

Supplementary Figures

Supplementary Figure 1 - Quality control and preprocessing

Supplementary Figure 2 - Ambient RNA removal and UMAP embedding

Supplementary Figure 3 - Metadata distribution

Supplementary Figure 4 - Visualization of covariates on UMAPs

Supplementary Figure 5 - Covariate impact

Supplementary Figure 6 - Compositional analysis by scCODA

Supplementary Figure 7 - Data-driven overlap of neuron subtypes on a striatum dataset

Supplementary Figure 8 - Dataset-driven separation of neuron subtypes on a substantia nigra dataset

Supplementary Figure 9 - Cross validation of GLUergic_RELN subtype on a second dataset

Supplementary Figure 10 - Cell-cell interaction analysis

Supplementary Figure 11 - Differential expression of genes of interest

Supplementary Figure 12 - Microglia trajectories analysis

Supplementary Figure 13 - Top pseudotime-related genes activated microglia

Supplementary Figure 14 - Top pseudotime-related genes PVMs

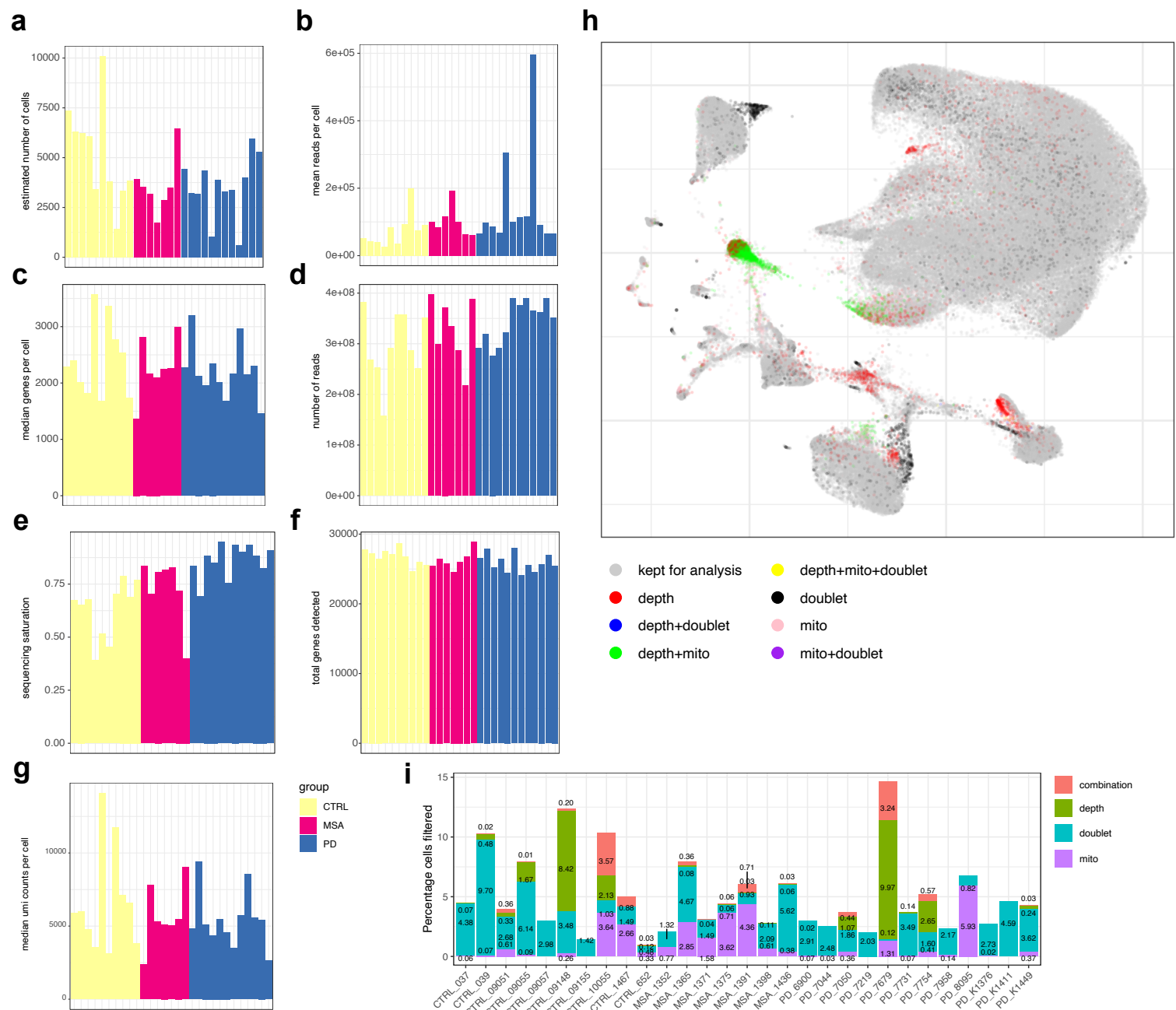
Supplementary Figure 15 - Cross validation of microglia compositional changes on a second dataset

Supplementary Figure 16 - HLA and cytokine expression

Supplementary Figure 17 - Gene regulons

Supplementary Figure 18 - Single-cell disease-relevance GWAS scoring

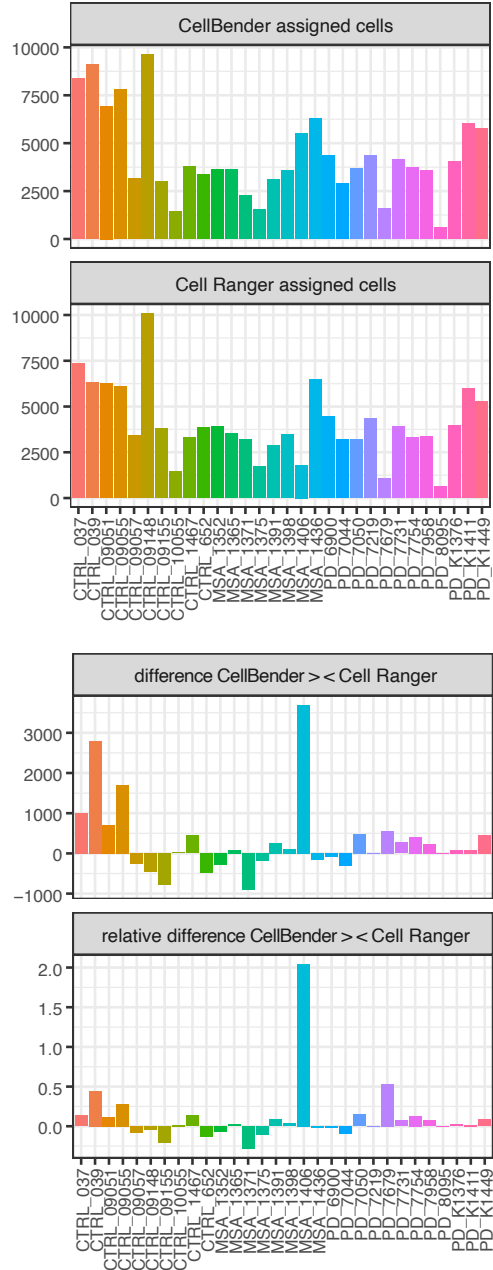
Supplementary Figure 19 - Phagocytosis assay on human derived microglia iPSC



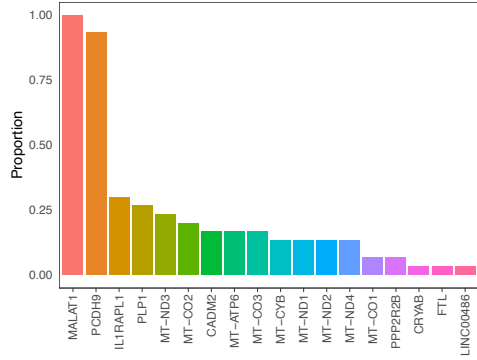
Supplementary Figure 1 - Quality control and preprocessing

a-g Sequencing overview for included samples shown as a number of nuclei per sample, **b** mean reads per nucleus, **c** median genes per cell, **d** number of reads per sample, **e** sequencing saturation per sample, **f** genes detected per sample or **g** median UMI counts per nucleus. **h** UMAP embedding highlighting filtered nuclei. Nuclei were omitted based on depth (< 500 UMIs per nucleus), fraction of mitochondrial (mito) genes per nucleus ($< 5\%$), or doublet labelling. **i** Filtered nuclei per sample by filtering criteria. CTRL: healthy controls, MSA: multiple system atrophy, PD: Parkinson's disease.

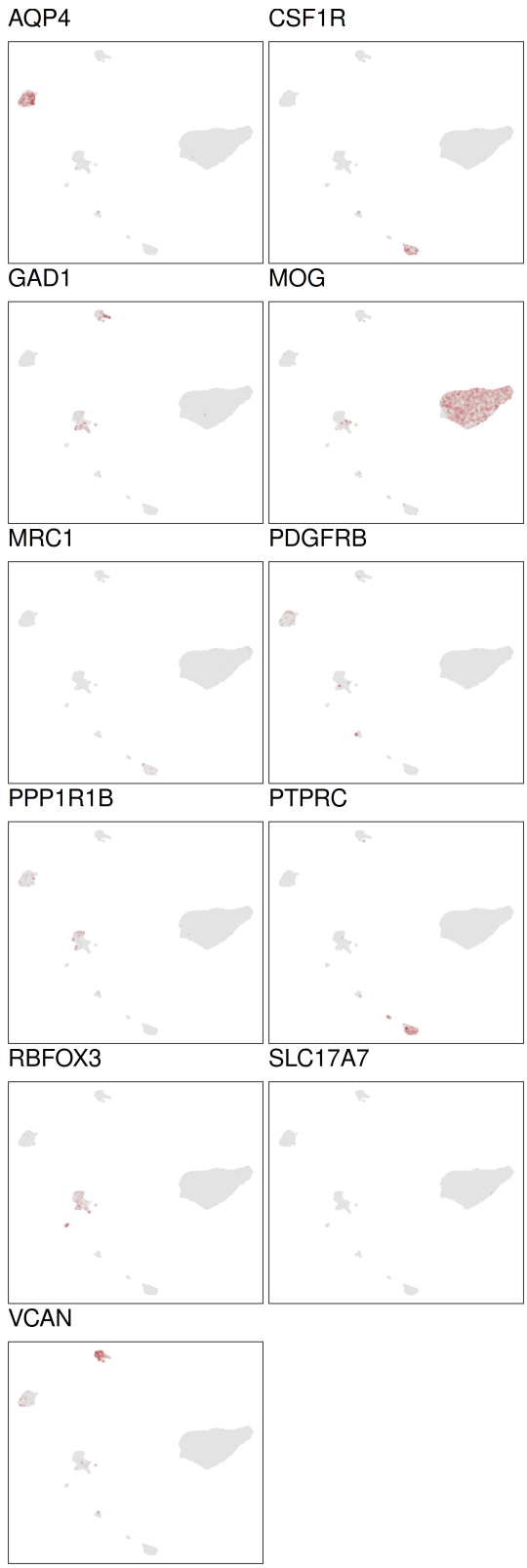
a



b

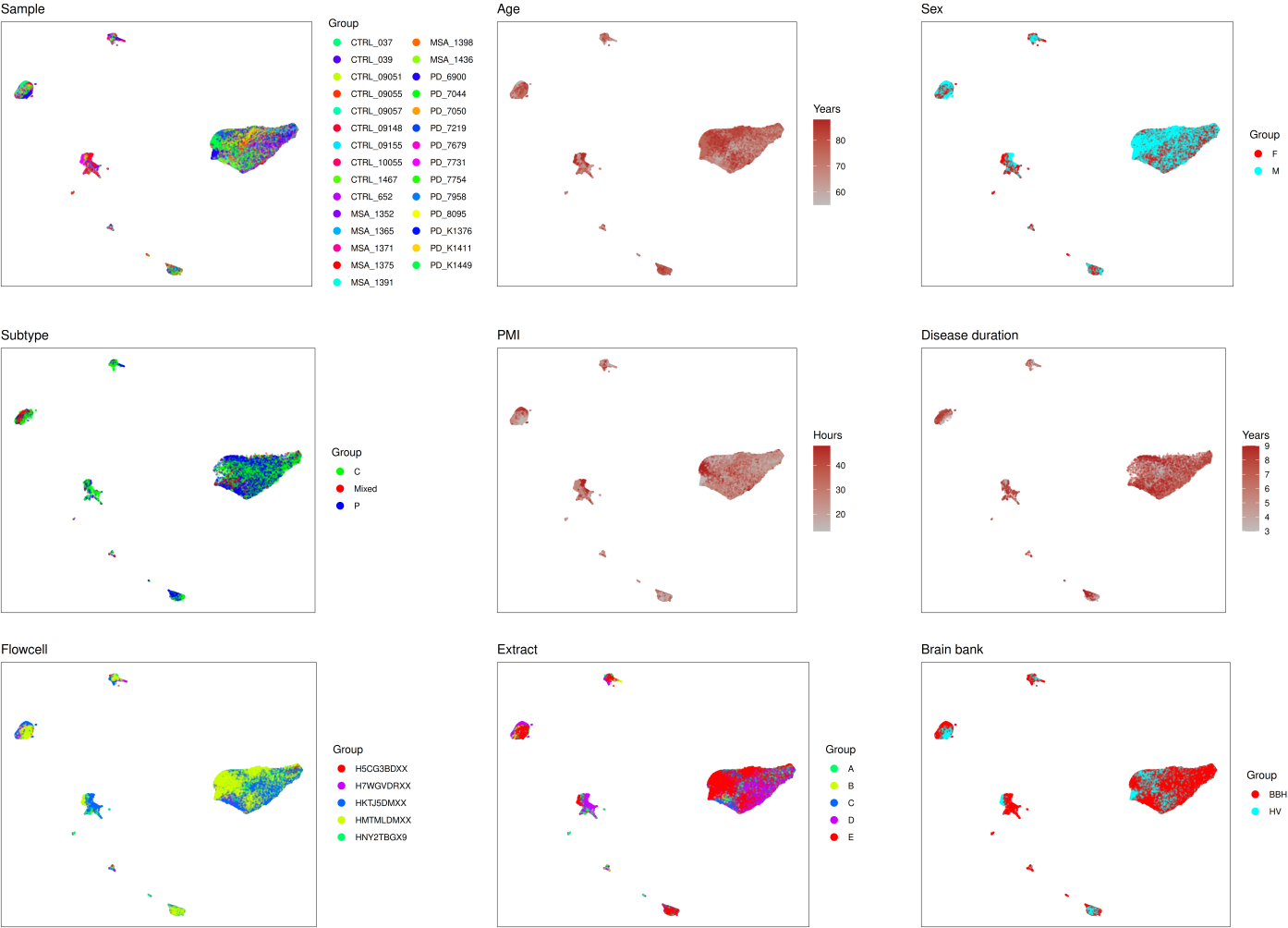


c



Supplementary Figure 2 - Ambient RNA removal and UMAP embedding

Overview of results from ambient RNA removal using CellBender129 calculated on raw alignment data. The CellBender pipeline assigns barcodes as cells or nuclei in a slightly different manner than Cell Ranger, explaining why the number of cells is somewhat different between the two methods as seen in **a**. **a** Estimated number of cells by CellBender (top) or expected cells estimated by Cell Ranger (second from top). Differences are summarized as nominal differences (second from bottom) or proportional differences (bottom). **b** Top ambient genes plotted as proportion of samples with each ambient gene. **c** Expression of marker genes for the major cell types, visualized in the UMAP embedding: astrocytes (AQP4), microglia (CSF1R), GABAergic neurons (GAD1), oligodendrocytes (MOG), perivascular macrophages (MRC1), pericytes/endothelial and immune cells (PDGFRB), medium spiny neurons (PPP1R1B), immune cells (PTPRC), neurons (RBFOX3), GLUergic neurons (SLC17A7), oligodendrocyte precursor cells (VCAN); CTRL: healthy controls; MSA: multiple system atrophy; PD: Parkinson's disease.



