
Supplementary information

Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease

In the format provided by the
authors and unedited

Table of Contents

<i>Supplementary Results</i>	2
iNMF analysis	2
Cell subpopulation annotation	2
Cell subpopulation differential analysis	3
Cell-type specific cCREs	4
Astrocyte FOSL2, microglia SPI1, and oligodendrocyte NRF1 dysregulation	4
Trajectory analysis to study NRF1 and SREBF1 in oligodendrocytes	5
Trajectory analysis to study SPI1 and ETV5 and the DAM gene signature in human microglia	5
Trajectory analysis to study CTCF and FOSL2 in astrocytes	7
Cell-type-specific cis-regulation at AD GWAS loci	7
Meta-analysis of single-cell co-expression networks using scWGNCA	8
<i>Supplementary Discussion</i>	9
<i>Supplementary Methods</i>	10
Immunohistochemistry	10
Processing sequencing reads from droplet-based single-nucleus RNA and ATAC -seq	11
Quality control for filtering of low-quality nuclei.....	12
Dimensionality reduction and initial clustering	13
Inference of gene activity from chromatin accessibility	13
Constructing multi-omic glial cell type pseudotime trajectories	14
Differential feature analysis	15
Modeling the pseudotemporal dynamics of gene expression and chromatin accessibility using a recurrent variational autoencoder (RVAE) neural network	16
Cross-species integrated analysis of microglia signatures	17
Estimating AD GWAS enrichment in the microglia pseudotime trajectory	18
Joint analysis of snRNA-seq with Mathys et al snRNA-seq	18
Processing sequencing reads from Bulk RNA-seq	19
<i>Supplementary Figures</i>	20
<i>References</i>	42

Supplementary Note

Supplementary Results

iNMF analysis

Integrative non-negative matrix factorization (iNMF) compresses gene expression data into a series of interpretable matrix factors that can be thought of as *metagenes*, and we observed that each *metagene* has high specificity to a cell-type or subset of cell-types (Extended Data Fig. 1). Additionally, we identified canonical cell-type markers that contribute to iNMF *metagene* loadings, such as *MBP* in iNMF 4 (Extended Data Fig. 1).

Cell subpopulation annotation

We annotated our excitatory and inhibitory neuronal subpopulations from snRNA-seq based on previously identified marker genes^{1,2}—EX1 as *LAMP5*⁺/*LINC00507*⁺, EX2 as *RORB*⁺, EX3 as *RORB*⁺/*THEMIS*⁺, EX4 as *FEZF2*⁺, EX5 as *NRGN*⁺, INH1 as *VIP*⁺, INH2 as *PVALB*⁺, INH3 as *SST*⁺, and INH4 as *LAMP5*⁺ (Supplementary Fig. 7). Similarly, we sought to annotate our glial subpopulations, separating our astrocyte clusters based on *GFAP* expression, which has been shown to increase in disease³ (ASC1 as *GFAP*^{medium}/*WIFI*⁺/*ADAMTS17*⁺, ASC2 as *GFAP*^{high}/*TNC*⁺, ASC3 as *GFAP*^{high}/*CHI3L*⁺, ASC4 as *GFAP*^{low}/*WIFI*⁺/*ADAMTS17*⁺), and our microglia clusters based on the expression of homeostatic and activation markers (MG1 as *SPP1*^{high}/*CD163*⁺, MG2 as *CX3CR1*^{high}, MG3 as *ETSI*⁺; Supplementary Fig. 6).

Due to the large number of oligodendrocyte clusters, we used hierarchical clustering to group our snRNA-seq and snATAC-seq oligodendrocyte clusters into two overarching groups (Methods, Fig. 2a-b). Based on the expression of *ITPR2*, *OPALIN*, and *CNP*, we labeled the

group comprised of snRNA-seq clusters ODC3, ODC5, ODC8, ODC10, ODC12, and ODC13 as “intermediate” and the other group of clusters as “mature” (Supplementary Fig. 6). *ITPR2* expression in oligodendrocytes has been shown to be a marker of newly differentiated oligodendrocytes, while *OPALIN* and *CNP* are genes involved in myelin formation⁴. Interestingly, we found specific expression of *CD74*, a marker of immune oligodendrocytes⁵, in ODC13 (Supplementary Fig. 6). We similarly labeled our snATAC-seq oligodendrocyte clusters as “intermediate” and “mature”; however, we were unable to identify an immune oligodendrocyte cluster equivalent to the snRNA-seq cluster ODC13 (Fig. 2c, Supplementary Fig. 6).

