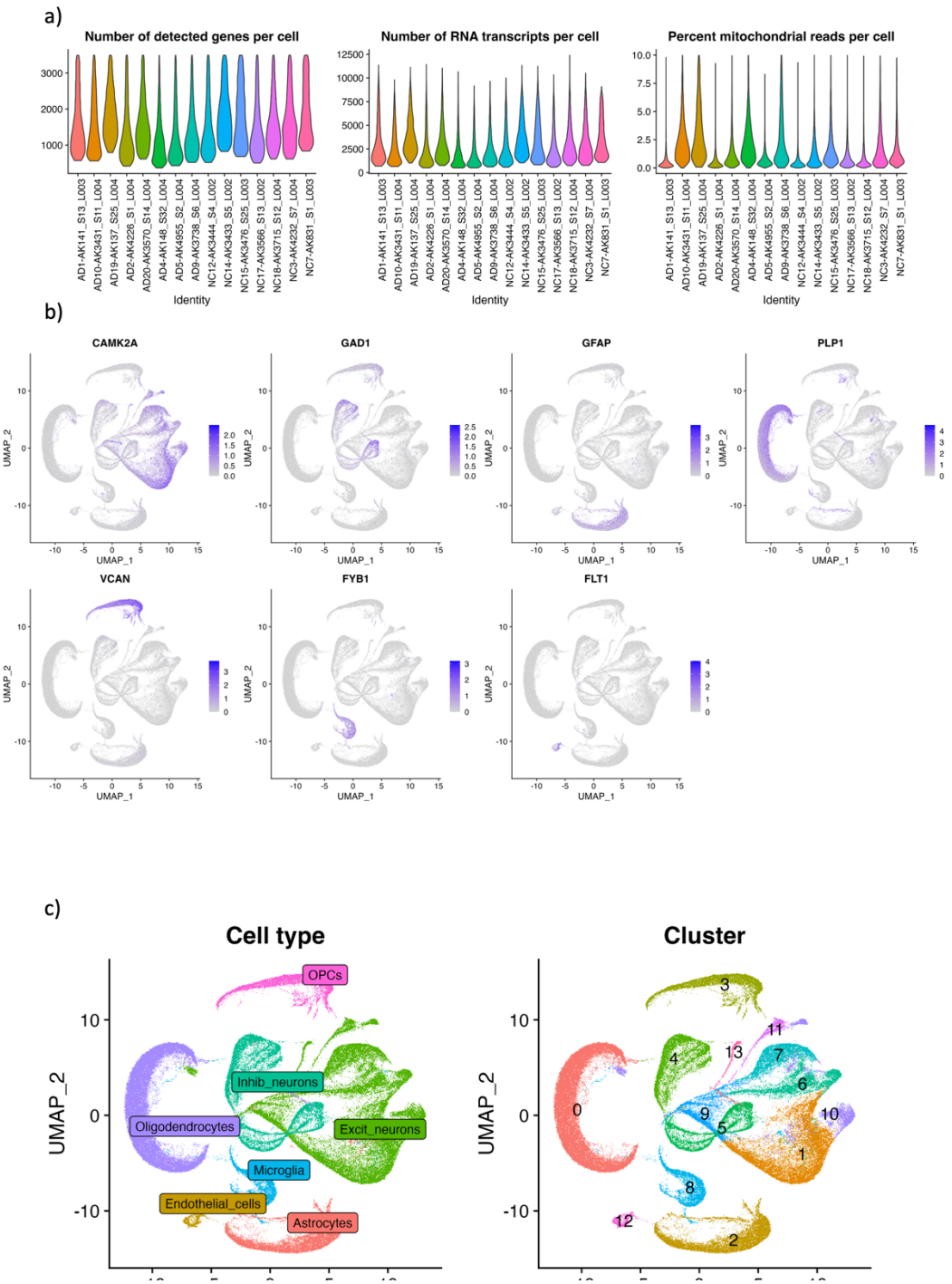

Figure 11: (a) Violin plots showing quality control variables in AD cortex samples (from left: number of genes detected in each cell, total number of RNA reads detected in each cell, and % of reads coming from mitochondria in each cell) (b) UMAPS showing cell type specific biomarkers (CAMK2A = excitatory neurons, GAD1 = inhibitory neurons, GFAP = astrocytes, PLP1 = oligodendrocytes, VCAN = OPCs, FYB1 = microglia, FLT1 = endothelial cells). (c) UMAPs showing assigned cell type (left) and Seurat cluster (right).

52,937 were from control samples (Table 3, Appendix 3). There was no major difference in transcript number between AD samples and control samples. Cell types were characterized with the same biomarkers used for the PD prefrontal cortex dataset (Figure 11b, 11c).

Table 3: Cell type counts in AD and NC prefrontal cortex samples

	Astrocytes	Endothelial cells	Excitatory neurons	Inhibitory neurons	Microglia	Oligodendrocytes	OPCs	Total
AD	6585	775	20897	7990	2704	17285	4422	60658
NC	5667	296	18766	7390	2716	13514	4588	52937
Total	12252	1071	39663	15380	5420	30799	9010	113595

No upregulated genes related to microglia and astrocytes activation in AD

Several biomarkers associated with microglial activation – FTL, SPP1, CD74, and FCGR3A – were downregulated in AD microglia, while just one – APOE – was upregulated (Figure 12a). Additionally, many biomarkers of astrocyte reactivity – MT2A, HSPB1, NFAT5, MAOB, CRYAB, S100B, and CHI3L1 – were shown to be downregulated in AD astrocytes, while one – C3 – was upregulated (Figure 12b) (45–49).

To perform a broader investigation of if there was an inflammatory-like response in microglia and astrocytes in AD as well, GSEA was performed in each cell type. In astrocytes, there were far fewer hallmarks of cellular activity, with the most significant changes in GO terms being suppressions. The most suppressed terms were related to extracellular interactions, such as “cell periphery,” “plasma membrane,” and “extracellular region” (Figure 12c).