Supplementary Figure 3 - Metadata distribution

Distribution of nuclei colored by metadata. Nuclei are colored by condition (CTRL: healthy controls; MSA: multiple system atrophy; PD: Parkinson's disease), age (in years), brain bank source (BBH: Bispebjerg Hospital Brain Bank; HV: Harvard Brain Tissue Resource Center), tissue processing batch (extract), sequencing flowcell, subtype of multiple system atrophy either P for parkinsonian type or C for cerebellar type, post-mortem interval (PMI; in hours), sex (F: female; M: male), or disease duration (in years).

Condition CTRL - MSA

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

Age

color
55
60
65
70
75
80
85

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

Disease duration

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

Extract batch

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

Flowcell

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

PMI

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

Sex

color
• F
• M

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

Subtype

color
• C
• Mixed
• P

condition
• CTRL
▲ MSA

Num. cells
● 2000
● 3000
● 5000

Association significance

$-\log_{10}(\text{adj. } p)$

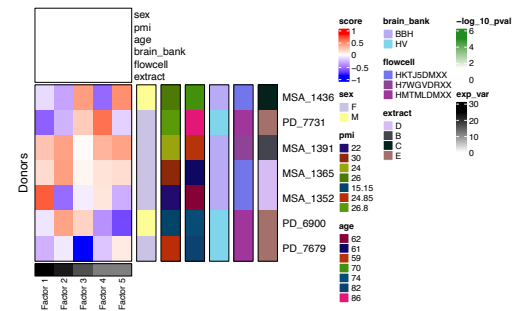
age, disease_duration, extract, flowcell, pmr, sex, subtype

Figure 1 displays a series of scatter plots showing the relationship between various clinical and genetic factors and the number of cells in the MSA and PD conditions. The plots are arranged in a 2x10 grid. The top row shows the relationship between the number of cells and the condition (MSA or PD) for various factors: Age, PMI, Brain bank, Sex, Disease duration, Subtype, Extract batch, and Association significance. The bottom row shows the relationship between the number of cells and the condition (MSA or PD) for various factors: Flowcell, Age, PMI, Brain bank, Sex, Disease duration, Subtype, Extract batch, and Association significance. The y-axis for all plots is 'Num. cells' (1000, 3000, 5000). The x-axis for all plots is 'condition' (MSA, PD). The 'Association significance' plot shows the relationship between the number of cells and the condition (MSA or PD) for various factors: age, brain_bank, disease_duration, extract, flowcell, pmr, sex, and subtype. The y-axis for this plot is $-\log_{10}(\text{adj. } p)$ (0.0, 0.5, 1.0, 1.5, 2.0). A red horizontal line indicates a threshold at $-\log_{10}(\text{adj. } p) = 1.3$.

Supplementary Figure 4 - Visualization of covariates on UMAPs

UMAP embeddings showing metadata-based separation of samples colored by condition for healthy control (CTRL) and **a** multiple system atrophy (MSA), or **b** Parkinson's disease (PD) or **c** between diseases. age (in years), brain bank source (BBH: Bispebjerg Hospital Brain Bank; HV: Harvard Brain Tissue Resource Center), tissue processing batch (extract), sequencing flowcell, subtype of multiple system atrophy either P for parkinsonian type or C for cerebellar type, post-mortem interval (PMI; in hours), sex (F: female; M: male), or disease duration (in years). Significance of associations was calculated using Cacao's `estimateMetadataSeparation` that employs a two-sided permutation test (see Methods) followed by Benjamini-Hochberg correction. PVMs: perivascular macrophages; OPCs: oligodendrocyte precursor cells.

Neurons



Heatmap visualization showing donor characteristics and gene expression data. The main heatmap displays expression levels for 10 donors (rows) across 10 genes (columns). The genes are: HKT1J5DMXX, H7WQVDRXX, HNY2TGBXX, and HMT1MLDMXX. The donor characteristics are color-coded by sex (M/F), PMI (0-30), age (22-84), brain bank (BBH/HV), extract (D/C/E), log10 pval (0-6), exp var (0-15), and flowcell (1-6).

Heatmap visualization showing gene expression data across 15 donors (rows) and 15 genes (columns). The color scale represents expression levels, ranging from 0 (blue) to 1 (red). The genes are listed on the right, and the donors are listed on the left.

Genes (Columns):

- PD_7958
- PD_7731
- PD_7050
- PD_7219
- PD_7679
- PD_K1449
- PD_7754
- PD_K1411
- MSA_1398
- MSA_1436
- MSA_1365
- PD_7044
- PD_6900
- PD_K1376
- PD_K1376

Donors (Rows):

- Donor 1
- Donor 2
- Donor 3
- Donor 4
- Donor 5
- Donor 6
- Donor 7
- Donor 8
- Donor 9
- Donor 10
- Donor 11
- Donor 12
- Donor 13
- Donor 14
- Donor 15

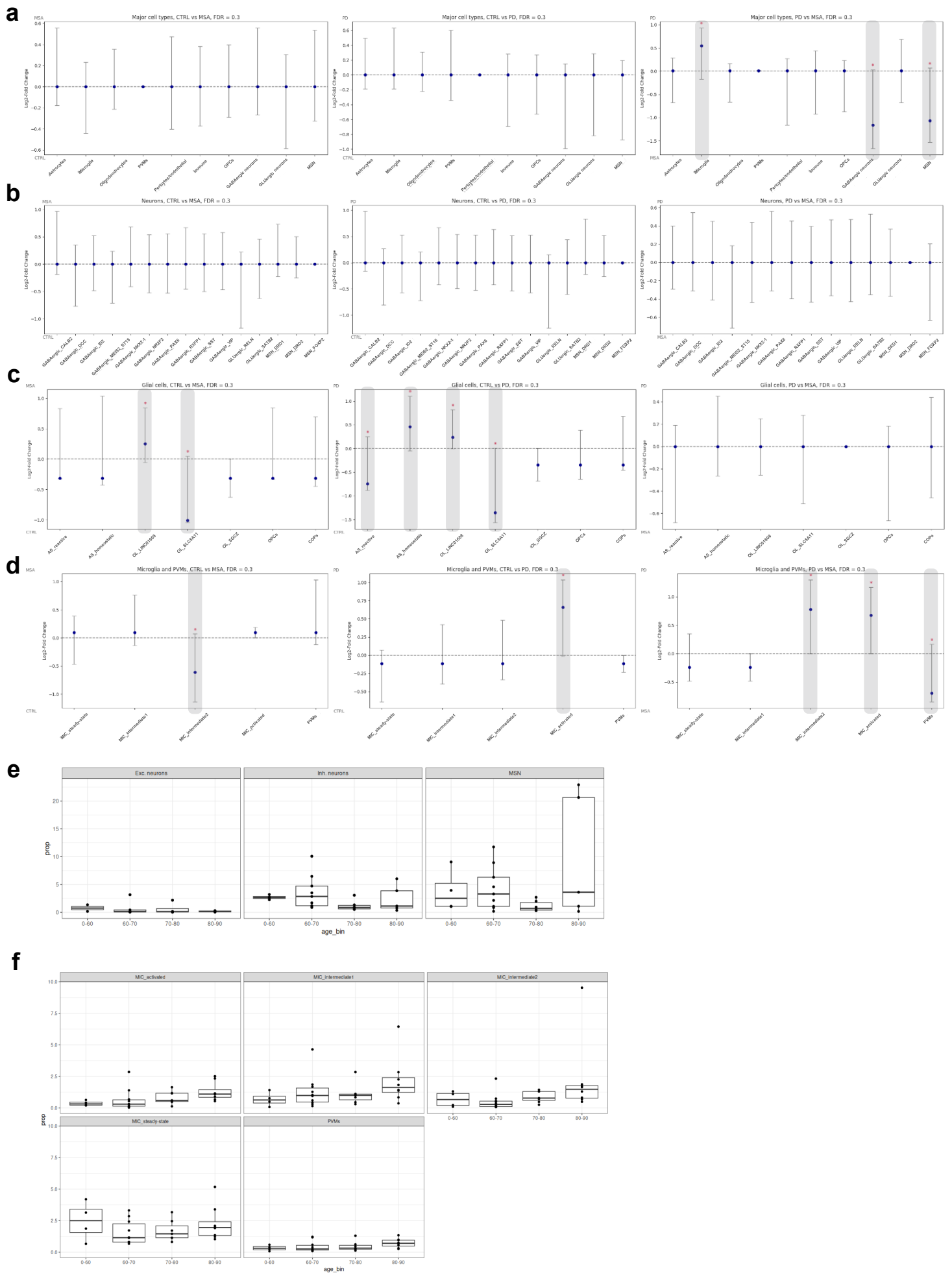
Color Scale:

- 0 (Blue)
- 1 (Red)

[illegible]

Supplementary Figure 5 - Covariate impact

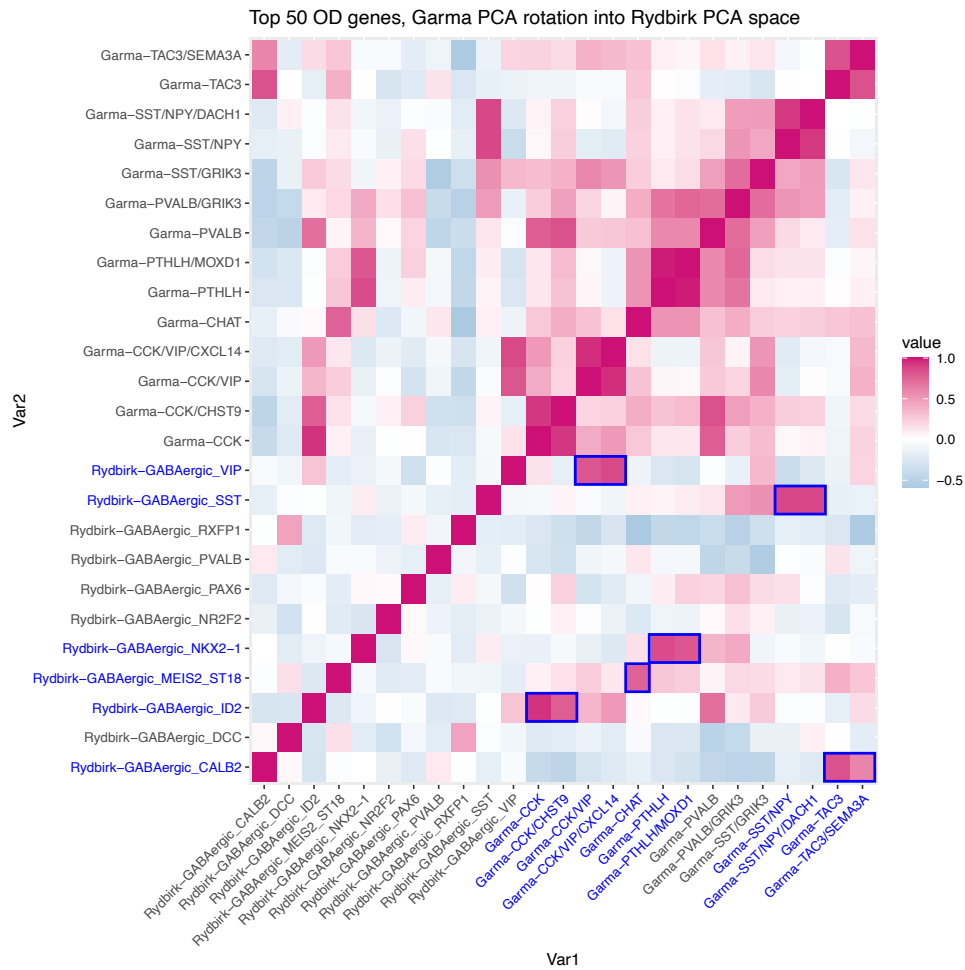
Tensor decomposition results from scITD analysis with plots showing the variation captured by the top factors across different cell types and conditions. Analyses were divided into estimations for each set of conditions investigated in the downstream analyses described in the present study per major cell type. Please note, the brain bank covariate could not be included for estimations for healthy controls (CTRL) and multiple system atrophy (MSA) patients as all samples originated from the same brain bank. M: male, f: female, pmi: post-mortem interval in hours, age in years.