Cell subpopulation differential analysis

Downregulated gene-proximal differentially accessible regions (DARs) in microglia clusters MG.a and MG.b, which we identified as upregulated in disease, were associated with small nucleolar RNA (snoRNA) catabolism and protein K11-linked deubiquitination, respectively (Supplementary Fig. 7). Downregulated distal DARs in mature oligodendrocyte cluster ODC.d were related to oligodendrocyte apoptosis, while in another mature oligodendrocyte cluster ODC.h, they were related to positive regulation of myelination (Supplementary Fig. 8). In addition, genes upregulated in ODC8 and ODC2 were related to protein folding, whereas genes upregulated in ODC5 and ODC10 were associated with cell adhesion, suggesting differential responses to disease by oligodendrocyte subpopulations (Supplementary Fig. 9). Moreover, we observed that many top differentially expressed genes were dependent on cell-type and cluster (Supplementary Fig. 9).

Cell-type specific cCREs

In microglia we identified an intronic candidate cis-regulatory element (cCRE) linked to *CD44* at a distance of 12.37 kb away from the promoter, while in oligodendrocyte progenitors we identified a distal cCRE linked to *CD44* at a distance of 308.55 kilobases from the promoter. We also identified biological processes regulated by each of the cCRE NMF modules by performing GO term enrichment in the set of target genes for each module (Fig. 3g). We found that regulation of amyloid fibril formation was enriched in NMF module 27, which is associated with oligodendrocytes, and contains cCREs linked to *APOE* and *CLU*, suggesting that this module may be involved in the regulation of A β neurotoxicity. Furthermore, we found that positive regulation of autophagosome maturation was enriched in NMF module 25, which is most associated with late-stage AD nuclei in the *GFAP*^{high} ASC.b cluster based on its NMF coefficient, indicating that cCREs may be mediating astrocyte autophagy in late-stage AD.

Astrocyte FOSL2, microglia SPI1, and oligodendrocyte NRF1 dysregulation

Fos related antigen 2 (FOSL2), a member of the FOS family, interacts with members of the JUN family to form AP-1, whose role in neurodegeneration is still largely unclear⁶. We found that FOSL2 had significantly increased motif variability in late-stage AD across multiple snATAC-seq astrocyte clusters (Supplementary Fig. 10). TF footprinting corroborated disease associated FOSL2 motif accessibility in astrocytes (Supplementary Fig. 10). Additionally, we found that gene expression of FOSL2's downstream targets was significantly increased in late-stage AD in our snRNA-seq astrocyte clusters, with the notable exception of upregulated cluster ASC3, suggesting cell subpopulation-specific regulatory effects of FOSL2 in disease (Supplementary Fig. 10).

We also found SPI1 has a promoter link with AD downregulated *IL6R*. Interleukin 6 (IL6) signaling has been previously implicated in AD; however, its role has been unclear since IL6 can be pro-inflammatory (trans-signaling pathway) or anti-inflammatory (classical pathway)⁷. SPI1 regulation of *IL6R* expression in microglia may be a potential area of interest for future AD research. Additionally, NRF1 has a promoter link in oligodendrocytes to AD downregulated *AK5*, which encodes a brain-specific adenylate kinase, reinforcing our hypothesis that NRF1 may be affecting oligodendrocyte energy resources.

Trajectory analysis to study NRF1 and SREBF1 in oligodendrocytes

Similar to the oligodendrocyte TF regulatory network, we specifically examined genes with promoters or gl-cCREs with accessible TF binding sites, and therefore this analysis predicts candidate target genes that are activated by TF binding events (positive trajectory correlation), as well as target genes that are repressed by TF binding (negative trajectory correlation) (Fig. 5f). Interestingly, one of the candidate target genes of SREBF1 is AD DEG *SELENOP*, previously implicated in AD but not explored specifically in oligodendrocytes⁸. Finally, we identified enriched GO terms throughout the trajectory, revealing biological processes changing throughout the oligodendrocyte trajectory, such as terms related to neuronal maintenance, ERK and MAPK signaling, and cell motility (Extended Data Fig. 6). Together, these results further unravel the oligodendrocyte-specific gene regulatory changes in late-stage AD and highlight genes of interest for further research into oligodendrocyte disease biology.

Trajectory analysis to study SPI1 and ETV5 and the DAM gene signature in human microglia

We found that motif variability of both SPI1 and ETV5 increases with the *TREM2*-independent DAM signature and the proportion of AD nuclei (Extended Data Fig. 6). We found ETV5, which has been implicated in the regulation of A β deposition and gliosis in AD mouse models⁹, was positively correlated with genes at the end of the trajectory (Fig. 6f). Through GO term enrichment, we identified microglia t-DEGs at the beginning and end of the trajectory are unsurprisingly related to neuroinflammation, with enriched GO terms such as positive regulation of cytokine production and cellular response to interferon-gamma; however, we also discovered GO terms towards the end of the trajectory related to the extracellular matrix (Extended Data Fig. 6). A recent study has demonstrated the loss of perineuronal nets, extracellular matrix structures, in AD is mediated by microglia¹⁰.