Differences in gene set expression were noted in microglia as well. As with the astrocytes, there were fewer signs of increased cellular activity in the AD samples over the control samples. However, the profile of suppressed GO terms was quite similar to the profile seen in PD cortex samples. In both cases, terms related to ribosomes and protein translation were suppressed. Such terms in AD samples include “cytoplasmic translation,” “cytosolic ribosome,” and “ribonucleoprotein complex” (Figure 12c).

Oligodendrocytes in both AD and PD cortexes showed several GO terms suppressed. However, while most of these terms were related to cell development and differentiation in PD prefrontal cortex samples, they were largely related to cellular locomotion in AD prefrontal cortex samples. Particularly suppressed terms include “chemotaxis,” “regulation of cell motility,” and “locomotion” (Figure 12c).

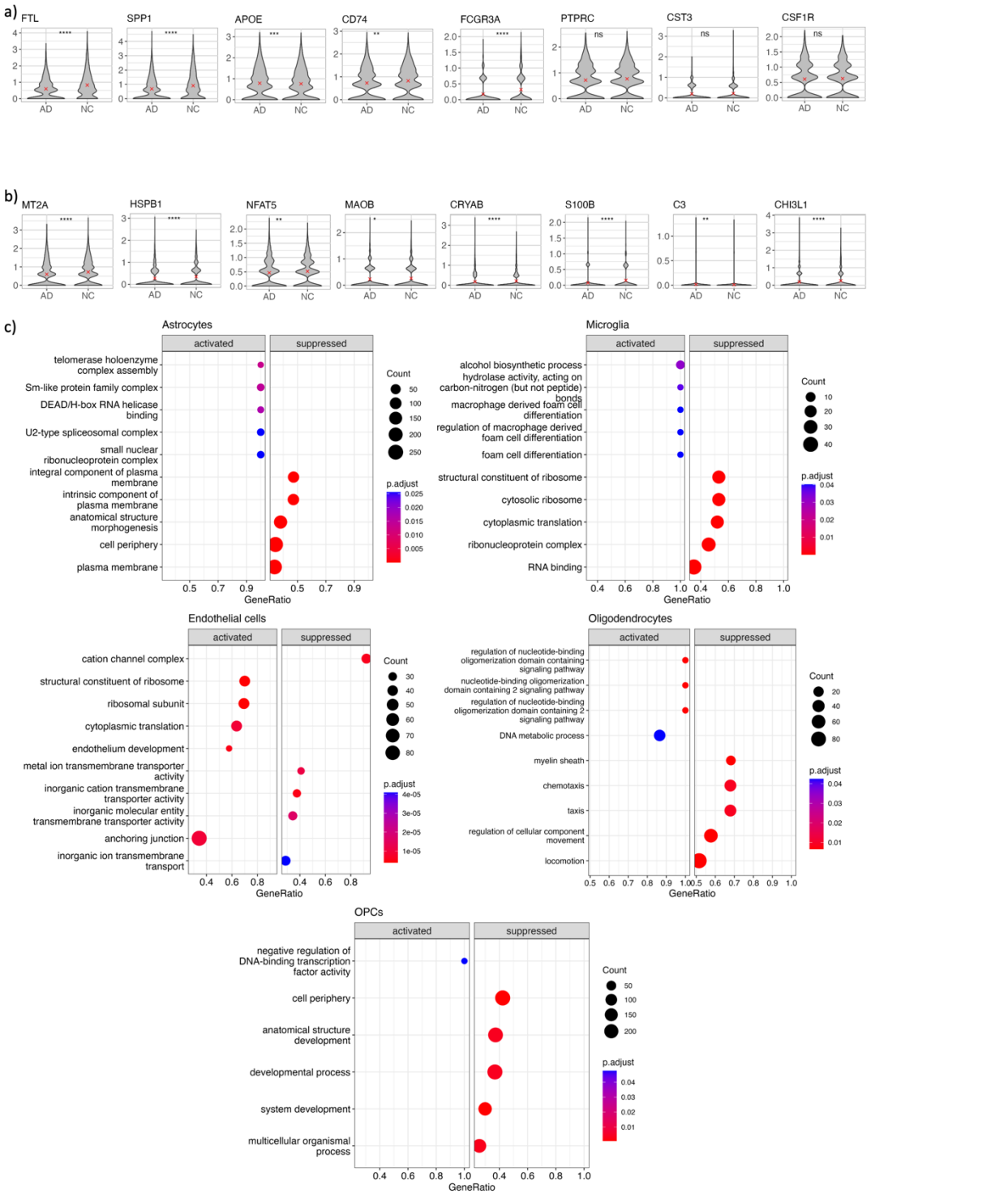

*Figure 12: (a) GSEA results for each cell type, showing the top five most significantly activated and repressed terms in PD (if more than five terms were significant; otherwise, all terms are shown) Excitatory and inhibitory neurons not shown because no terms showed a significant change. (b) Violin plots showing expression of microglial activation biomarkers in PD (left) and normal control (right) individuals. (c) Violin plots showing expression of astrocyte activation biomarkers in PD (left) and control (right) individuals. (For (b) and (c): Wilcoxon rank-sum test, ns = ($p > 0.05$); * = ($p < 0.05$); ** = ($p < 0.01$); *** = ($p < 0.001$); **** = ($p < 0.0001$).*

Interestingly, OPCs in AD samples showed an almost opposite GO profile from the OPCs in PD

411 samples, as GO terms associated with extracellular interaction were largely suppressed (Figure
412 12c). Such terms included “cell periphery” and “multicellular organismal process” (Figure 12c).
413 Just one GO term, “negative regulation of DNA-binding transcription factor activity,” was
414 activated (Figure 12c).
415
416 As with the PD cortex samples, endothelial cells from AD cortexes showed a strikingly large
417 increase in the expression of terms associated with extracellular interaction and tissue
418 development. Such terms in AD samples include “endothelium development” and “anchoring
419 junction” (Figure 12c). No GO terms were activated or suppressed in either excitatory or
420 inhibitory neurons.
421
422 *TE expression – HERVs show increased expression in endothelial cells and microglia*