Supplementary Figure 6 - Compositional analysis by scCODA

Compositional changes estimated using scCODA as a validation of the compositional changes identified using CoDA in Cacao. ScCODA models cell-type abundances with a hierarchical Dirichlet–Multinomial Bayesian framework with spike-and-slab priors to identify credible condition-associated shifts. Estimations were performed on **a** major cell types, **b** neurons, **c** glial cells, or **d** microglia (MIC) and perivascular macrophages (PVMs). Since scCODA is more conservative than CoDA in Cacao, we lowered the false discovery rate (FDR) to 0.3 to highlight the most significant changes. Significant differences are highlighted in gray and with an asterisk. The estimated log2 fold changes are shown on the second axis, bars show 3-97% highest density interval, dots show median values. Proportions of cells binned according to age for **e** major neuronal cell types, or **f** microglial cell subtypes. In neither **e** nor **f** did we observe an age-dependent effect of cell proportions (adj. $p > 0.05$, Kruskal-Wallis test with Benjamini-Hochberg post-hoc). Box plots show 1st and 3rd quantiles and medians, whiskers extend to the smallest and largest values within 1.5 x IQR, and points beyond the whiskers represent outliers. AS_homeostatic: homeostatic astroglia; AS_reactive: reactive astroglia; OL: oligodendroglia; OPCs: oligodendrocyte precursor cells; COPs: committed OPCs; MSN: medium spiny neurons; CTRL: healthy controls; MSA: multiple system atrophy; PD: Parkinson's disease.

a



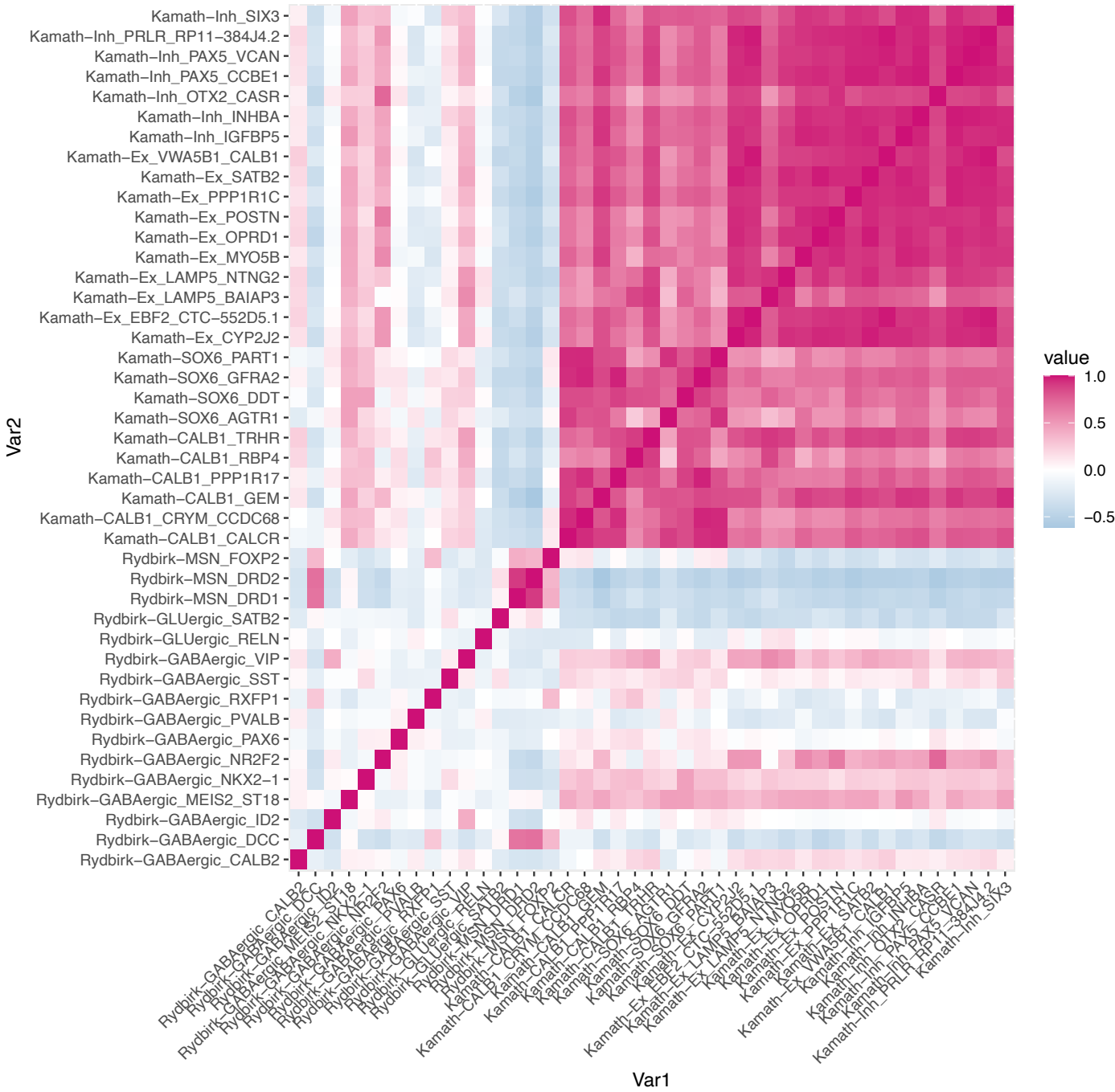
b



Supplementary Figure 7 - Data-driven overlap of neuron subtypes on a striatum dataset

Similarities between our neuronal annotation and the annotation on interneurons from the striatum used by Garma et al. **a** Cosine similarities between annotations from the two studies where high values/warm colors indicate high similarity. For computational approach, see Methods. As expected, SST and VIP subtypes show high similarity between studies, whereas our GABAergic_ID2 subtype resembles Garma's CCK subtypes, our GABAergic_CALB2 subtype resembles Garma's TAC3 subtypes, our GABAergic_MEIS2_ST18 subtype resembles Garma's CHAT subtype, and our GABAergic_NKX2-1 subtype resembles Garma's PTHLH subtype. **b** Dot plot showing expression of subtype markers from both studies on an integrated dataset containing (inter)neurons from both studies.

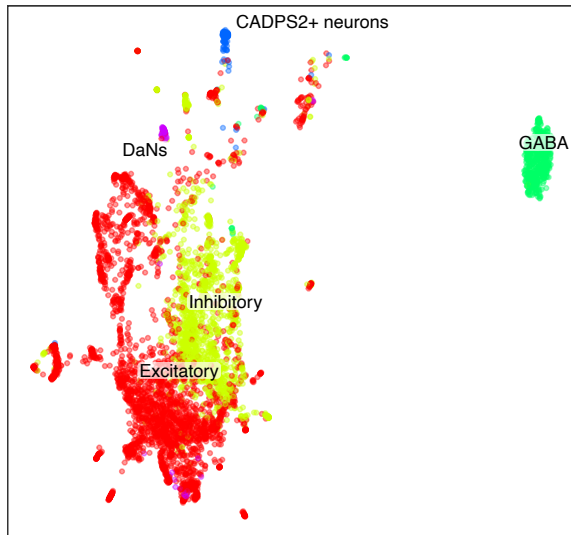
Top 100 OD genes, Kamath PCA rotation into Rydbirk PCA space



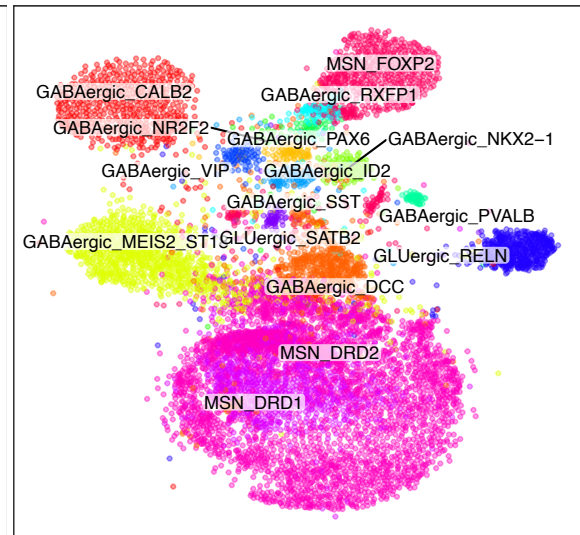
Supplementary Figure 8 - Dataset-driven separation of neuron subtypes on a substantia nigra dataset

Similarities between our neuronal annotation and the annotation used by Kamath et al. on neurons from the substantia nigra pars compacta. Cosine similarities between annotations from the two studies where high values/warm colors indicate high similarity. For computational approach, see Methods. A few of our neuronal subtypes showed weak similarity to almost all the neuronal subpopulations in the dataset from Kamath et al. highlighting the regional specificity of neuronal subtypes.

