Single-Nucleus Accessible Chromatin Profiling Highlights Distinct Astrocyte Signatures in Progressive Supranuclear Palsy and Corticobasal Degeneration

Nils Briel, Viktoria C Ruf, Katrin Pratsch, Sigrun Roeber, Jeannine Widmann, Janina Mielke, Mario M Dorostkar, Otto Windl, Thomas Arzberger, Jochen Herms*, Felix L Struebing*

* These authors contributed equally.

– Supplementary figures –
**Suppl. Fig.01 Cluster validation and cluster purity**

**A** Lineplot indicating the *gap statistics* for different $k$-values between 1 and 15 that were drawn from a subset of all barcodes and bin-matrix, *post hoc* confirming a reliable cluster differentiation by the previous graph-based clustering approach. Points and error bars correspond to the deviation of intra-cluster variation at $k = 1,…,15$ from a randomly distributed reference data set and their standard deviations, respectively. The dendrogram on the right visualizes the determined cluster similarity using centroid hierarchical clustering.

**B-D** Differentiation of cluster assigned cell type identities in UMAP, stratified by group entity (columns) and n=500,1000, 5000 cells (B,C,D). The color code and polygon fill both indicate cell types and clusters. With increasing cell numbers the observed cluster purity decreases slightly.

**F** Projection of case identity onto the UMAP embedding of all barcodes before (left) and after batch effect correction using *harmony*.

**Abbreviations:** Ast, astrocytes; DLN, deep-layer neurons; Exc., excitatory; Inh. N., Inhibitory neurons; Mic, microglia; Neu., neurons; ULN, upper-layer neurons; Oli Oligodendrocytes; OPC, oligodendrocytic precursor cells.
Suppl. Fig. 02 Epidemiological and sequencing metadata

A Projections of technical metadata onto the UMAP embedding of high quality-filtered barcodes indicating the respective variable as color code. Shown are read depth, fraction of reads in peaks ('FriP') score, duplicate likelihood, promoter and enhancer ratios. A pie chart displays group-wise contributions to the entire cell pool. As reference, the first an second panel include cluster identifiers (ID) and cell type assignments, respectively.

B Boxplot comparisons of nuclei sequencing read depth across included cases. Outliers are depicted as black dots. The hinges of each box correspond to the 25th and 75th percentiles with medians drawn as black bar. The 1.5-times inter-quartile ranges are shown as black whiskers. Red dots show means of the distributions.

C Boxplot comparisons of patient-specific and disease-relevant parameters, such as age at death, disease duration and post mortem interval. The plot structure equals to B. Color coding indicates the neuropathological diagnose, while asterisks denote the degree of significance with ‘ns’ = ‘not significant’, *p<.05, and **p<.01. Analysis of variance (Anova) results are shown at the bottom of each triple comparison.

D Correlation matrix heatmap of epidemiological and technical parameters between average-aggregated barcodes of single human cases. Pearson’s R is displayed by the color shading and labels. Paired correlations that do not comply with a p<.05 are shown as empty boxes.

Abbreviations: PMI, post mortem interval; TSS, transcription starting site; UMAP, uniform manifold approximation and projection.
Suppl.Fig.03 Cell type identification by marker gene activity

Projections of accessibility at canonical neuronal (A-E) and glial (F-H) marker genes onto the UMAP embedding of barcodes indicating the respective variable as color code from low (blue) to high (yellow). Depicted are the three most distinguishing marker genes of each cell type or sub-type. GA scores are scaled to each gene's range.

**Abbreviations**: DLN, deep-layer neurons; Exc., excitatory; Inh. Inhibitory neurons; Neu., neurons; Oligodendro, oligodendrocytes; ULN, upper-layer neurons.
case
PSP1
CBD1
PSP2
CBD2
CBD3
CBD4
PSP3
PSP4
C1
C2
C3
C4