We found some evidence that L1HS, L1PA2, L1PA3, and L1PA4 are mildly upregulated in AD

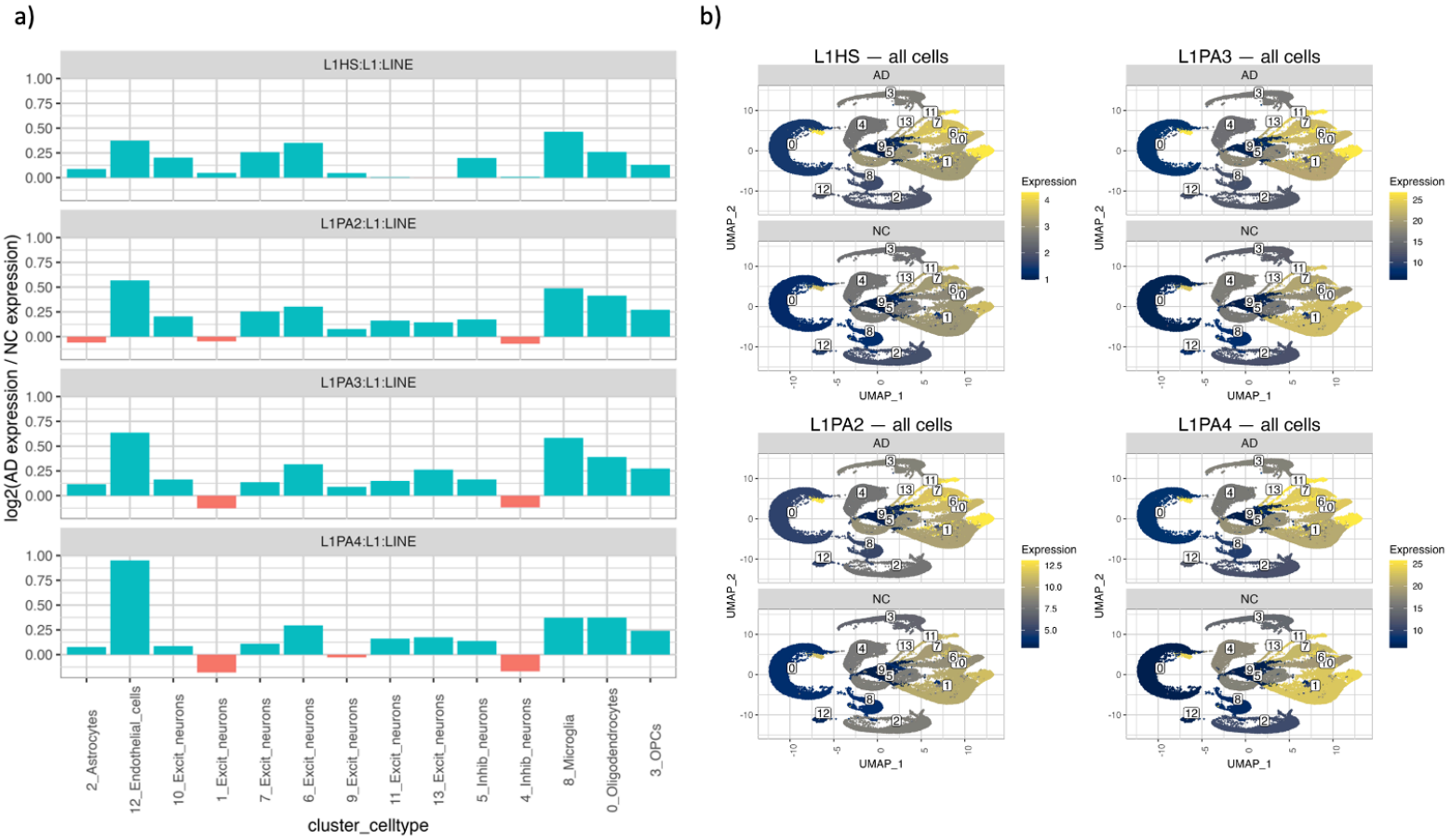

Figure 13: (a) Bar charts showing the expression of various L1 subfamilies on a per-cluster basis; data for prefrontal cortex in normal control (NC) and PD samples shown. (b) UMAPs showing the same information.

OPCs, microglia, and excitatory neurons (Figure 13a, 13b). However, in contrast to PD, astrocytes showed no change in any subfamilies, oligodendrocytes showed an increase in the expression of all subfamilies as opposed to no change seen in PD, and endothelial cells showed an increase in the expression of all subfamilies, with L1PA2, L1PA3, and L1PA4 all showing LFDs above 0.5 in AD samples when compared with NC samples – as opposed to the consistent decrease seen in PD (Figure 13a, 13b).