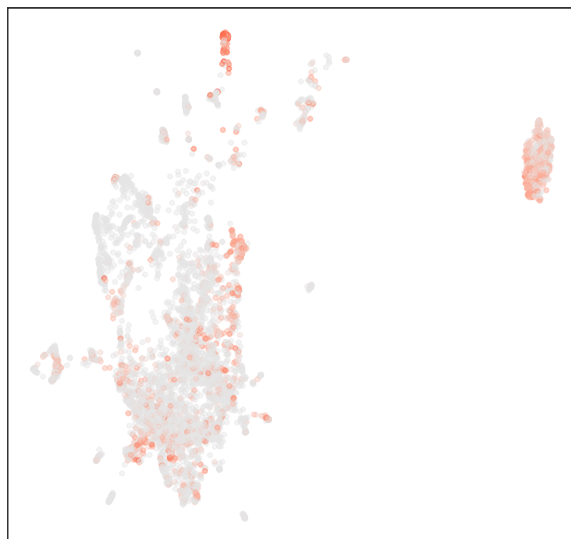
Smajic neurons, annotation



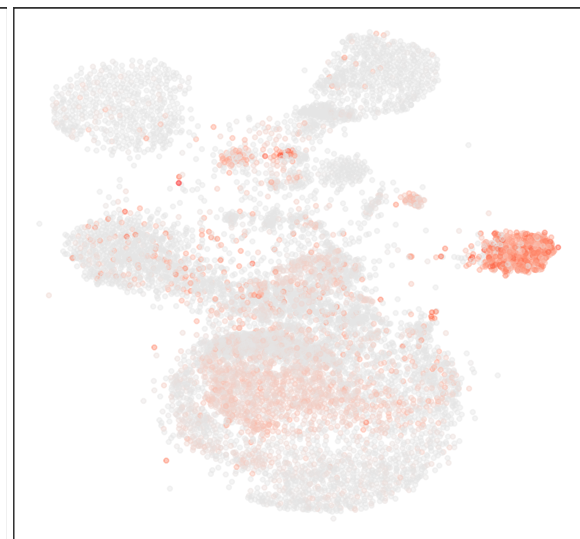
Rydbirk neurons, annotation



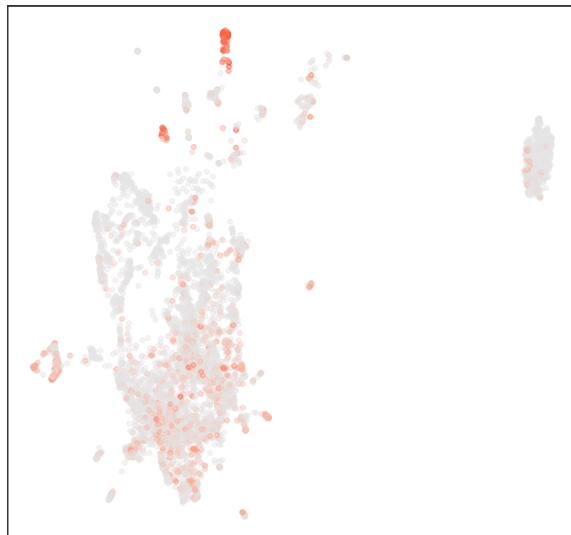
RELN expression



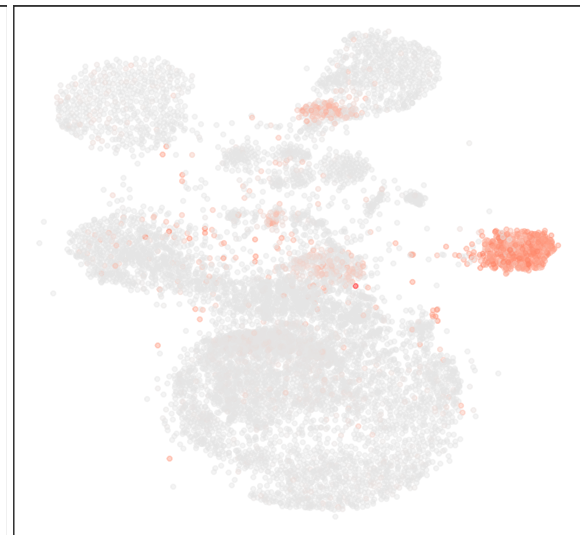
RELN expression



CADPS2 expression



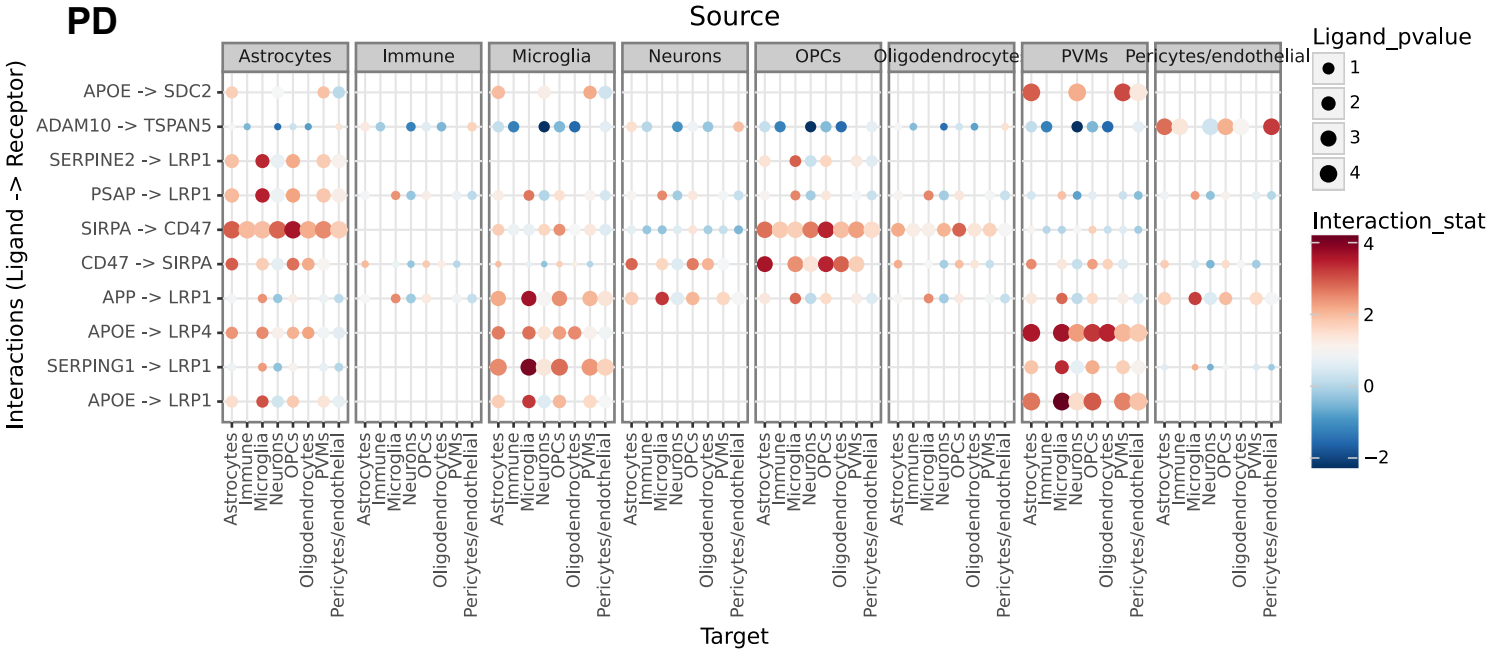
CADPS2 expression



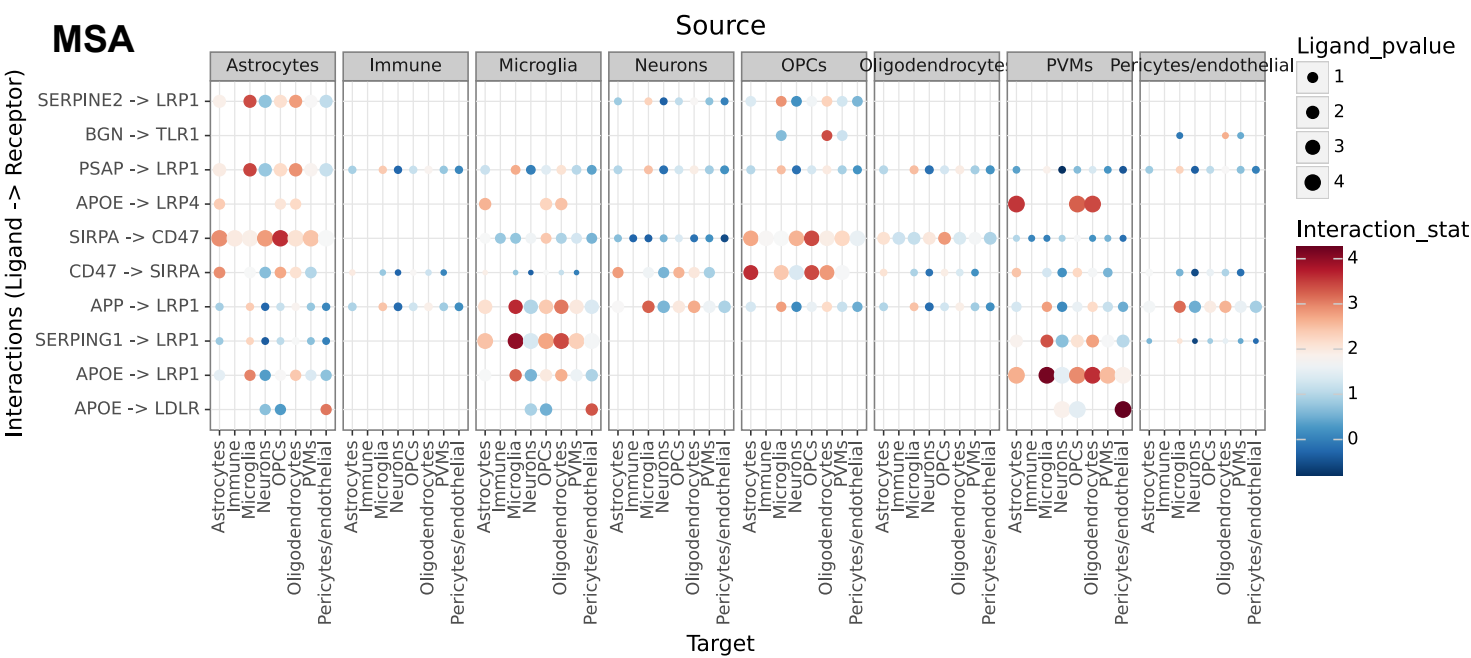
Supplementary Figure 9 - Cross validation of GLUergic_RELN subtype on a second dataset

Cross validation of our GLUergic_RELN neuronal subtype in the dataset from Smajic et al. Smajic data are shown on the left, our data are shown on the right. We isolated the neuronal cells in both datasets and plotted the related annotations (top). We observed that cells expressing high levels of RELN (middle) also express high levels of CADPS2 (bottom) underlining how these cells are shared across datasets.