R = −0.03, p = 0.92

H = −0.14, p = 0.67

H = −0.18, p = 0.56

H = −0.15, p = 0.63

H = −0.32, p = 0.31

H = −0.37, p = 0.24

H = 0.65, p = 0.017

H = −0.031, p = 0.92

H = 0.2, p = 0.5

H = −0.31, p = 0.33

H = −0.28, p = 0.38

# cells: Ast

# cells: Exc. DLN

# cells: Exc. N. NEFM|BCL11B

# cells: Exc. ULN

# cells: Inh. N. SST|PVALB

# cells: Inh. N. VIP

# cells: Mic

# cells: Oli #1 MOBP

# cells: Oli #2

# cells: Oli #3

# cells: OPC

Suppl.Fig.04
Suppl.Fig.04 Cross-case correlation of cell type frequencies and *Thal* phases

Stratified by previously defined cell types and subpopulations, each scatter plot depicts the relation between *Thal*-phase (Aβ+ plaque distribution) on the x-axis and the relative cell type frequency across cases in a group-agnostic way. A regression line is drawn in black, while the confidence interval is shown in grey. *Pearson’s* R and p-values are depicted in the upper right corners. Color code indicates case identity.
**Suppl.Fig.05 Gene ontology (GO) of the major brain cell types**

**A-C** GO tile maps of the top 100 results of each category Molecular Function (MF, A), Biological Process (BP, B) and Cellular Compartment (CC, C). The results were ranked by their *Benjamini Hochberg*-corrected $p$-values of the binomial enrichment test in *rGREAT*. Every column corresponds to one of these terms and every row to the cell type under investigation. The color fill indicates the binomial test fold enrichment, while the colored outlines show the negative decadic logarithm of the adjusted $p$-values (Adj. $P$).

**Abbreviations:** Adj. $P$, adjusted $p$-value; Ast, astrocytes; binom, binomial; DLN, deep-layer neurons; Exc., excitatory; Inh. N., Inhibitory neurons; Mic, microglia; Neu., neurons; ULN, upper-layer neurons; Oli, oligodendrocytes; OPC, oligodendrocytic precursor cells.
Suppl. Fig.06 TF profile in the tauopathy data set

A Motif enrichment heatmap of a TF subset that characterizes brain cell types. Every column corresponds to a single barcode and every row to a TF motif, while the color gradient indicates the extent of TF motif enrichment (TFME) from negative (black) through intermediate (purple) to high (yellow) values. Rows were clustered hierarchically (Manhattan distance, Ward-D2 method) and results indicated as dendrogram on the left. The column order was fixed while the colored bars at the top inform about the case, disease diagnosis and cell type. Gene names comply with the Ensembl identifiers.

B Heatmap of correlation testing between cell type-wise average TFME values. Spearman correlation coefficient is displayed by the color shading and labels. Paired correlations that do not comply with a p<.05 are shown as empty boxes.

Abbreviations: Ast, astrocytes; Corr, correlation coefficient; DLN, deep-layer neurons; Exc., excitatory; Inh. N., Inhibitory neurons; Mic, microglia; Neu., neurons; ULN, upper-layer neurons; Oli Oligodendrocytes; OPC, oligodendrocytic precursor cells.
Suppl. Fig. 07

A

B

Difference of GA-medians:  
PSP - Ctrl  
Total = 93 variables

Difference of GA-medians:  
CBD - Ctrl  
Total = 94 variables
Suppl.Fig.07 Accessibility alterations at tauopathy-associated genes

A Volcano plot of differentially accessible tauopathy-associated genes in the PSP frontal cortex resolved by cell type assignments. Resulting from pair-wise Wilcoxon rank sum tests, the 93 significant tauopathy gene - cell type pairs (GCPs) are depicted in the plot. The x-axis differentiates the GA difference between the PSP GCPs and the reference GCPs of the Ctrl group. The y-axis displays the negative decadic logarithm of the FDR-corrected p-values. Color codes indicate whether GCPs comply with an adj. p-value <= .05 and an absolute difference of GA medians >=1 (red) or not (blue).

B Volcano plot of differentially accessible tauopathy-associated genes in the CBD frontal cortex resolved by cell type assignments. Resulting from pair-wise Wilcoxon rank sum tests, the 94 significant GCPs are distributed in the plot. The general plot structure corresponds to C.