431 ERV expression was far less consistent between subfamilies in AD samples than L1 expression.

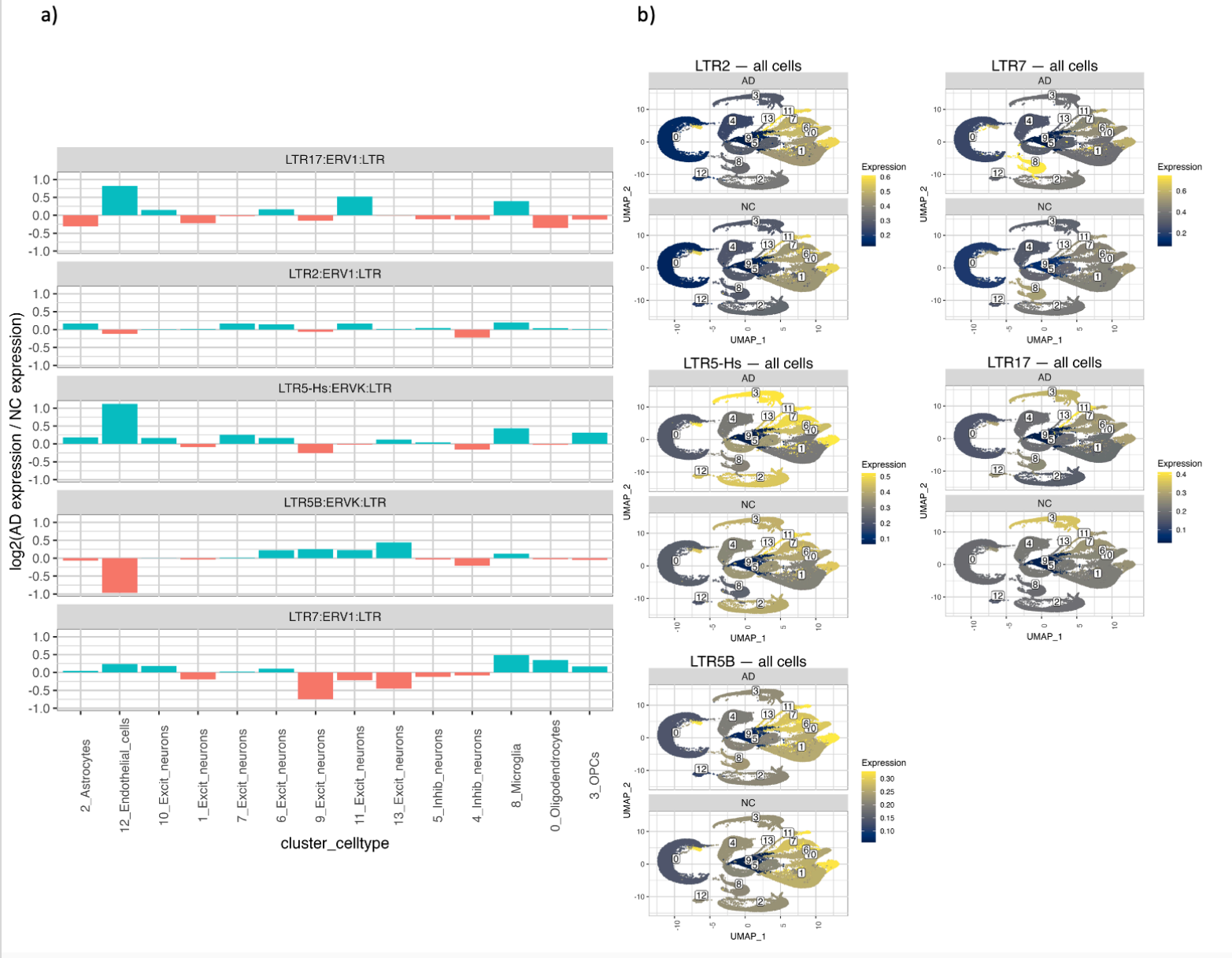

Figure 14: (a) Bar charts showing the expression of various LTR subfamilies on a per-cluster basis in substantia nigra samples from PD and normal control (NC) individuals. (b) UMAPs showing the same information.

432 In LTR2, expression by clusters was generally quite similar between the two conditions (Figure

433 14a, 14b). LTR5-Hs showed a dramatic increase in expression in AD endothelial cells and

434 modest increases in expression in AD astrocytes, OPCs, and microglia; however, there was no

435 major difference in expression in oligodendrocytes or inhibitory neurons, and there were only

436 modest increases in expression in excitatory neuron clusters 6, 7, and 10 (Figure 14a, 14b). The

only difference in LTR5B expression was a decreased expression in AD endothelial cells. LTR7, on the other hand, showed increased expression in both endothelial cells (mildly) and microglia in individuals with AD (Figure 14a, 14b). Finally, LTR17 was found to be upregulated in AD endothelial cells and excitatory neurons from cluster 11, although some other clusters of excitatory neurons showed the opposite results, with only clusters 6 and 10 supporting the upregulation of LTR17. The rest of the cell types shown minor differences such as the upregulation in microglia and the downregulation in astrocytes, OPCs, and excitatory neurons in cluster 1 (Figure 14a, 14b).

Discussion

Understanding the etiologies behind age-related neurodegenerative diseases such as AD and PD has become crucial as global human life expectancies rise and these conditions become more common. Differential gene expression analyses in both substantia nigra and prefrontal cortex samples in our PD dataset point towards microglial activation and astrocyte reactivity. This result corroborates previous observations of neuroinflammation in neurodegenerative diseases (20,43).

Using trusTEr, we were able to assess TE subfamily expression and find differences in the expression of several HERV and L1 subfamilies between diseased brains and control brains in both PD and AD (1,2). L1s, which make up 17% of the human genome, are relevant to study as they are the only TE family which still actively retrotransposes in humans (5,6). This means they have the potential to cause genomic instability, potentially leading to genetic dysregulation and the expression of mutated genes. Nonetheless, we found no support of upregulation of L1s in substantia nigra PD samples. That said, we did find a downregulation of the surveyed L1s in PD

oligodendrocytes; however, considering deleterious health effects that are associated with *increased* L1 expression, this finding is difficult to explain and will require further investigation.

In the PD prefrontal cortex experiment, we found a moderately increased expression of the tested young L1s in microglia, neurons, astrocytes, and OPCs. This finding, in combination with the presence of reactive astrocytes and activated microglia, raise the question if increased L1 expression may play a role in the immune response seen in PD.

The HERV expression study in the PD substantia nigra experiment yielded some interesting results. For instance, we found that LTR17 is upregulated in PD microglia. LTR17s, as well as proteins derived from the *env* gene in HERV-Ws, have been previously shown to cause the degradation of microglia in multiple sclerosis (51). Moreover, several previous studies have correlated HERV expression with neuroinflammation and activated microglia (11,12). However, the cluster where this change was the most dramatic had very few cells in it, meaning this observation needs to be further examined to discard the possibility of it being a technical artifact (given that TE quantification requires many reads due to mapping ambiguity), and to understand its relevance in a PD context. For example, an experiment using immunohistochemical staining for both the *env*-derived protein and microglial activation biomarkers might validate this observation.