a

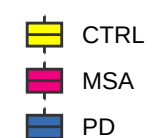
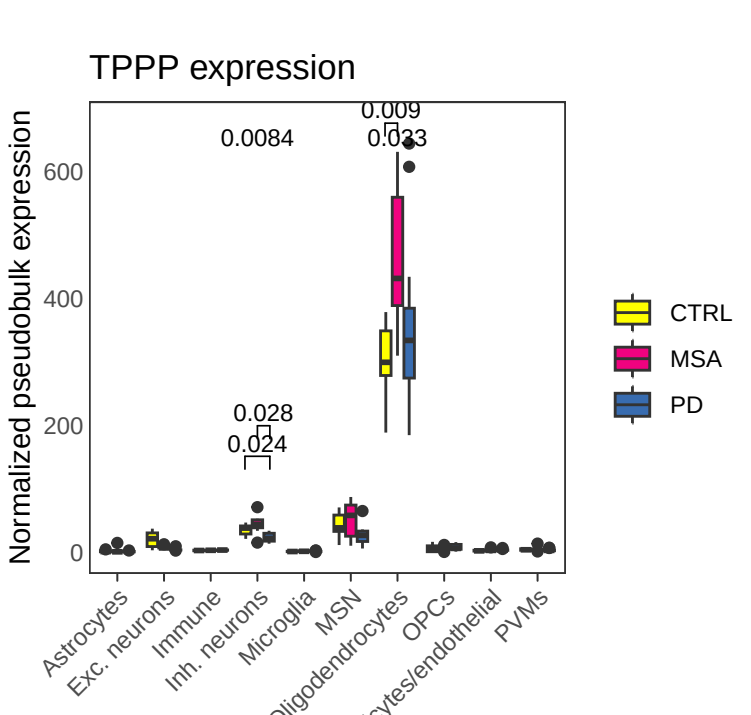
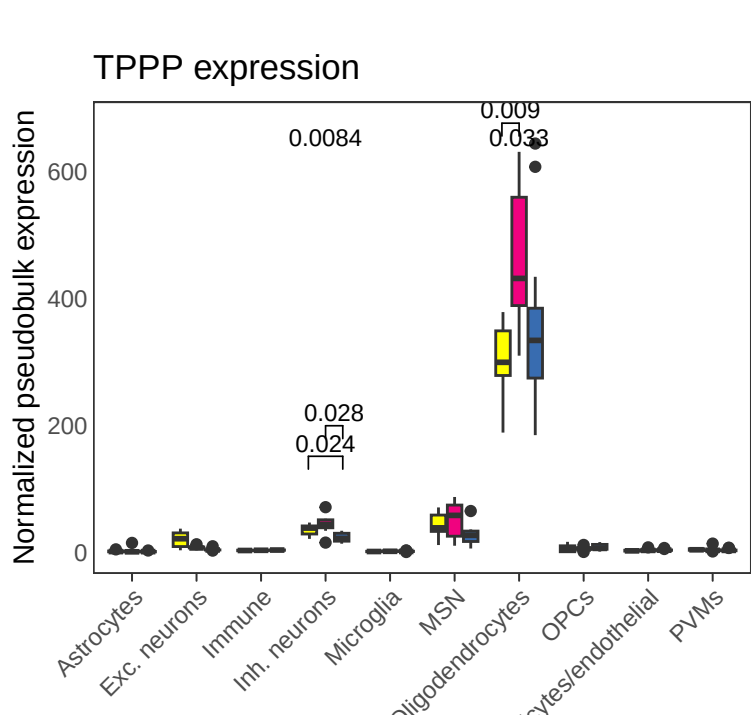
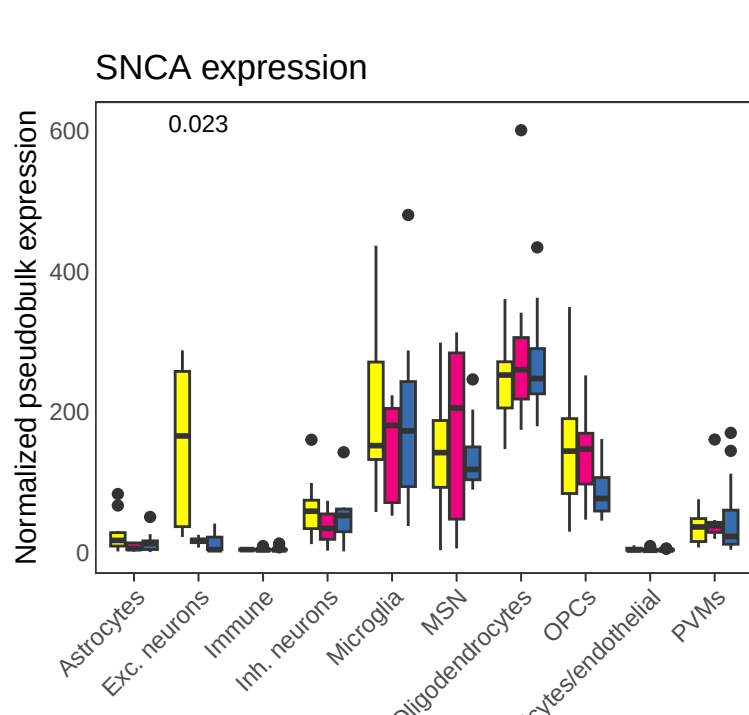
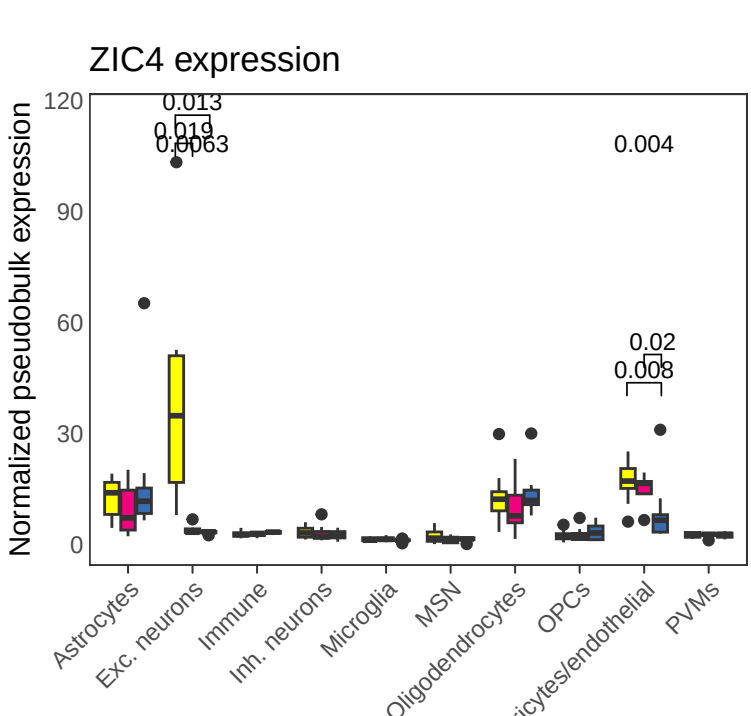
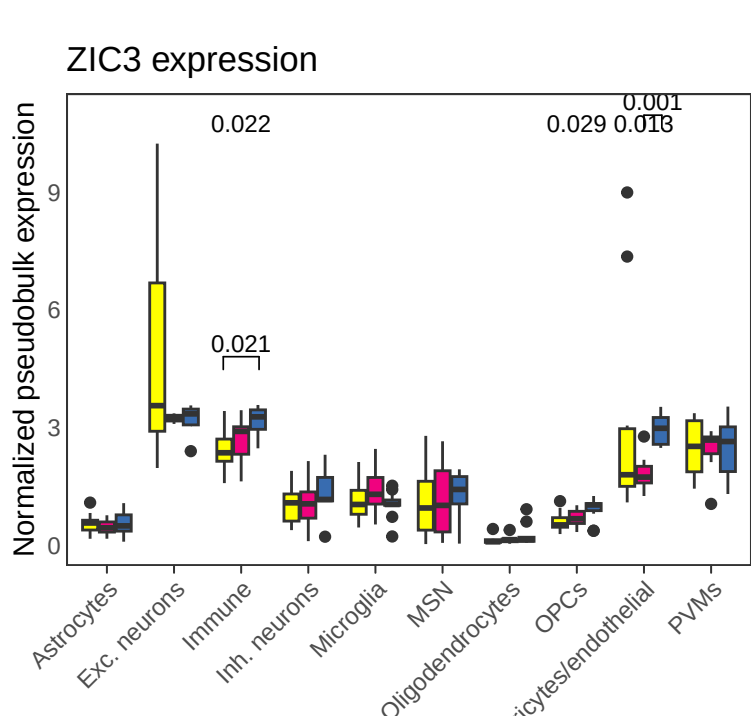
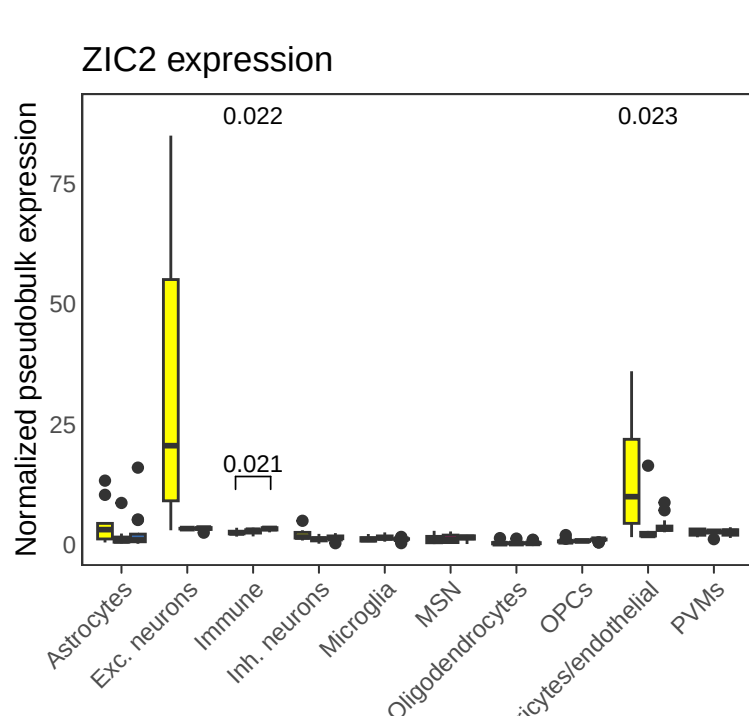
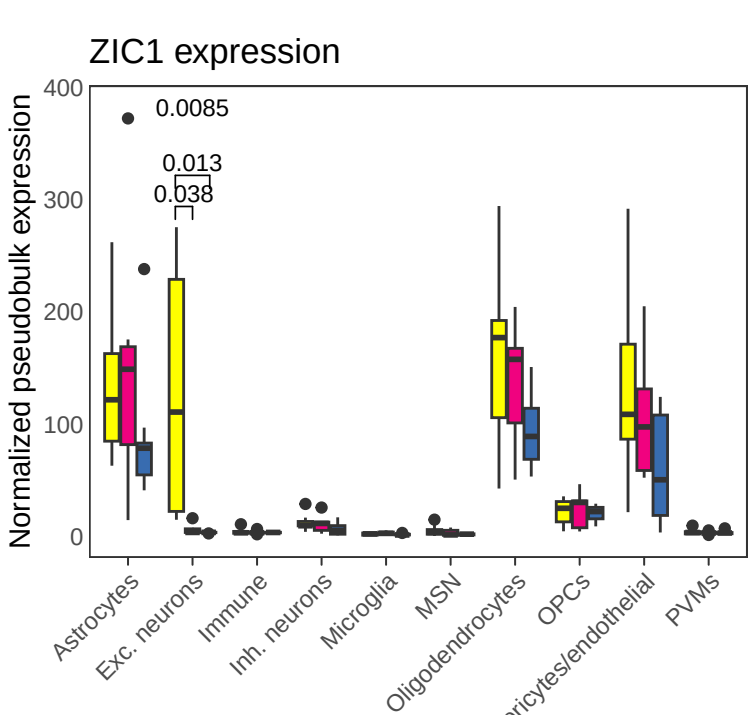
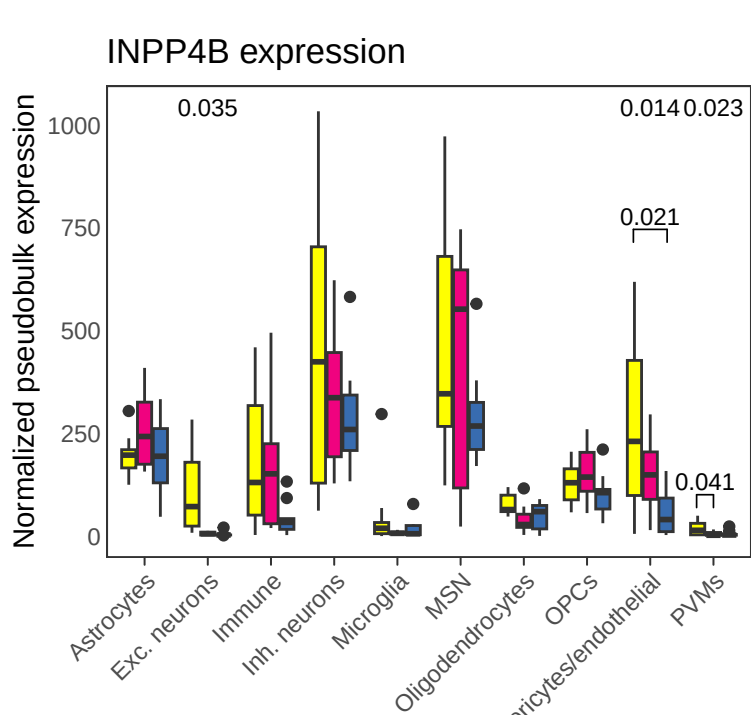
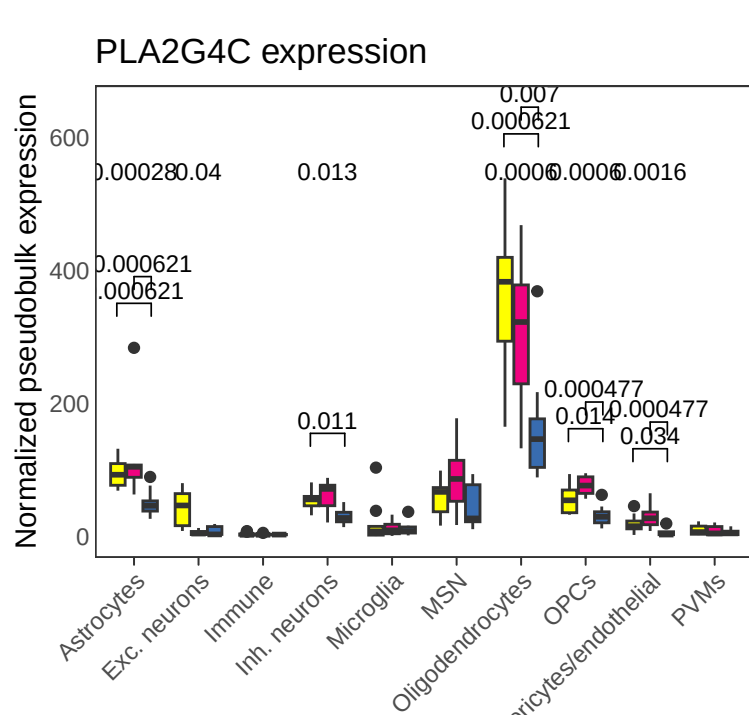
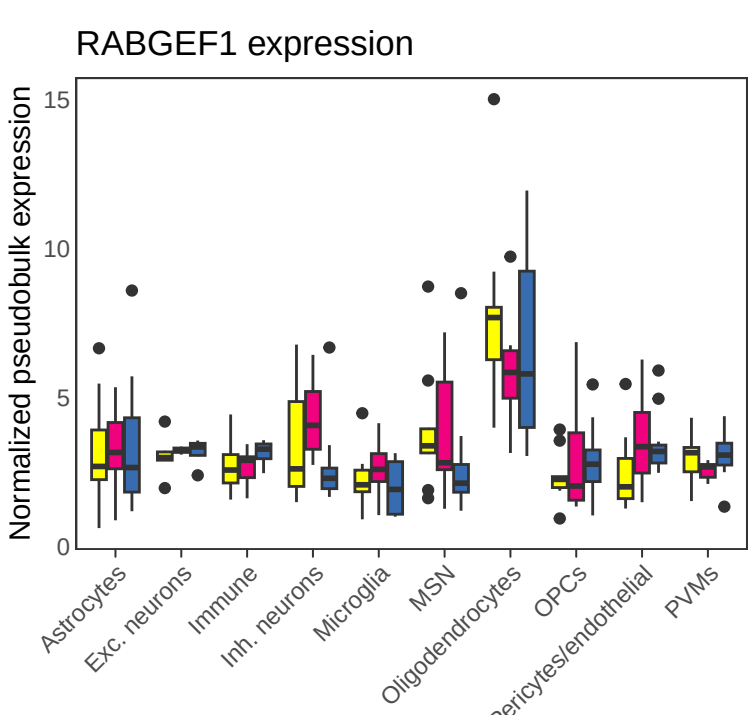
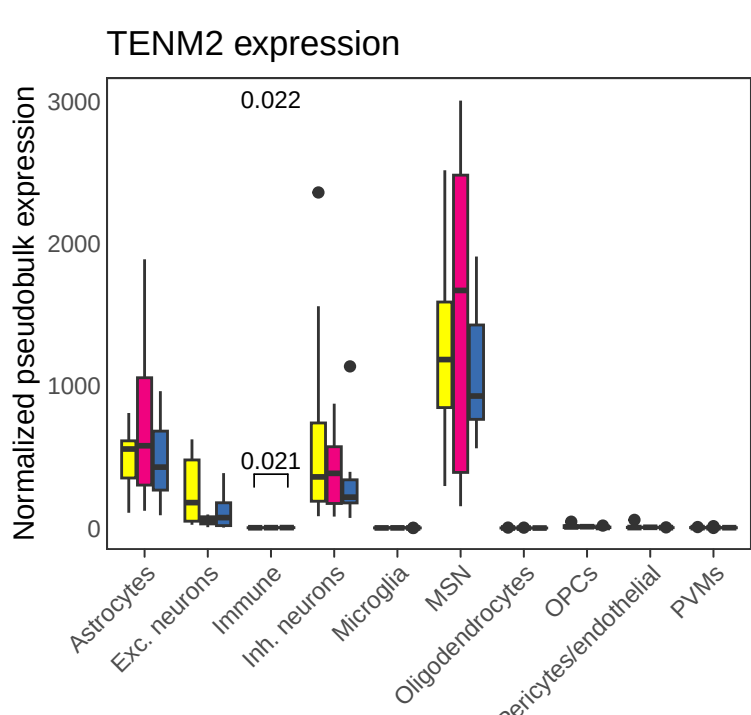
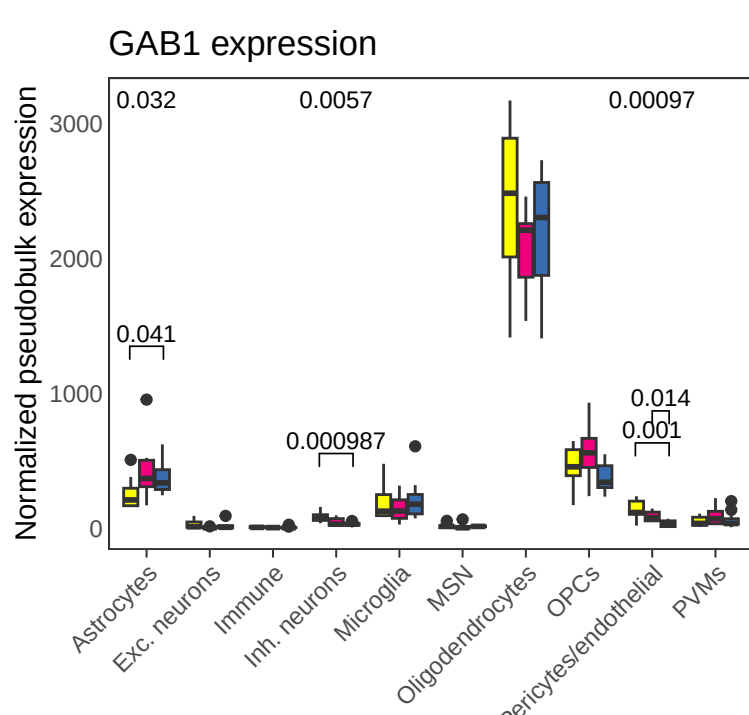
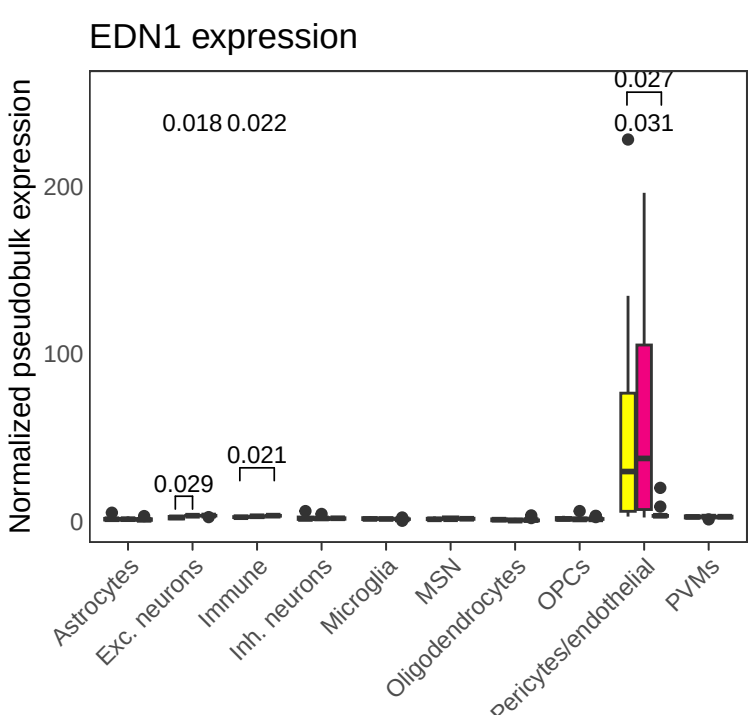
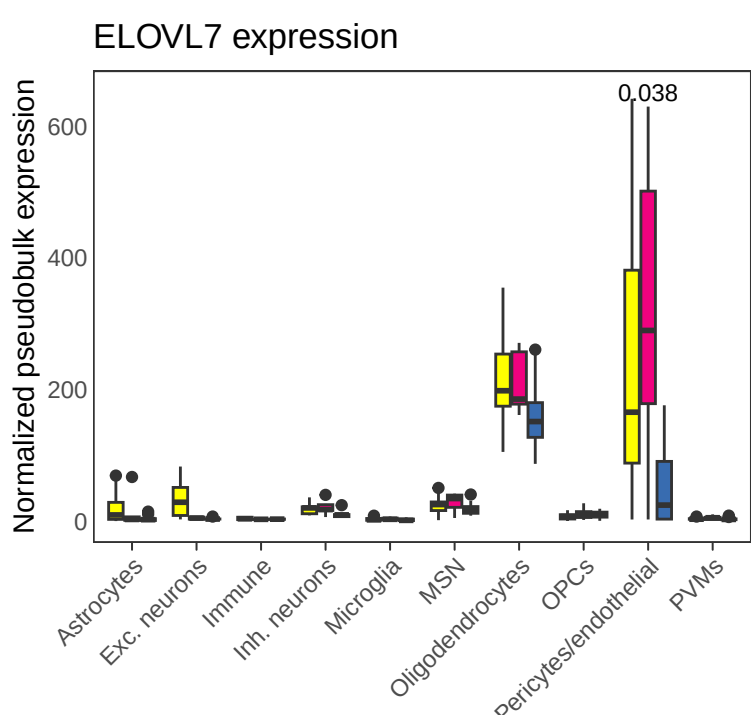
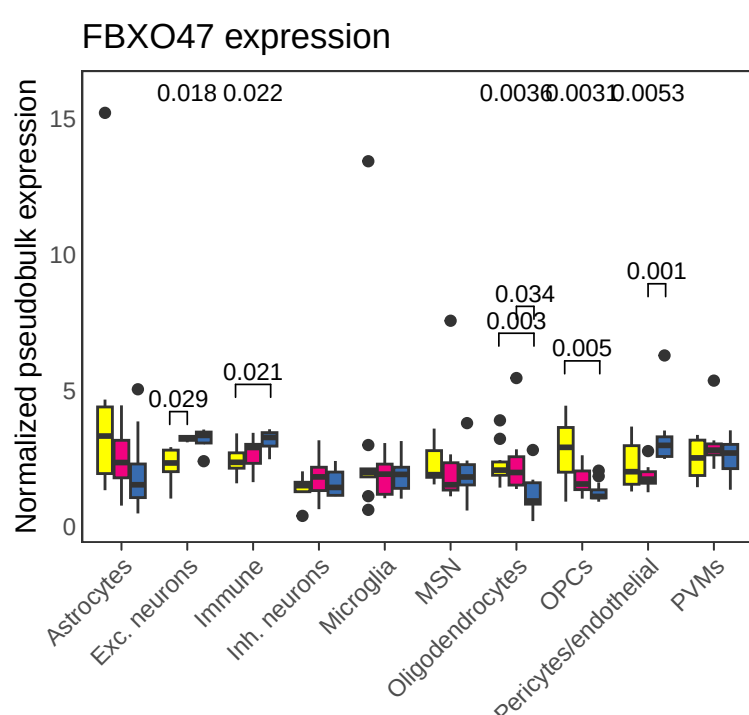
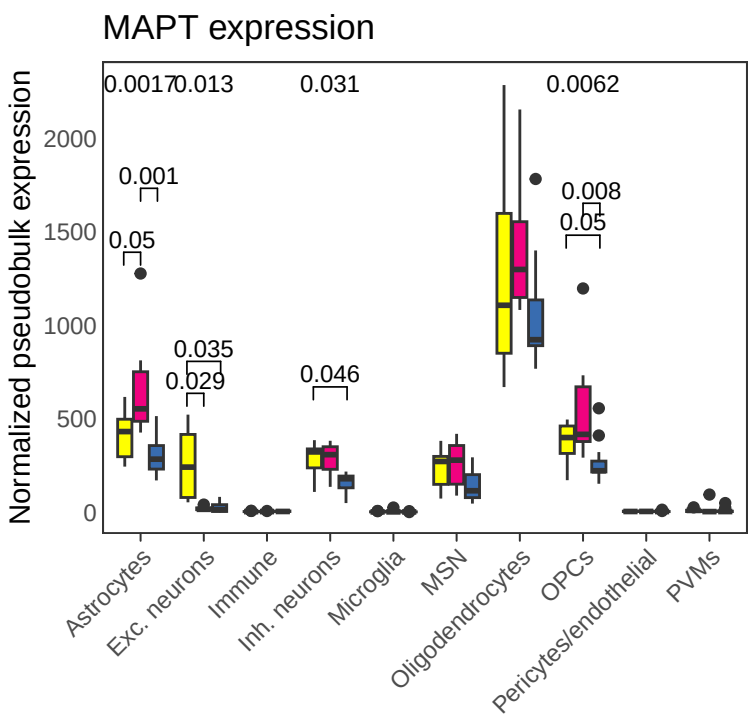
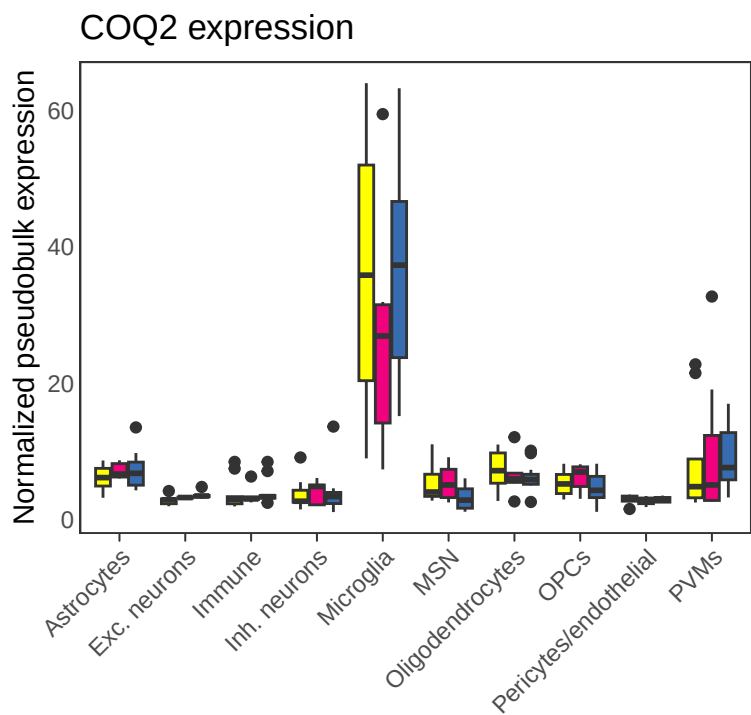
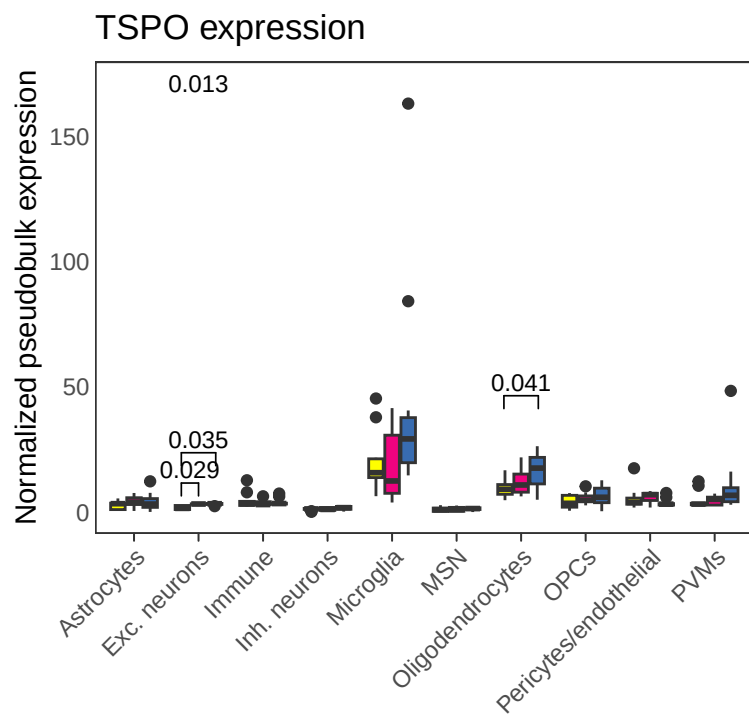