Abbreviations: Adj. P, adjusted p-value; Ast, astrocytes; FDR, false discovery rate; GA, gene accessibility.
Suppl. Fig. 08

A

Molecular Function

- signal transducer activity
- receptor activity
- signaling receptor activity
- transmembrane receptor activity
- transmembrane signaling receptor activity
- G-protein coupled receptor activity
- olfactory receptor activity

-log(Adj. P)

Binom. fold enrichment

B

TFME: EMX2

Pseudotime

Entity

Ctrl  PSP
Suppl. Fig.08 GO of the astrocyte subclusters

A GO analysis results in the astrocytic subcluster 3 after k-means clustering of the PSP, CBD, and Ctrl astrocytes-comprising UMAP. Bar plots of the top results of each category MF (left), BP (mid) and CC (right) were ranked by their Benjamini Hochberg-corrected p-values of the binomial enrichment test in rGREAT. Every row corresponds to one of the GO terms and the color fill indicates the binomial test fold enrichment, while the bar length represents the negative decadic logarithm of the adjusted p-values (Adj. P).

B Subset of PSP and Ctrl astrocytes re-embedded in UMAP. Color code indicates EMX2 TF motif enrichment (TFME) (left), inferred pseudotime (mid), or group entity (right). The black line indicates the pseudotemporal trajectory from the ‘root’ (white label) toward the ‘end’ (grey label) cell. A clear transition trajectory could not be inferred.

Abbreviations: Adj. P, adjusted p-value; Ast, astrocytes; binom, binomial.
A) transcription factor binding
- cellular response to cadmium ion
- regulation of hematopoietic stem cell differentiation
- protein heterodimerization activity
- cellular senescence
- transcription initiation from RNA polymerase II promoter
- cytokine-mediated signaling pathway
- cellular response to lipopolysaccharide
- inflammatory response
- regulation of hematopoietic stem cell differentiation
- blood coagulation
- negative regulation of gene expression
- regulation of alternative mRNA splicing, via spliceosome
- protein self-association

B) cytokine-mediated signaling pathway
- cellular response to lipopolysaccharide
- inflammatory response
- regulation of hematopoietic stem cell differentiation
- positive regulation of NF-kappaB transcription factor activity
- negative regulation of gene expression
- inflammatory response
- Fc-epsilon receptor signaling pathway
- cytokine-mediated signaling pathway

C) chromatin binding
- cellular response to cadmium ion
- regulation of hematopoietic stem cell differentiation
- blood coagulation
- negative regulation of gene expression
- protein heterodimerization activity
- regulation of alternative mRNA splicing, via spliceosome

D) cytokine-mediated signaling pathway
- cellular response to lipopolysaccharide
- inflammatory response
- regulation of hematopoietic stem cell differentiation
- positive regulation of NF-kappaB transcription factor activity
- negative regulation of gene expression
- inflammatory response
- Fc-epsilon receptor signaling pathway
- cytokine-mediated signaling pathway

E) Combined Terms Graph

F) p = 2e-05
Suppl. Fig.09 Biological pathways indicated by altered TFs in tauopathy brains

A GO enrichment analysis of TFs that exhibited significant TFME deviations in PSP astrocytes. Only the top 25 enrichment results with an adjusted p-value <.05 are depicted. Rows correspond to the GO terms comprising MF, BP, and CC, while TF *Ensembl* IDs are given on the x-axis. The color shading indicates the enrichment score from negative (blue) through zero (grey) to high (red) values.

B GO enrichment analysis of TFs that exhibited significant TFME deviations in CBD astrocytes. The plot structure equals to A.

C Combined heatmap-upset plot of the top 10 GO terms enriched in the set of differentially active TFs in PSP astrocytes. The lower matrix shows the co-enriched GO term logic. The resulting intersection of GO terms involves those TFs that are highlighted in the heatmap above. Whether these TFs exhibit positive or negative TFME values changes is indicated by the color code.

D Combined heatmap-upset plot of the top 10 GO terms enriched in the set of differentially active TFs in CBD astrocytes. The plot structure equals to C.

E Bubble-connection graph depicting the top 10 co-enriched GO terms of significant TF alterations in PSP and CBD. The yellow dots represent specific GO terms, while their size equal to the number of term-assigned genes (# genes) that are found in the queried data sets. These genes are displayed as red (co-downregulated), green (co-upregulated), or grey dots (conflicting direction).