In oligodendrocytes of PD substantia nigra, we found a moderate LTR downregulation in PD compared to control, which is challenging to explain and warrants further investigation. In the PD prefrontal cortex samples, LTR2, LTR5-Hs, LTR5B, and LTR7 all showed increased

expression in neurons, while LTR7 and LTR5-Hs showed modest upregulation in astrocytes and microglia.

Interestingly, AD prefrontal cortex samples also showed increased expression of LTR7 and LTR5-Hs in the microglia, a finding that may require further research. In these samples, we did not find evidence of widespread microglial activation or astrocyte reactivity in AD samples; however, a protein detection method such as immunohistochemical staining would be needed to confirm the lack of activated microglia or reactive astrocytes. In addition, we found some evidence of HERV dysregulation in endothelial cells from AD samples, as LTR17 and LTR5-Hs are both modestly upregulated and LTR5B is modestly downregulated.

Though the endothelial cell findings appear important in AD, the small population size of this cell type is worth noting. In contrast to all the other cell types (which had multiple thousands of cells) there were just 296 endothelial cells coming from control samples and 775 endothelial cells coming from AD samples. TE analysis requires many reads due to the challenges in subfamily identification, which makes it difficult to draw any firm conclusions from the patterns seen in endothelial cells.

The results of this study suggest that the overexpression of TEs, and in particular HERVs, might be of importance in the study of neuroinflammation in PD. For example, LTR17 being upregulated in activated PD microglia raises the question if TEs could play a role in the immune response in PD, leading to neuroinflammation. Additionally, the relatively low expression of TEs in PD oligodendrocytes in substantia nigra uncover new potential areas of research.

The implications of the data remain unclear for AD. Although some of the results from this study indicate that TE overexpression in certain cell types (such as microglia) may play a role in PD development, further research is required to draw any conclusions about the role of TEs in PD and AD.

Study Limitations and Further Research

The next step for the research would be to map TE reads uniquely, rather than using multimapping to perform our analysis. Unique mapping would allow us to analyze the data on a per-element level rather than a per-subfamily level. This would allow us to further interrogate the roles that TEs play in altering gene expression on a per-locus level, but it would come at the cost of discarding any reads that map ambiguously to different locations.

TE analysis softwares such as TEtranscripts give us the exciting opportunity to examine how TEs impact human health, and snRNA-seq is a powerful tool for analyzing how different cell subpopulations differ in their transcriptomic profiles. However, high sequencing depth is crucial for TE quantification softwares to produce accurate results, which has traditionally limited us to aggregate RNA-seq methods when studying TEs (4,42). Our approach – to pool related nucleus transcriptomes together from as many samples as possible and to analyze TE expression per cell population – allows us to solve the issues of TE investigation while also allowing us to perform the subpopulation-level analysis afforded to us by snRNA-seq. That said, this approach introduces its own issues, perhaps the most notable one being that each cell cluster only generates one expression value per TE subfamily. This means that, although the bioinformatics approach we chose is the most likely one to generate accurate expression values, we cannot

assess the statistical significance of any TE expression difference. As a result, while our experiments do suggest some cell type-specific relationships between AD/PD and retrotransposon expression, further research is required to confirm or disprove these relationships.

One way these downsides could be mitigated would be to perform a similar analysis on high-depth long-read snRNA-seq data. Such an analysis would allow us to map a greater proportion of TE reads to specific loci on the reference genome, allowing us to examine which loci specifically may be related to neurodegenerative pathologies. Additionally, sufficiently high sequencing depth would make it unnecessary to pool together reads like we did here, meaning that we can determine the statistical significance of our observations.

Methods

Preparation of samples

The PD data were collected from the Cambridge Brain Bank in the UK. The biopsies taken from the substantia nigra and prefrontal cortex were homogenized and analyzed using the 10X Chromium 3' workflow. In this workflow, cells are first multiplexed (labeled based on sample and then pooled together), then separated from each other using 10X's proprietary Gel Beads in Emulsion (GEM) technology. From there, each GEM is sequenced using Illumina short-read sequencing and each transcript is given a unique molecular identifier (UMI).

The preparation protocol of AD samples can be found in the original study (50).

trusTEr

TrusTEr is an open-source Python module developed by the Molecular Neurogenetics Laboratory to create a convenient pipeline for retrotransposon expression analysis. The program takes raw FASTQ files generated by 10X sequencing as input and outputs an RData file containing a Seurat object with retrotransposon expression data appended. The steps of the pipeline are as follows:

1. Quantification of reads

This step runs Cell Ranger's count function, which takes raw FASTQ data as input and outputs data in a variety of formats, including an HTML summary, a BAM file, and a Market Exchange (MEX) format matrix containing barcode information. The version of Cell Ranger used was 3.1.0 for both experiments. The reference genome used in this step is the Genome Reference Consortium Human Build 38 (GRCh38).

2. Generation of cell clusters

This step uses Seurat to cluster cells together using *k*-nearest neighbors clustering on a per-sample basis. It executes the ScaleData() (which scales the expression data in preparation to perform principal component analysis), RunPCA() (which finds principal components of the data), FindNeighbors() (which calculates which cells are most similar to each other based on Euclidean distance), FindClusters() (which creates clusters based on *k* nearest neighbor clustering), and RunUMAP() (which generates a UMAP for easier viewing of distinct cell types) commands sequentially for each sample and saves the generated Seurat objects as output. Additionally, it creates TSV files for each cluster containing a list of all the cellular barcodes associated with each specific cluster. The full program can be found in the trusTEr GitHub

repository. The parameters used that differ from the default are “resolution” (changed to 0.7) and the normalization method (changed to CLR). The Seurat version used is 3.1.5 and the R version used is 4.0.0.

3. Merging and integration of samples

This step merges all the Seurat objects from the previous step into one object using Seurat’s merge() function and reruns the clustering algorithm, creating similar barcode TSVs in the process. As with the previous step, the full program can be found in the trusTEr repository. The custom parameters used here are “integrate” (set to True), “resolution” (set to 0.1 in the AD experiment and 0.5 in the PD experiment), and “normalization_method” (set to “CLR”). The Seurat version used is 3.1.5 and the R version used is 4.0.0.