b



Supplementary Figure 10 - Cell-cell interaction analysis

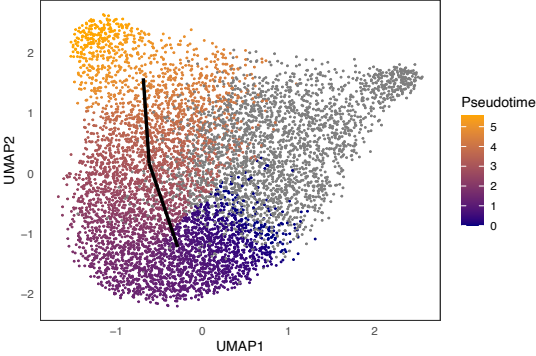
Dot plot showing LIANA cell-cell interaction analysis results for **a** healthy controls (CTRL) vs. Parkinson's disease (PD) or **b** CTRL vs multiple system atrophy (MSA). The results were based on LIANA's OmniPath-based ligand-receptor resources, and the ranked interaction scores are used to display the top interactions by magnitude and specificity. OPCs: oligodendrocyte precursor cells, PVMs: perivascular macrophages.



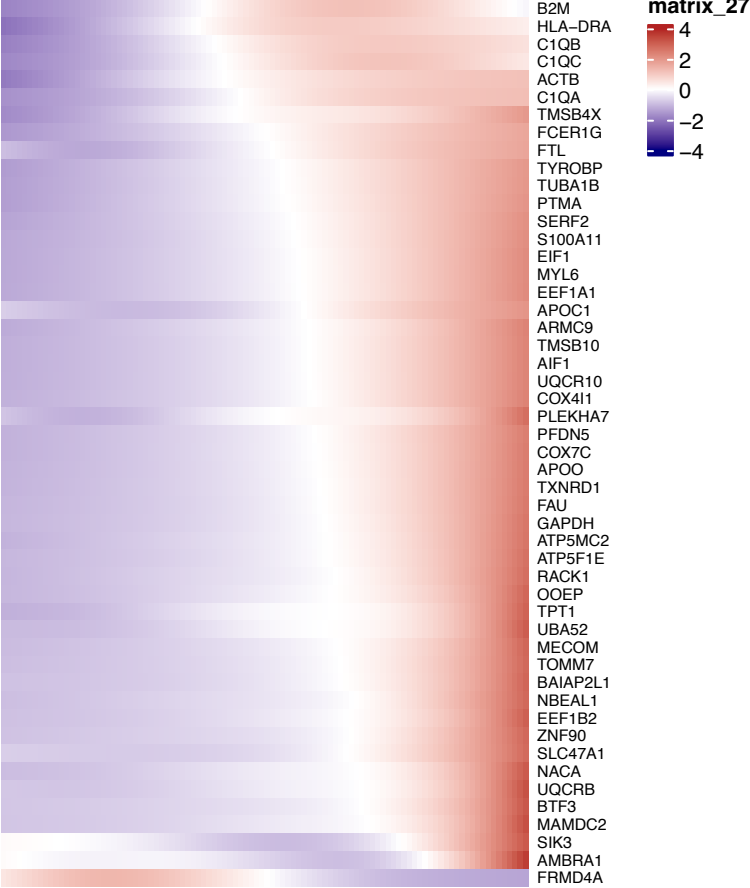
Supplementary Figure 11 - Differential expression of genes of interest

Selected genes expression for all major cell types. Significance was assessed using the one-way Kruskal-Wallis omnibus test followed by the Wilcoxon test. Box plots show 1st and 3rd quantiles and medians (center line), whiskers extend to the smallest and largest values within $1.5 \times \text{IQR}$, and points beyond the whiskers represent outliers. CTRL (n = 10): healthy controls, MSA (n = 7): multiple system atrophy, PD (n = 12): Parkinson's disease; OPCs: oligodendrocyte precursor cells; MSN: medium spiny neurons; PVMs: perivascular macrophages; exc. neurons: excitatory neurons; inh. neurons: inhibitory neurons.