F The extent of TFME alterations differs between astrocytes of PSP and CBD origin. Y-axes depict TFME median value deviations from Ctrls’ medians in astrocytes. The dots represent single TFMs, while violin convexities indicate the distribution over TFME deviations. For a more intuitive comprehension a horizontal line marks the zero line. The degree of statistical significance is given for the CBD vs. PSP comparison (*Wilcoxon* rank-sum test).
Suppl. Fig. 10 Loss of immaturity markers and acquisition of a reactive inflammatory state

A Heatmap displaying the pseudotime changes in TFME of the TFs belonging to the two major, inverse running clusters as determined by k-means clustering in tradeSeq. Results relate to Ctrl/CBD astrocytes and their pseudotemporal transition only. TFME courses (rows) were clustered hierarchically (Euclidean distance, complete method) and results indicated as dendrogram on the left.

B Heatmap displaying the pseudotime changes in GA of tauopathy-associated as well as significantly altered genes in the start-vs.-end and association test in tradeSeq. Candidates had to comply fulfill the criterion of p<.05 (Wald-statistic, Bonferroni correction). The results relate to Ctrl/CBD astrocytes and their pseudotemporal transition only. GA courses (rows) were clustered hierarchically (Euclidean distance, complete method) and results indicated as dendrogram on the left.
Suppl. Fig. 11 Analysis of discretized TFME changes across pseudotime in CBD astrocytes

A Projections of pseudotime steps (left) and the MA0099.2_FOS::JUN enrichment (right) onto the UMAP embedding of Ctrl/CBD astrocytes indicating the respective variable as color code.

B-E Boxplots showing the TFME values of selected TFMs in Ctrl/CBD astrocytes over 5 discrete pseudotime steps. The color code emphasizes the assigned pseudotime bin. Statistical comparison was conducted between each time step > 1 and the first one (Wilcoxon rank-sum test) and results expressed as asterisks where *p<.05, **p<.01, ***p<.001 and ****p<.0001. On the right, the position weight matrix is displayed via motif sequence logos. The information content is depicted as bits for every base position of the motif sequence.
Suppl.Fig.12 XGB model performance and most discriminative features for predicting the astrocyte disease entity

A The sensitivity vs. counter-probability of specificity is drawn as receiver-operator curves (ROC) of the trained XGB model performance on an unseen astrocytic TF matrix subset. Differently colored curves display the three comparisons between Ctrl, PSP, and CBD assigned astrocytes.

B Feature matrix boxplots depicting the TFME distribution of the 5 most important features in the XGB model representation for each group entity (top: Ctrl, middle: CBD, lower: PSP) as extracted from Lime. Outliers are depicted as blue circles. The hinges of each box correspond to the 25th and 75th percentiles while medians drawn as black dots. The 1.5-times inter-quartile ranges are shown as black whiskers. The entire range is scaled to values from -0.4 to 0.8 for all features.

Abbreviations: col, color.
Suppl.Fig.13 GO analysis of TA-associated regulons

Clustered depiction of GO term enrichment in all genes regulated by the top 50 TA-associated regulons. The x-axis differentiates the GO term fold enrichment, the negative decadic logarithm of respective p-values is indicated as color gradient, and the gene set size is expressed as point size.

**Abbreviations**: # genes, number of genes.
Suppl. Fig. 14 Assessing the involvement of protein degradation pathways on a system level

A-C Gene accessibility (GA) heatmaps of genes associated with three major protein homeostasis pathways (UPS, A; CMA, B; UPR, C) in astrocytes. Every column corresponds to a single nucleus and every row to a specific gene, while the color gradient indicates the extent of GA from negative (blue) through intermediate (black) to high (red) values. Rows were clustered hierarchically (Manhattan distance, Ward-D2 method) and results indicated as dendrogram on the left. The column order was fixed while the colored bars at the top inform about the cell type and disease diagnosis. Gene names comply with the Ensembl identifiers.

D Heatmap displaying the pseudotime changes in GA of tauopathy and degradation pathways-associated, significantly altered genes in the group-wise comparisons (Fig. 6H, main manuscript). The results relate to Ctrl/CBD astrocytes and their pseudotemporal transition exclusively. Gene assignments to these pathways are indicated by colored bars on the left.