4. TSV to BAM conversion

This step executes the subset-bam utility to use the barcode TSVs created in the last step to extract cluster-specific data from the BAM files created in the quantification step and generate new cluster-specific BAM files to be used in further analysis. The subset-bam version used here is 1.0.

5. UMI filtering

This step filters reads based on the UMI appended to each molecule during Illumina sequencing to eliminate duplicate molecules recorded due to PCR replication. The custom program filterUMIs loops through each read in the BAM file from the previous step and removes any reads with UMIs that have appeared before. This deduplication allows us to be confident that multiple identified copies of a gene or TE transcript truly do come from separate molecules in the cell.

6. BAM to FASTQ conversion

This step runs the bamtofastq utility from the Cell Ranger suite on the BAM files outputted in the previous step to create FASTQ files compatible for downstream analysis.

7. Lane concatenation

This step creates a simple Bash script to concatenate all the lane-specific R2 FASTQ files from the previous step into a single FASTQ file.

8. Cluster merging

This step reruns the merge_samples.R script to group all genetic information from certain populations into distinct Seurat objects. In this case, AD and control samples were merged into two separate Seurat objects in the AD experiment, and PD and control samples were merged into two separate Seurat objects in the PD experiment.

9. Cluster mapping

This step executes the Spliced Transcript Alignment to a Reference (STAR) software to align the cluster-specific FASTQ files to a reference genome and produce a BAM file as a result (52). The STAR version used is 2.7.8a.

10. TE quantification

This step uses the TEtranscripts software to quantify the TE expression in each cluster (42). TEtranscripts was preferred over traditional high-throughput RNA-seq analysis softwares, such as HTSeq and Cufflinks, because it is programmed to account for the large amount of repetition in TE sequences and thus provides more accurate results than its competitors when analyzing TE data (42).

11. Normalization of TE results

This step takes the raw output from TEtranscripts, divides the TE count per cluster by the number of cells in each cluster, and appends the result to the Seurat object created previously.

This prevents larger clusters from being overrepresented in TE expression when compared to smaller clusters.

Local data analysis

Local data analysis was performed in R v. 4.1.2 using the packages Seurat v. 4.1.0, tidyverse v. 1.3.1, ggpubr v. 0.4.0, and viridis v. 0.6.2 (53–56). Biomarker UMAPs were generated using the FeaturePlot() function in Seurat. GSEA was performed using the clusterProfiler package v. 4.2.2 as well as the org.Hs.eg.db human genome annotation package v. 3.14.0 (57,58). More details can be found in the associated GitHub repository, located at https://github.com/SteinAcker1/MSc_thesis_code.

Quality control

To ensure that low-quality cells and doublets were not included in our analyses in the PD samples, we removed cells where the number of total reads was less than the mean for the sample minus one standard deviation, as well as cells where the number of total reads was greater than the mean plus two standard deviations. Additionally, cells where more than 10% of reads came from the mitochondria were excluded.

The quality control protocol used for the AD samples was somewhat different from the protocol used for the PD experiments. The same criteria were used to eliminate cells with too few genes or too high of a mitochondrial percentage; however, doublets filtered out by eliminating any cell with more than 3500 genes. The change in protocol was done because many of the samples had

fat-tailed distributions, meaning (*mean + 2SD*) was not necessarily a surefire way to eliminate doublets.

Data access

The AD data were downloaded from the Gene Expression Omnibus (GEO) (accession no. GSE157827)(50). We only included the largest replicate from each sample in our analysis. These replicates were AD10-AK3431_S11_L004, AD19-AK137_S25_L004, AD1-AK141_S13_L003, AD20-AK3570_S14_L004, AD2-AK4226_S1_L004, AD4-AK148_S32_L004, AD5-AK4955_S2_L004, AD9-AK3738_S6_L004, NC12-AK3444_S4_L002, NC14-AK3433_S5_L002, NC15-AK3476_S25_L003, NC17-AK3566_S13_L002, NC18-AK3715_S12_L004, NC3-AK4232_S7_L004, and NC7-AK831_S1_L003.

These samples were generated via paired-read RNA-seq using an Illumina NovaSeq 6000. Further details regarding the preparation of these samples can be found in the paper and in the experiment's GEO page.

Acknowledgements

First and foremost, I would like to thank Raquel Garza and Johan Jakobsson for their incredible guidance throughout this thesis project. Their support has been immensely helpful through the many joys and challenges this challenge has brought, and I am extremely grateful for it. I would also like to thank Anita Adami, Diahann Atacho, and Jenny Johansson at Lund University, as well as Roger Barker at Cambridge University, for performing the laboratory phase of the experiment. I would also like to thank Vivien Horvath for reviewing this paper and giving her

666 feedback. Additionally, I would like to thank all the other M.Sc. students in A11 this year for
667 helping to make the lab such a wonderful environment. I would also like to thank Aligning
668 Science Across Parkinson's (ASAP), Lund Stem Cell Center, MultiPark, Hjärnfonden,
669 Cancerfonden, Parkinson fonden, Barncancer fonden, and Vetenskapsrädet for their support.
670 Finally, I would like to thank Dag Ahrén, Eran Elhaik, and Anna Runemark for leading such a
671 challenging, fulfilling, and exciting M.Sc. program.

References

1. Jönsson ME, Garza R, Johansson PA, Jakobsson J. Transposable Elements: A Common Feature of Neurodevelopmental and Neurodegenerative Disorders. Trends Genet. 2020 Aug 1;36(8):610–23.

2. Saleh A, Macia A, Muotri AR. Transposable Elements, Inflammation, and Neurological Disease. Front Neurol. 2019 Aug 20;10:894.

3. Huang CRL, Burns KH, Boeke JD. Active Transposition in Genomes. Annu Rev Genet. 2012 Dec 15;46(1):651–75.

4. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018 Nov 19;19(1):199.

5. Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 Elements in Structural Variation and Disease. Annu Rev Genomics Hum Genet. 2011 Sep 22;12(1):187–215.