a

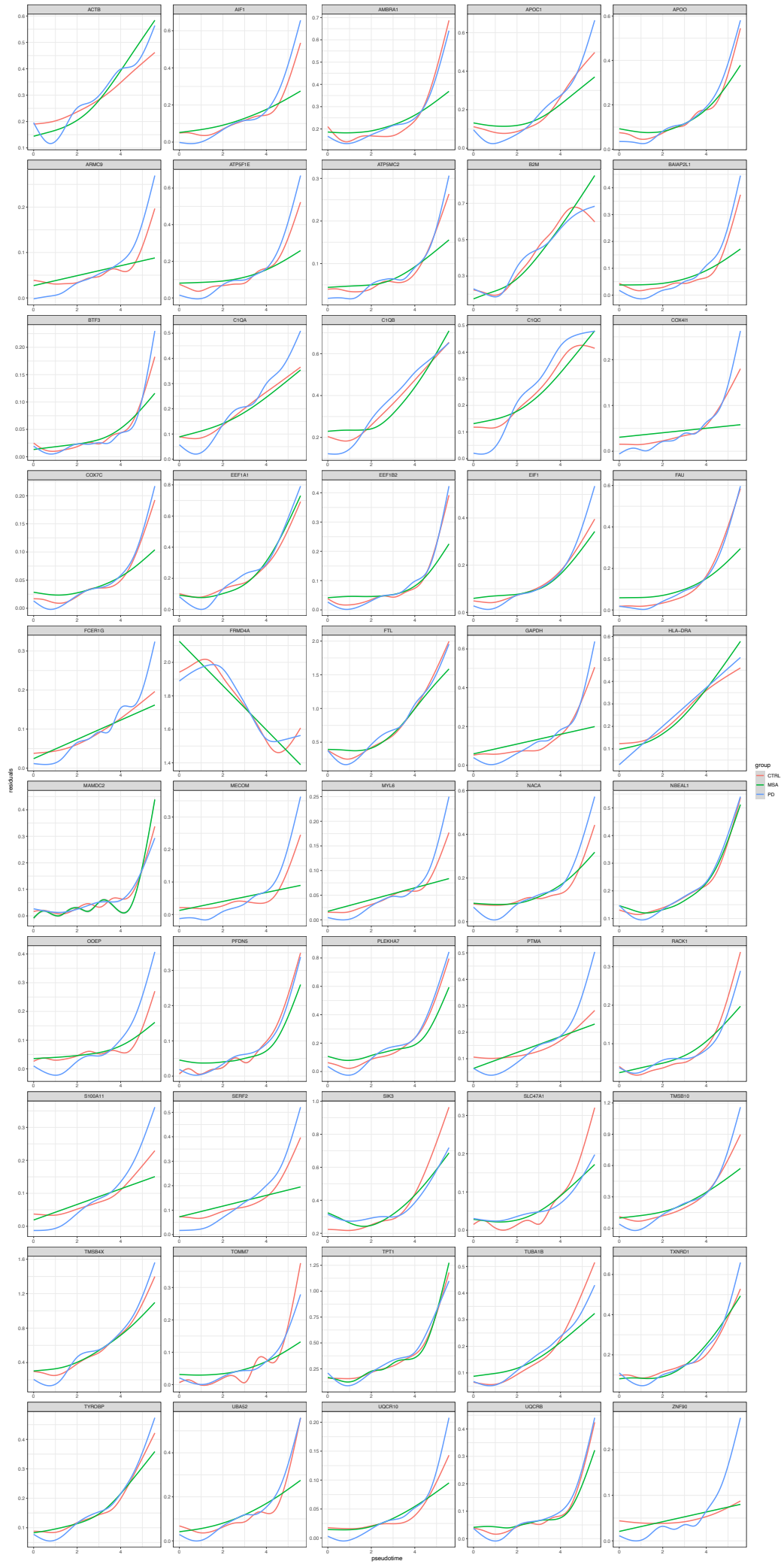


b



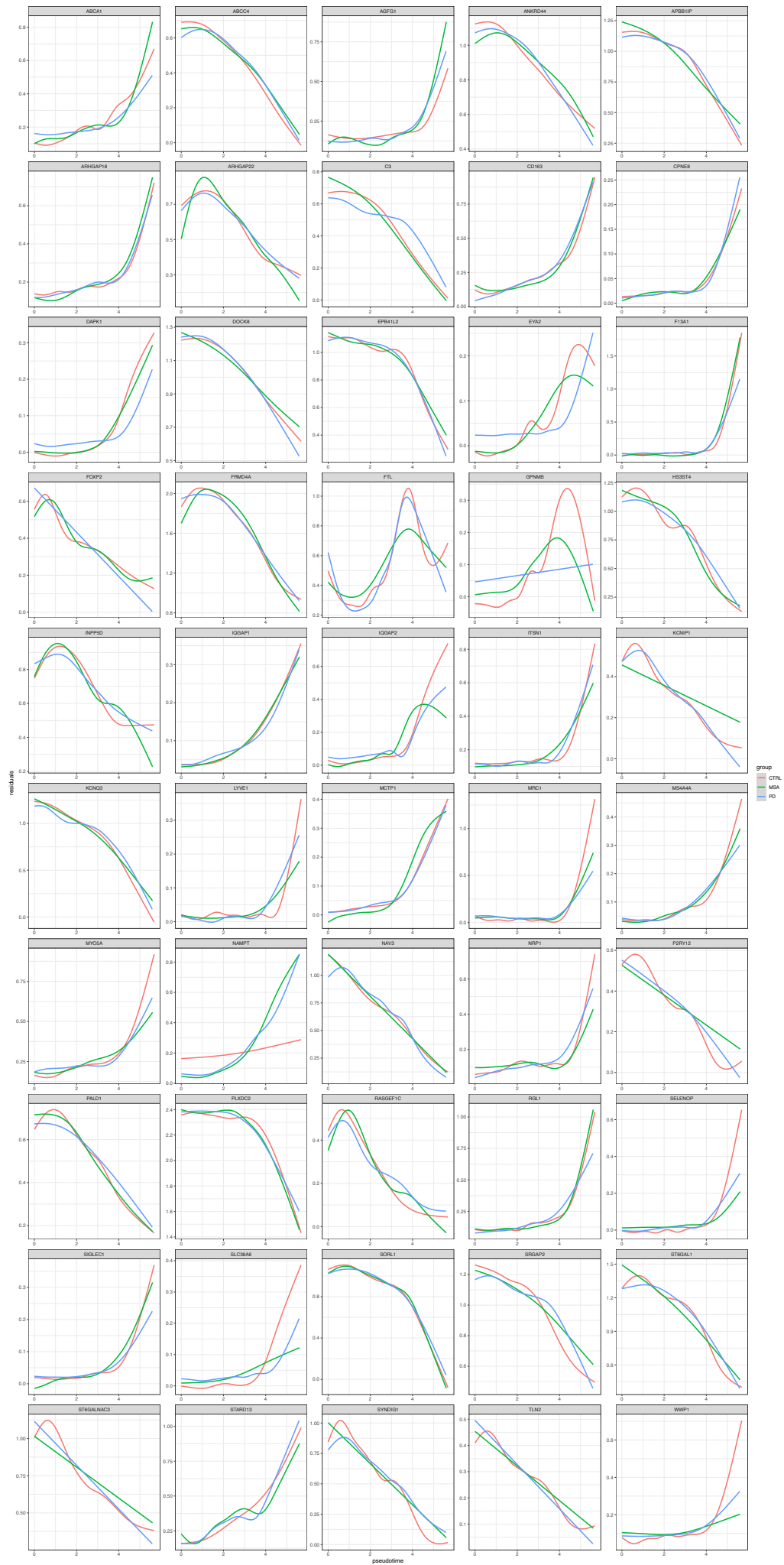
Supplementary Figure 12 - Microglia trajectories analysis

a Pseudotime shown for microglia trajectory towards activated microglia. Gray nuclei are not in the highlighted trajectory. **b** Smoothed heatmap of top 50 pseudotime genes for trajectory towards activated microglia.



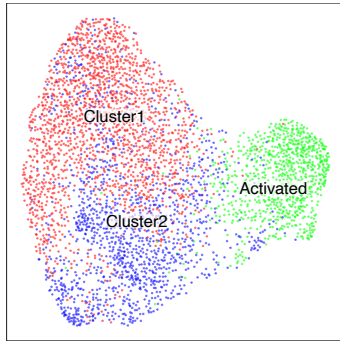
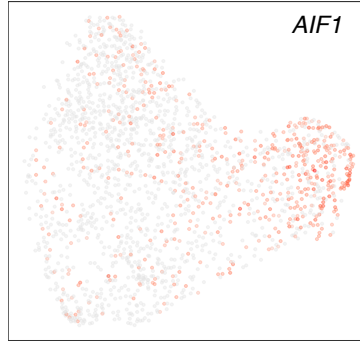
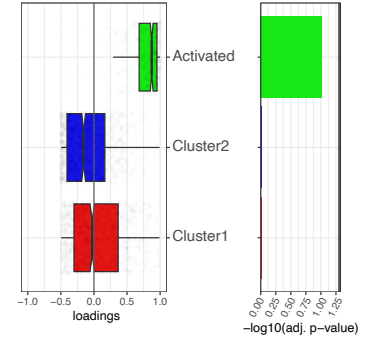
Supplementary Figure 13 - Top pseudotime-related genes activated microglia

Residuals per condition against pseudotime for top 50 pseudotime genes for trajectory from steady-state microglia towards activated microglia. CTRL: healthy controls; MSA: multiple system atrophy; PD: Parkinson's disease.



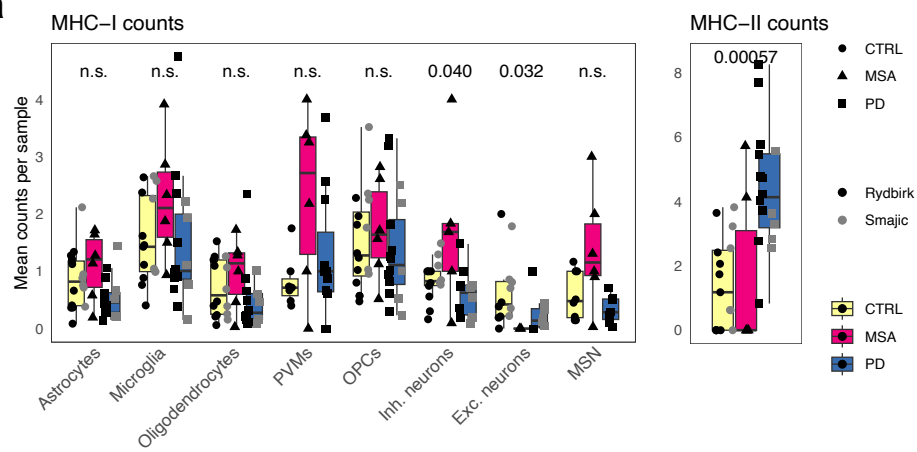
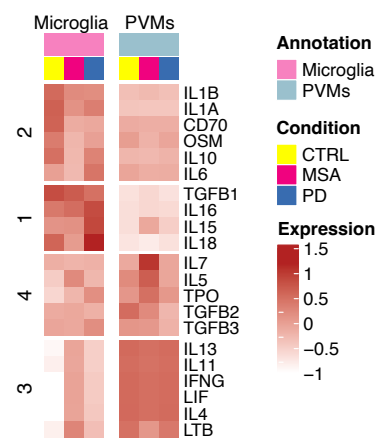
Supplementary Figure 14 - Top pseudotime-related genes PVMs

Residuals per condition against pseudotime for top 50 pseudotime genes for trajectory from steady-state microglia towards perivascular macrophages. CTRL: healthy controls; MSA: multiple system atrophy; PD: Parkinson's disease.

a**b****c**

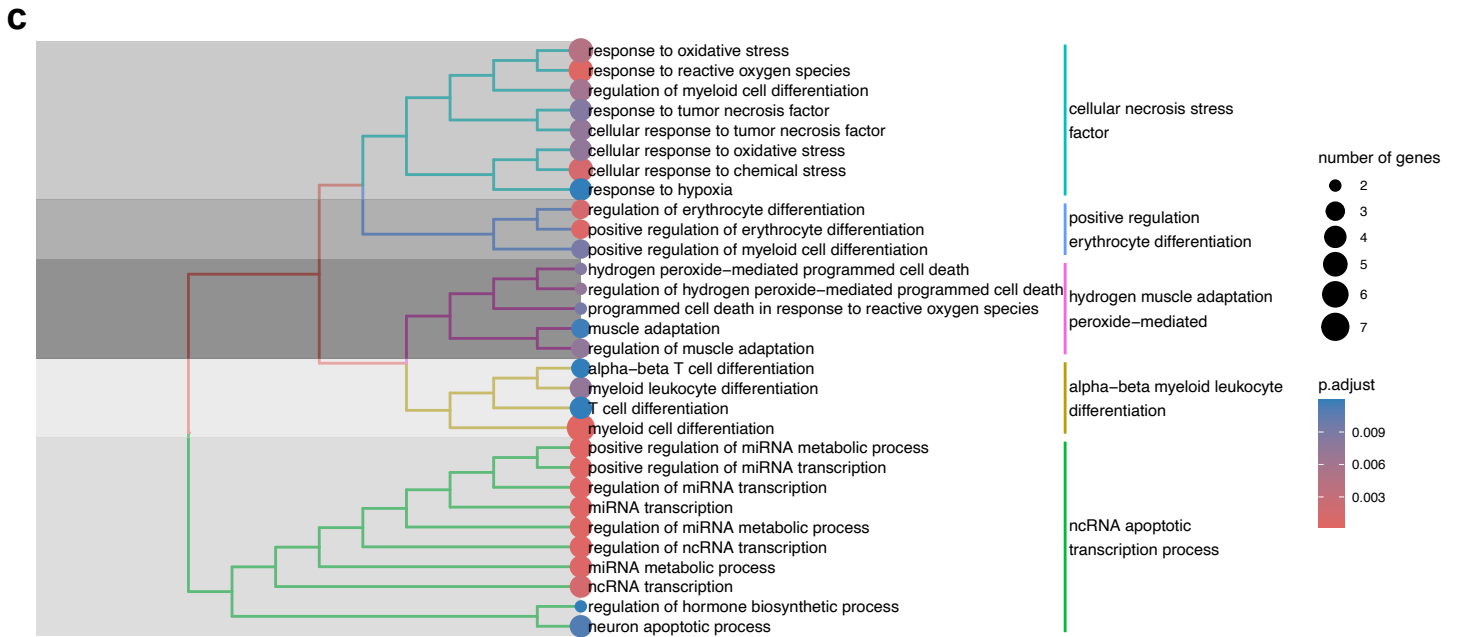
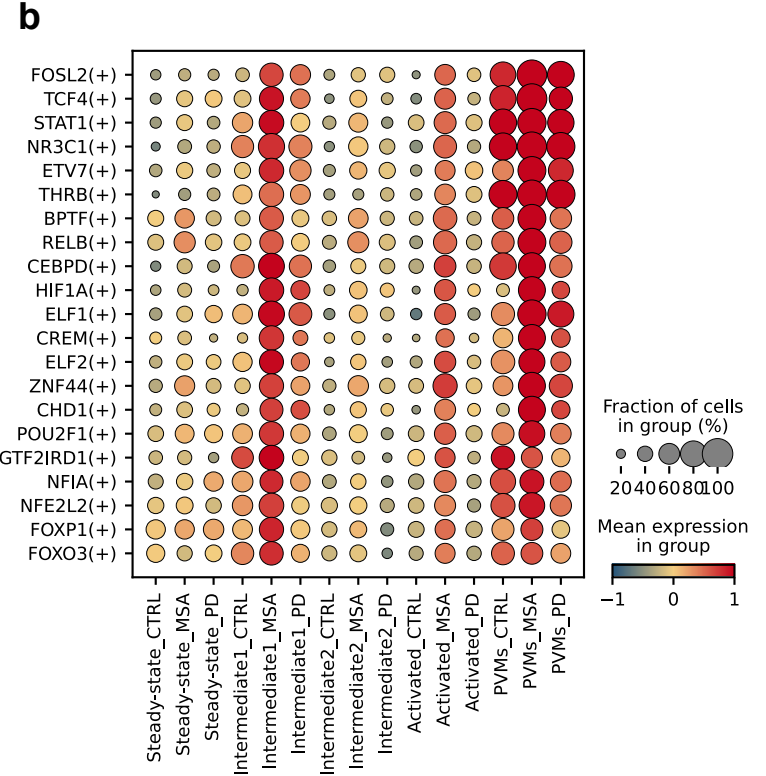
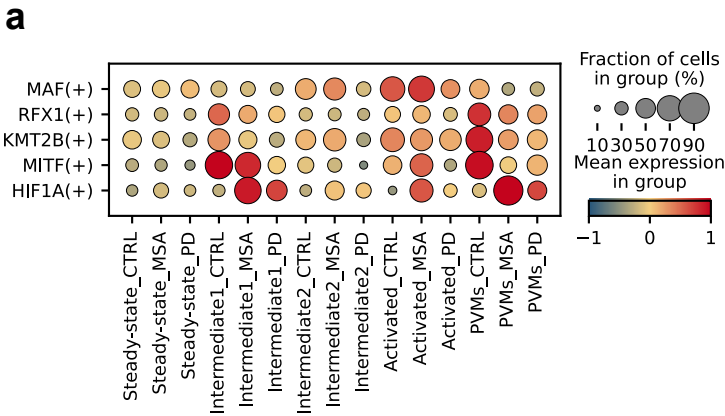
Supplementary Figure 15 - Cross validation of microglia compositional changes on a second dataset