E Bubble-connection graph depicting the top 3 co-enriched and top 2 exclusively enriched GO terms of TF signatures in PSP and CBD. The yellow dots represent specific GO terms, while their size equal to the number of term-assigned genes (# genes) that are found in the queried data sets. These genes are displayed as red (co-downregulated), green (co-upregulated), or grey dots (conflicting direction). ‘A-only terms’ correspond to PSP-related and ‘B-only terms’ to CBD-related ones.

Abbreviations: CMA, chaperon-mediated autophagy; UPS, ubiquitin-proteasome-system; UPR, unfolded-protein-response.
Suppl. Fig.15 Re-analysis of the dataset with exclusion of #CBD3 – part 1: dataset characterization

A Plot is analogous to Fig.2a. Projections of cluster-cell type assignments and metadata onto the UMAP embedding. Color coding and labels indicate the cell type or sub-cell type identity where applicable.

B Plot is analogous to Fig.2c. Boxplots of relative cell type frequencies show reductions in excitatory neurons in PSP (left) and in all neuronal populations in CBD. Higher astrocyte and oligodendrocyte frequencies can be detected in the remaining CBD samples, when compared to the Ctrl’s mean (vertical dashed line). Total numbers of cells (# of cells) are indicated as bar plots on the right.

C Plot is analogous to Fig.3c. Protein homeostasis-related genes across all cell types differentiated by group entity and degradation pathway show reduced system-level GA in Ast, most neuronal, and Oli populations, while Mic exhibit marked CMA and UPS inductions in PSP. Color coding shows aggregated mean scores of accessibility values at genes that were altered significantly and associated with either the CMA (top), UPR (middle), or UPS degradation systems (bottom). P-values are given for each group vs. Ctrl comparison (2-tailed Welch t-test, if p <= .05).

D Plot is analogous to Fig.3d. Heatmap of genetic risk variant enrichment results in tauopathy cortices resolved by cell type assignments (x-axis) and GWAS data set (y-axis) highlight Ast, which exhibit strong enrichment. Color code indicates z-scores and text inserts depict the uncorrected p-value (italic) as well as the BH-corrected p-values (bold, Wilcoxon rank-sum test).

Abbreviations: CMA, chaperon-mediated autophagy; Exc. DLN, excitatory deep-layer neurons; Exc. ULN, excitatory upper-layer neurons; Inh. N., inhibitory neurons; Mic, microglia; Oli, oligodendrocytes; OPC, oligodendrocytic precursor cells; UPS, ubiquitin-proteasome-system; UPR, unfolded-protein-response.
Suppl. Fig. 16

A) PSP, CBD, Ctrl, Clusters

B) END, ROOT, Ctrl, CBD

C) MA0099.2_FOS::JUN

D) MA0799.1_RFX4, MA0612.1_EMX1, MA0701.1_LHX9, MA0886.1_EMX2, MA0894.1_HESX1, MA0630.1_SHOX, MA0841.1_EMX2, MA0894.1_HESX1, MA0652.1_IRF8, MA0502.1_IRF1, MA0653.1_IRF9, MA0722.1_IRF7, MA0051.1_IRF2, MA0099.2_FOS::JUN, MA0018.2_CREB1, MA0839.1_CREB3L1, MA0839.1_CREB3, MA0692.1_TFEB, MA0462.1_BATF::JUN, MA0477.1_FOSL1, MA0478.1_FOSL2, MA0490.1_JUNB, MA0491.1_JUND, MA0841.1_NFE2, MA0476.1_FOS, MA0489.1_JUN(var.2), MA0655.1_JDP2

TFME

AP-1 complex | IER-related | homeostasis | immune response | early stages

Sign. Start-vs.-End

log2(TFME:EMX2) -1.0 -0.5 0.0 0.5 1.0

Pseudotime 0 5 10 15

Group Entity Ctrl CBD

Clusters 0.25 0.30 0.35 0.40 0.45

Pseudotime 0 5 10 15
Suppl.Fig.16 Re-analysis of the dataset with exclusion of #CBD3 – part 2: Pseudotime trajectory of CBD astrocytes

A Plot is analogous to Fig.4a. All PSP-, CBD-, and Ctrl-derived astrocytes (excluding #CBD3) re-embedded in UMAP, stratified by group entity (first, second, third panel), and depicted after k-means clustering in a merged UMAP (fourth panel). One cluster (#3) remains specific for CBD astrocytes. Color code indicates group entity or cluster assignments in the first three or the fourth panel, respectively.