6. Suarez NA, Macia A, Muotri AR. LINE-1 retrotransposons in healthy and diseased human brain. Dev Neurobiol. 2018 May 1;78(5):434–55.

7. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, et al. LINE-1 Retrotransposition Activity in Human Genomes. Cell. 2010 Jun 25;141(7):1159–70.

8. Brouha Brook, Schustak Joshua, Badge Richard M., Lutz-Prigge Sheila, Farley Alexander H., Moran John V., et al. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci. 2003 Apr 29;100(9):5280–5.

9. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MCN, et al. Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. Cell. 2015 Oct;163(3):583–93.

10. Belshaw Robert, Pereira Vini, Katzourakis Aris, Talbot Gillian, Pačes Jan, Burt Austin, et al. Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci. 2004 Apr 6;101(14):4894–9.

11. Küry P, Nath A, Créange A, Dolei A, Marche P, Gold J, et al. Human Endogenous Retroviruses in Neurological Diseases. Trends Mol Med. 2018 Apr 1;24(4):379–94.

12. Jönsson ME, Garza R, Sharma Y, Petri R, Södersten E, Johansson JG, et al. Activation of endogenous retroviruses during brain development causes an inflammatory response. EMBO J. 2021 May 3;40(9):e106423.

13. Licastro F, Porcellini E. Activation of Endogenous Retrovirus, Brain Infections and Environmental Insults in Neurodegeneration and Alzheimer's Disease. Int J Mol Sci. 2021 Jul 6;22(14):7263.

14. Li W, Lee MH, Henderson L, Tyagi R, Bachani M, Steiner J, et al. Human endogenous retrovirus-K contributes to motor neuron disease. Sci Transl Med. 2015 Sep 30;7(307):307ra153-307ra153.

15. Sosa-Ortiz AL, Acosta-Castillo I, Prince MJ. Epidemiology of Dementias and Alzheimer's Disease. Arch Med Res. 2012 Nov;43(8):600–8.

16. Savica R, Grossardt BR, Rocca WA, Bower JH. Parkinson disease with and without Dementia: A prevalence study and future projections. Mov Disord. 2018 Apr 1;33(4):537–43.

17. Cunningham EL, McGuinness B, Herron B, Passmore AP. Dementia. Ulster Med J. 2015 May;84(2):79–87.

18. Twohig D, Nielsen HM. α-synuclein in the pathophysiology of Alzheimer's disease. Mol Neurodegener. 2019 Dec;14(1):23.

19. Custodia A, Ouro A, Romaus-Sanjurjo D, Pías-Peleteiro JM, de Vries HE, Castillo J, et al. Endothelial Progenitor Cells and Vascular Alterations in Alzheimer's Disease. Front Aging Neurosci. 2022 Jan 26;13:811210–811210.

20. Grammas P. Neurovascular dysfunction, inflammation and endothelial activation: implications for the pathogenesis of Alzheimer's disease. J Neuroinflammation. 2011 Mar 25;8:26–26.

21. Zlokovic BV. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. Nat Rev Neurosci. 2011 Dec 1;12(12):723–38.

22. Zaman V, Shields DC, Shams R, Drasites KP, Matzelle D, Haque A, et al. Cellular and molecular pathophysiology in the progression of Parkinson's disease. Metab Brain Dis. 2021 Jun;36(5):815–27.

23. Sanford AM. Lewy Body Dementia. Alzheimer Dis Dement. 2018 Nov 1;34(4):603–15.

24. Lai TT, Kim YJ, Ma H il, Kim YE. Evidence of Inflammation in Parkinson's Disease and Its Contribution to Synucleinopathy. JMD. 2021 Nov 3;15(1):1–14.

25. Diaz K, Kohut ML, Russell DW, Stegemöller EL. Peripheral inflammatory cytokines and motor symptoms in persons with Parkinson's disease. Brain Behav Immun - Health. 2022 Mar 7;100442.

26. Ozben T, Ozben S. Neuro-inflammation and anti-inflammatory treatment options for Alzheimer's disease. Clin Biochem. 2019 Oct;72:87–9.

27. Zhao J, Zhao D, Wang J, Luo X, Guo R. Inflammation—Cause or consequence of late onset Alzheimer's disease or both? A review of the evidence. Eur J Inflamm. 2022 Jan;20:1721727X2210953.

28. De Cecco M, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA. Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. Aging. 2013 Dec 7;5(12):867–83.

29. Wood JG, Jones BC, Jiang N, Chang C, Hosier S, Wickremesinghe P, et al. Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in *Drosophila*. Proc Natl Acad Sci. 2016 Oct 4;113(40):11277–82.

30. Ramirez P, Zuniga G, Sun W, Beckmann A, Ochoa E, DeVos SL, et al. Pathogenic tau accelerates aging-associated activation of transposable elements in the mouse central nervous system. Prog Neurobiol. 2022 Jan;208:102181.

31. Guo C, Jeong HH, Hsieh YC, Klein HU, Bennett DA, De Jager PL, et al. Tau Activates Transposable Elements in Alzheimer's Disease. Cell Rep. 2018 Jun 5;23(10):2874–80.

32. Sturm Á, Ivics Z, Vellai T. The mechanism of ageing: primary role of transposable elements in genome disintegration. Cell Mol Life Sci. 2015 May;72(10):1839–47.

33. Jiang S, Guo Y. Epigenetic Clock: DNA Methylation in Aging. Stem Cells Int. 2020 Jul 8;2020:1–9.

34. Mansuroglu Z, Benhelli-Mokrani H, Marcato V, Sultan A, Violet M, Chauderlier A, et al. Loss of Tau protein affects the structure, transcription and repair of neuronal pericentromeric heterochromatin. Sci Rep. 2016 Dec;6(1):33047.

35. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, et al. Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 2009/05/13 ed. 2009 Sep;19(9):1516–26.

36. Petri R, Brattås PL, Sharma Y, Jönsson ME, Pircs K, Bengzon J, et al. LINE-2 transposable elements are a source of functional human microRNAs and target sites. PLOS Genet. 2019 Mar 13;15(3):e1008036.