Compositional changes of microglia from Feleke et al. **a** We isolated microglia cells from controls and Parkinson's disease patients in the Feleke dataset and **b** identified activated microglia through their expression of AIF1. **c** Using CoDA in Cacao with 1,000 bootstraps per comparison we estimated compositional changes and identified a nominal significant increase ($p = 0.034$) of activated microglia in Parkinson's disease compared to controls. Significance was assessed using a two-sided permutation test.

a**b**

Supplementary Figure 16 - HLA and cytokine expression

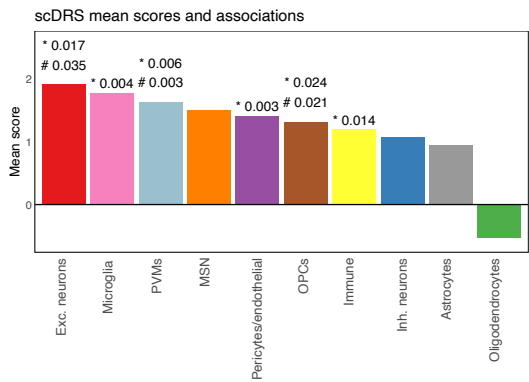
a Molecule counts of human leukocyte antigens in major histocompatibility complex (MHC) class I (left) or class II (right) per cell type. For comparison, publicly available data from Smajic et al. were included. These data underwent the same preprocessing as the one applied to our own data, and labels from our data were transferred to the data from Smajic et al. using built-in functions in Conos. Statistical significance was assessed using the one-way Kruskal-Wallis test on integrated data. MHC-I expression was generally increased in multiple system atrophy (MSA), significant for inhibitory (inh.) neurons. MHC-II expression was significantly increased in Parkinson's disease (PD) microglia, not in MSA. No difference was identified between any comparisons of our data with Smajic et al. per condition per cell type (Wilcoxon signed-rank test, $p > 0.05$). **b** Heatmap showing pseudo-bulk expression per condition of cytokines in microglia and perivascular macrophages (PVMs). Clusters were identified using row-wise clustering in ComplexHeatmap. CTRL: healthy controls; MSA: multiple system atrophy; PD: Parkinson's disease; OPCs: oligodendrocyte precursor cells; MSN: medium spiny neurons.



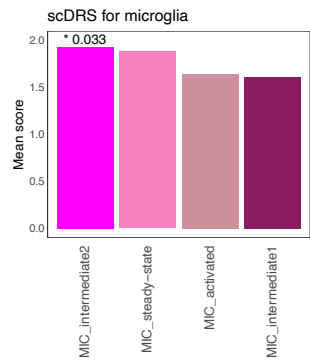
Supplementary Figure 17 - Gene regulons

Regulons identified using SCENIC showing **a** disease-specific differences in Intermediate1 microglia and/or perivascular macrophages (PVMs) or **b** multiple system atrophy (MSA) specific regulons in Intermediate1 microglia. **c** Gene ontology analysis on MSA-specific regulons. CTRL: healthy controls; PD: Parkinson's disease.

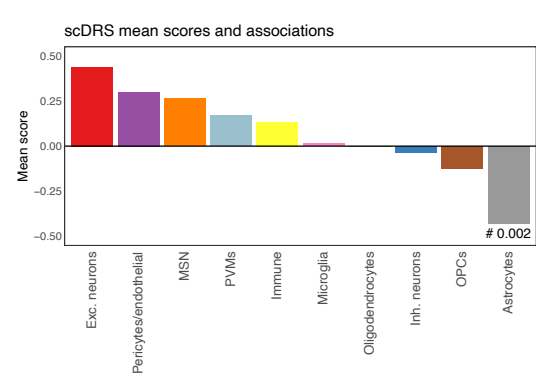
a



b

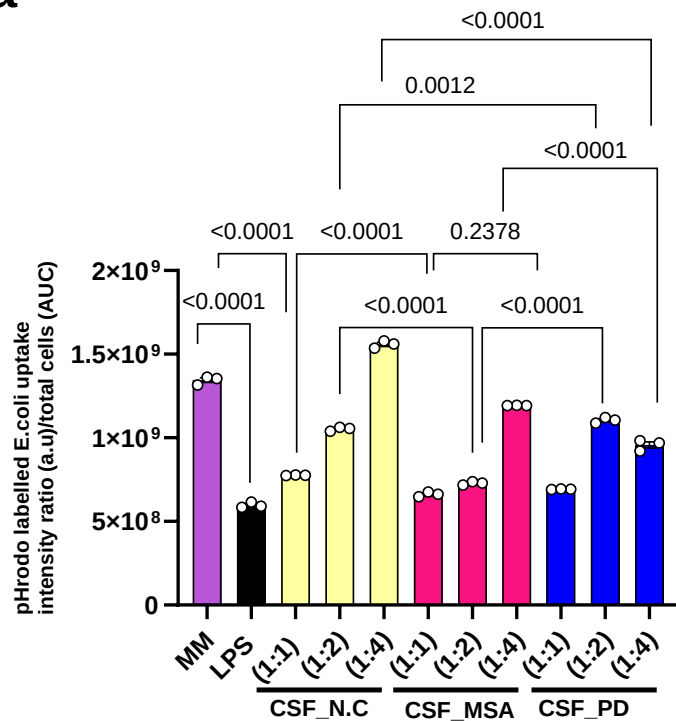
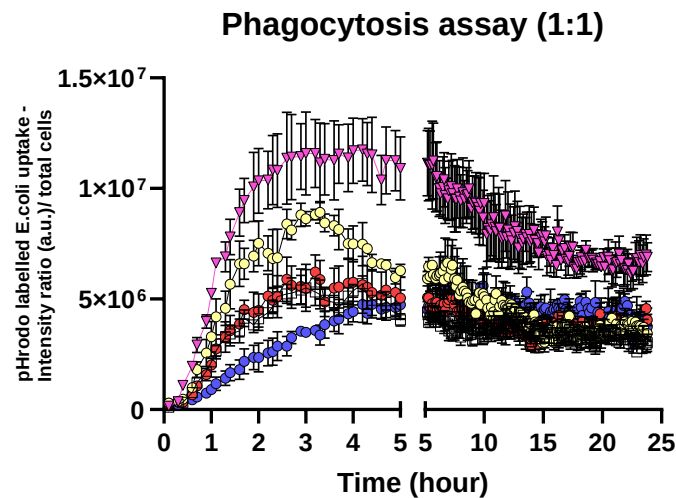
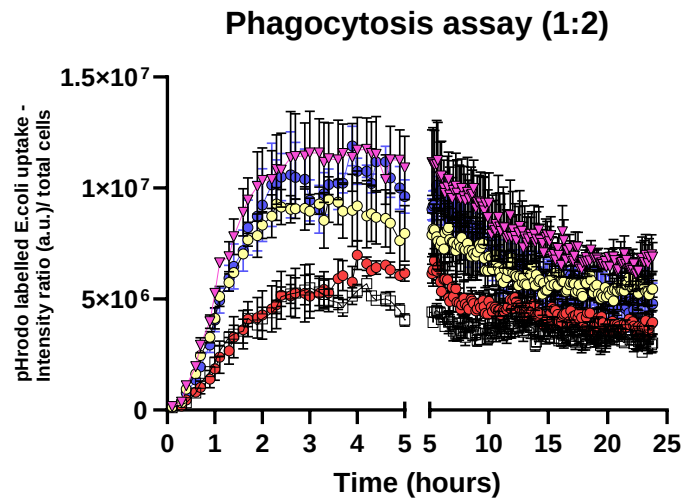
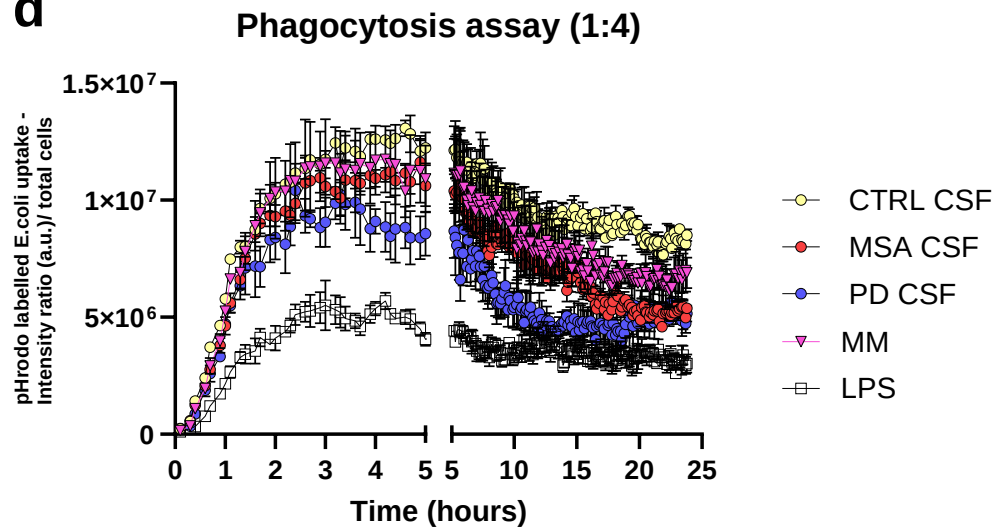


c



Supplementary Figure 18 - Single-cell disease-relevance GWAS scoring

a GWAS single-cell disease-relevance scoring based on scDRS shows nominally significant cell-type disease association for microglia, (MIC), perivascular macrophages (PVMs), immune cells and pericytes (Monte Carlo significance test) using GWAS data on Parkinson's disease from Nalls et al. **b** This association was nominally significant for the intermediate-2 microglia. **c** Similarly, GWAS associations using data on multiple system atrophy from Chia et al. Asterisks indicate significance level of cell-type disease associations, sharps indicate significance level of heterogeneity in association with disease across individual cells within a given cell type. OPCs: oligodendrocyte precursor cells; MSN: medium spiny neurons; exc. neurons: excitatory neurons; inh. neurons: inhibitory neurons.

a**b****c****d**

Supplementary Figure 19 - Phagocytosis assay on human derived microglia iPSC

a Total phagocytosis for all dilutions in experiment 2 quantified as the area under the time-course curve (AUC) across 166 time-points. Each dot represents one technical replicate trajectory; bars indicate mean $\pm$ SEM. Statistical analysis was performed on the full time-course data using two-way ANOVA (time x condition) followed by Tukey's multiple comparison test. **b** Phagocytosis over time in experiment 2 when microglia cells were incubated with undiluted cerebrospinal fluid (CSF; 1:1). Data are shown as mean $\pm$ SEM of three technical replicate trajectories. **c** Phagocytosis over time when microglia cells were incubated with CSF diluted 1:2 in microglia medium (MM). Data are shown as mean $\pm$ SEM of three technical replicate trajectories. **d** Phagocytosis over time when microglia cells were incubated with CSF diluted 1:4 in MM. Data are shown as mean $\pm$ SEM of three technical replicate trajectories. LPS: lipopolysaccharides. CTRL: healthy controls; MSA: multiple system atrophy; PD: Parkinson's disease.