B Plot is analogous to Fig.4b-d. Exclusively CBD- and Ctrl-derived astrocytes (excluding #CBD3) re-embedded in UMAP. A pseudotime trajectory leads from a non-specific Ast pool towards a CBD-enriched population. Color code indicates log2(EMX2 TFME) (left), pseudotime (mid, dimensionless), or group entity (right). The black line indicates the pseudotemporal trajectory from the root towards the end cell.

C Plot is analogous to Fig.4e-f. Generative additive model non-linear fits of TFME values over pseudotime of the FOS-JUN (upper) or NFE2 (lower) motifs indicate parallel increments during the astrocytic transition towards a CBD-state.

D Plot is analogous to Fig.4g. Pseudotime heatmap displaying the TFME values of significantly altered TFMs in the start-vs.-end comparison (Wald-testing, BH-adjusted p<.05), as well as markers of early astrocytic development or immune regulation. TFME of astrocytic early-stage TFs is gradually decreasing, while immunologically relevant and AP-1 complex-related TFs gain in motif enrichment.

**Abbreviations:** IER, immediate-early response; TFME, transcription factor motif enrichment; Sign, significant.
Suppl.Fig.17

A

Overall Accuracy 0.817
Balanced Accuracy 0.83
F1 0.784
Sensitivity 0.766
Specificity 0.893
Pos Pred Value 0.81
Neg Pred Value 0.9
Kappa 0.681

B

|          | PSP | Ctrl | CBD | Total |
|----------|-----|------|-----|-------|
| Case: AGGCGTCTCGACTATG.1 | Label: CBD | Probability: 1.0e+00 | Explanation Fit: 0.61 |
| Probability: 1.0e+00 | Explanation Fit: 0.61 |

C

Case: AGGCGTCTCGACTATG.1
Label: CBD
Probability: 1.0e+00
Explanation Fit: 0.61

Case: TGTGCGCAGGCAATG.1
Label: PSP
Probability: 1.0e+00
Explanation Fit: 0.42

Case: AGTTTGGCACGGCCAT.1
Label: Ctrl
Probability: 9.9e-01
Explanation Fit: 0.25

-7.54e-02 < MA0059.1_MAX::MYC <= -4.95e-02
-0.11692 < MA0025.1_NFIL3 <= -0.08705
0.02060 < MA0014.2_PAX5 <= 0.02865
-0.1240 < MA0477.1_FOSL1 <= -0.1063
0.02060 < MA0014.2_PAX5 <= 0.02865
-0.11692 < MA0025.1_NFIL3 <= -0.08705
0.02060 < MA0014.2_PAX5 <= 0.02865
-0.1240 < MA0477.1_FOSL1 <= -0.1063
0.02060 < MA0014.2_PAX5 <= 0.02865
-0.11692 < MA0025.1_NFIL3 <= -0.08705
Suppl.Fig.17 Re-analysis of the dataset with exclusion of #CBD3 – part 3: TF networks associated with an astrocytic tauopathy state

A Plot is analogous to Fig.5b. Evaluation parameters of classification performance of the trained XGB model on the 20% test set-split of astrocytic nuclei (excluding #CBD3). Overall, more than 81% of predictions were correct (overall accuracy) and the model performs substantially with a Cohen kappa of 68.1%.

B Plot is analogous to Fig.5a. Confusion matrix displaying the intersections of the XGB model's predictions (rows) and the actual labels (columns). Each square contains the percentual proportion (large digits) and total numbers (small digits) of test set samples (=20% astrocytic nuclei excluding #CBD3) with the assigned prediction-label-relation. The sums of each row or column are depicted in the rightmost column or undermost row, respectively. The total sample number (i.e., nuclei of the 20% test set-split) is shown in the downright corner.