37. Brattås PL, Jönsson ME, Fasching L, Nelander Wahlestedt J, Shahsavani M, Falk R, et al. TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells. Cell Rep. 2017 Jan;18(1):1–11.

38. Grassi DA, Jönsson ME, Brattås PL, Jakobsson J. TRIM28 and the control of transposable elements in the brain. Brain Res. 2019 Feb;1705:43–7.

39. Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jönsson ME, Männe C, et al. TRIM28 Represses Transcription of Endogenous Retroviruses in Neural Progenitor Cells. Cell Rep. 2015 Jan;10(1):20–8.

40. Schön Ulrike, Diem Olivia, Leitner Laura, Günzburg Walter H., Mager Dixie L., Salmons Brian, et al. Human Endogenous Retroviral Long Terminal Repeat Sequences as Cell Type-Specific Promoters in Retroviral Vectors. J Virol. 2009 Dec 1;83(23):12643–50.

41. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2012 Jan;13(1):36–46.

42. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics. 2015 Nov 15;31(22):3593–9.

43. Römer C. Viruses and Endogenous Retroviruses as Roots for Neuroinflammation and Neurodegenerative Diseases. Front Neurosci. 2021;15:648629.

44. Peze-Heidsieck E, Bonnifet T, Znaidi R, Ravel-Godreuil C, Massiani-Beaudoin O, Joshi RL, et al. Retrotransposons as a Source of DNA Damage in Neurodegeneration. Front Aging Neurosci. 2022 Jan 4;13:786897.

45. Escartin C, Galea E, Lakatos A, O'Callaghan JP, Petzold GC, Serrano-Pozo A, et al. Reactive astrocyte nomenclature, definitions, and future directions. Nat Neurosci. 2021 Mar;24(3):312–25.

46. Ochocka N, Kaminska B. Microglia Diversity in Healthy and Diseased Brain: Insights from Single-Cell Omics. Int J Mol Sci. 2021 Mar 16;22(6):3027.

47. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell. 2017 Jun;169(7):1276-1290.e17.

48. Absinta M, Maric D, Gharagozloo M, Garton T, Smith MD, Jin J, et al. A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis. Nature. 2021 Sep 30;597(7878):709–14.

49. Rangaraju S, Raza SA, Li NX, Betarbet R, Dammer EB, Duong D, et al. Differential Phagocytic Properties of CD45low Microglia and CD45high Brain Mononuclear Phagocytes-Activation and Age-Related Effects. Front Immunol. 2018;9:405.

50. Lau SF, Cao H, Fu AKY, Ip NY. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. Proc Natl Acad Sci. 2020 Oct 13;117(41):25800.

51. Kremer D, Gruchot J, Weyers V, Oldemeier L, Göttle P, Healy L, et al. pHERV-W envelope protein fuels microglial cell-dependent damage of myelinated axons in multiple sclerosis. Proc Natl Acad Sci U S A. 2019 Jul 23;116(30):15216–25.

52. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15–21.

53. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun 24;184(13):3573-3587.e29.

54. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.

55. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots [Internet]. 2020 [cited 2022 May 19]. Available from: https://CRAN.R-project.org/package=ggpubr

56. Garnier S, Ross N, Rudis B, Filipovic-Pierucci A, Galili T, Timelyportfolio, et al. viridis - Colorblind-Friendly Color Maps for R [Internet]. Zenodo; 2021 [cited 2022 May 19]. Available from: https://sjmgarnier.github.io/viridis/

57. Carlson M. org.Hs.eg.db: Genome wide annotation for Human [Internet]. 2021 [cited 2022 May 19]. Available from: https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html

58. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021 Aug;2(3):100141.

Supplemental Information

Appendix 1: Cell counts per cluster in the PD substantia nigra experiment.

Cell type	Cluster	PD	NC	Total
Astrocytes	11	842	635	1477
	6	1050	793	1843
Microglia	9	1132	641	1773
	13	454	229	683
	16	46	83	129
	17	65	51	116
Neurons	15	127	363	490
Oligodendrocytes	8	1017	788	1805
	2	1652	1291	2943
	3	1442	1405	2847
	1	1843	1194	3037
	0	1967	1438	3405
	4	1675	1070	2745
	12	487	428	915
	10	891	744	1635
	5	1105	785	1890
OPCs	7	922	890	1812
	14	314	258	572
Total	-	17031	13086	30117

Appendix 2: Cell counts per cluster in the PD prefrontal cortex experiment.

Cell type	Cluster	PD	NC	Total
Astrocytes	9	739	529	1268
	14	569	390	959
	11	723	534	1257
	10	760	499	1259
Endothelial cells	24	89	154	243
Excitatory neurons	6	708	1029	1737
	5	1034	794	1828
	4	970	867	1837
	7	891	689	1580
	21	292	197	489
	12	833	408	1241
	13	726	352	1078
	20	371	138	509
Inhibitory neurons	18	548	229	777

	3	1042	876	1918
	8	890	462	1352
	19	450	301	751
	23	188	96	284
	22	182	153	335
Microglia	16	587	301	888
Oligodendrocytes	0	1826	1074	2900
	1	1489	1381	2870
	2	1113	1102	2215
OPCs	15	523	384	907
	17	379	487	866
Total	_	17922	13426	31348

Appendix 3: Cell counts per cluster in the AD prefrontal cortex experiment.

Cell type	Cluster	AD	NC	Total
Astrocytes	2	6585	5667	12252
Endothelial cells	12	775	296	1071
Excitatory neurons	10	2693	1977	4670
	1	8307	7791	16098
	7	3089	2689	5778
	6	3260	2700	5960
	9	2295	2431	4726
	11	733	739	1472
	13	520	439	959
Inhibitory neurons	5	3694	3624	7318
	4	4296	3766	8062
Microglia	8	2704	2716	5420
Oligodendrocytes	0	17285	13514	30799
OPCs	3	4422	4588	9010
Total	_	60658	52937	113595