C Plot is analogous to Fig.5d-f. *Lime* feature importance bar diagrams of the most certainly correctly classified barcodes of each group entity. The bar direction and bar color indicate the feature weights (~importance) assigned to the TFM, which are given as y-axis breaks. Note, feature weight was assigned to specific TFME value ranges. Each panel is complemented by the group entity label, the models’ calculated probability, and the explanatory model's fit value.

*Abbreviations*: Neg Pred Value: negative predictive value, Pos Pred Value: positive predictive value; TFM(E), transcription factor motif (enrichment); XGB, extreme gradient boosting tree.
Suppl. Fig. 18
Suppl. Fig. 18 Re-analysis of the dataset with exclusion of #CBD3 – part 4: Astrocytic tauopathy signatures

**A** Plot is analogous to Fig. 6a. Upset plot illustrating TFs useful in distinguishing PSP/CBD from Ctrl astrocytes (when excluding #CBD3) that resulted from (i) group-wise TFME comparisons in the snATAC-seq data set (‘TFME’ comparison’), (ii) the pseudotime trajectory analysis in the snATAC-seq data set (‘Trajectory analysis’), (iii) the TA-related regulon activity profile in the bulkRNA-seq data set (‘RTN: TA-RAP’), and (iv) interpreting the XGB classification model (‘ML model: tauopathy’). Set sizes are indicated as blue bars, while the intersection logic is shown as vertical lines and dots. Column heights depict the extent of intersection for the given sets. The first three intersections were assigned a hierarchy of importance in defining the primary tauopathy context.

**B** Plot is analogous to Fig. 6b. Triangle plot indicating significance, absolute extent, and direction of TFME changes in **p** Tau signature TFs in tauopathy-assigned astrocytes. The triangle tips point towards the direction of change while the size represents the absolute difference from the TFME reference in Ctrls. Fill shading displays the negative decadic logarithm of the BH-corrected p-values from pairwise Wilcoxon rank-sum tests. Empty coordinates inform about non-significant comparisons in the test set.

**C**. Plot is analogous to Fig. 6c. Upset plot to identify TFs useful in differentiating CBD from PSP astrocytes (when excluding #CBD3). The single sets resulted (i) from the most important TFs for PSP or CBD prediction according to the XGB model explainer (‘ML model: PSP’, ‘ML model: CBD’) and (ii) from pairwise statistical TFME comparisons between PSP and CBD astrocytes (‘TFME comparisons’). The general plot structure equates to A.

**D-E** Plots are analogous to Fig. 6d-e. Triangle plot indicating significance, absolute extent, and direction of TFME changes in **PSP** (D), and **CBD** (E) signature TFs in tauopathy-assigned astrocytes (when excluding #CBD3). The general plot structure equates to B.

**Abbreviations**: Abs.diff., absolute difference; AP, astrocytic plaque; Dis.Ent., disease entity; FDR, false discovery rate; ML, machine learning; TA, tufted astrocyte; TFM(E), transcription factor motif (enrichment).
Suppl. Fig. 19

A  Chromosome 11  CTSD

B  Chromosome 10  MAP3K8

PSP Astrocytes

Ctrl Astrocytes

Co-accessibility

Peaks

protein_coding

lincRNAs

Gene Model
Suppl. Fig. 19 Co-accessibility at JUNB and TFEB target genes in CBD & PSP astrocytes.

**A-C Co-accessibility plots at genomic loci including genes that are correlated with TFEB and/or JUNB motif enrichment (A CTSD, B MAP3K8) in astrocytes. Chromosomal location is depicted at the top of each panel, followed by Cicero’s co-accessibility links in PSP/CBD (upper) and Ctrl (middle) astrocytes. The y-axes correspond to the extent of co-accessibility, while the value 0.1 is chosen as cut-off to reduce depiction of technical or biological noise. Annotations for this genomic frame (gene location +/- 10e4 bp) including protein coding gene isoforms as well as lincRNAs are given in the bottom panel. The distribution of summed ATAC peaks are given beneath each co-accessibility plot.**

**Abbreviations:** lincRNAs, long intervening/intergenic noncoding RNAs.