Additionally, due to the striking absence of a *TREM2*-dependent DAM gene signature in the microglia trajectory, we further investigated the DAM gene signature through a cross-species meta-analysis of the PFC in three human snRNA-seq datasets^{11,12} and one 5XFAD mouse snRNA-seq dataset¹², where we identified twelve microglial consensus clusters (Supplementary Fig. 12). Several of these clusters were enriched in human AD and 5XFAD samples (Supplementary Fig. 12), including microglia consensus cluster 6, which showed the highest enrichment of *TREM2*-dependent stage 2 DAM genes (Supplementary Fig. 12). Differential gene expression analysis between nuclei from human AD and nuclei from 5XFAD in this cluster demonstrated that most of the DAM genes and many of the homeostatic genes identified in Keren-Shaul *et al.* are significantly downregulated in human nuclei, suggesting that the DAM gene signature has been difficult to recapitulate in human AD due to species dependence (Supplementary Fig. 12, Supplementary Data 7).

Trajectory analysis to study CTCF and FOSL2 in astrocytes

Interestingly, we identified AD DEG *TNC* at the end of the astrocyte trajectory (Fig. 6l). *TNC*, which encodes Tenascin C, has been shown to be upregulated in reactive astrocytes¹³, and we found that both CTCF and FOSL2 have promoter links to *TNC*. However, the CTCF motif variability trajectory is negatively correlated with *TNC* expression, while FOSL2's is positively correlated, indicating that these TFs have contrasting regulatory roles for *TNC*. Additionally, GO term enrichment analysis of astrocyte t-DEGs identified immune related terms, such as response to cytokine and cellular response to interferon gamma (Extended Data Fig. 6).

Cell-type-specific cis-regulation at AD GWAS loci

We analyzed the cis-regulatory structure of select AD GWAS loci by comparing CCANs between late-stage AD and control in a cell-type-specific manner, identifying regulatory connections that are gained or lost with disease. Specifically, we analyzed *BIN1* (Fig. 7d) and *ADAM10* (Fig. 7e) in oligodendrocytes; *BIN1* (Fig. 7f) and *APOE* (Fig. 7g) in microglia; *SLC24A4* (Fig. 7h) and *APOE* (Fig. 7i) in astrocytes. By overlaying the co-accessibility map with chromatin accessibility signal and GWAS statistics along the genomic axis, we begin to unravel the potential regulatory mechanisms disrupted by causal disease variants. *BIN1*, which has been implicated in both inflammation and myelination^{14–16}, has several cell-type specific regulatory links in microglia and has links flanking the causal SNP rs4663105 in both microglia and oligodendrocytes. In oligodendrocytes, the *ADAM10* locus, one of the most recently identified risk loci^{17,18}, has an AD-specific co-accessible link near the causal SNP rs442495.

We also found an enrichment of other GWAS traits, such as multiple sclerosis (MS) and inflammatory bowel disorder (IBD), in our microglia populations, suggesting that these disease-

associated polymorphisms influence gene regulatory mechanisms involved in the immune response across disorders.

Meta-analysis of single-cell co-expression networks using scWGNCA

Previous single-cell transcriptomic studies that used WGCNA to build co-expression networks were limited by the high level of noise inherent to the data, resulting in gene modules that could only capture cell-population marker genes and associated biology corresponding to individual cell-types (Extended Data Fig. 7)^{19,20}. Our revised approach uses aggregated expression profiles in place of potentially sparse single cells, whereby *metacells* are constructed from specific cell-populations by computing mean gene expression from 50 neighboring cells using k-nearest neighbors (Methods, Extended Data Fig. 7). This approach is analogous to cicero's²¹ construction of CCANs in snATAC-seq data, which relies on the aggregation of chromatin accessibility profiles to combat data sparsity in single cells. We selected 50 neighboring cells, since fewer or greater cells resulted in a loss of gene modules that may be biologically relevant (Extended Data Fig. 7).

Increasing our analytical power, we re-processed published AD snRNA-seq data from Mathys *et al.*¹¹ and used iNMF to integrate with our snRNA-seq data (Methods, Extended Data Fig. 8). Additionally, we performed bulk RNA-seq in early- and late-stage AD cases, as well as pathological controls (UCI; Control n = 40, Early-pathology AD n = 24, Late-pathology AD n = 26 samples), and curated additional AD bulk-tissue RNA-seq samples from ROSMAP²² (Control n = 109, Early-pathology AD n = 207, Late-pathology AD n = 311 samples). Finally, we used consensus WGCNA²³, a meta-analytical approach, to jointly form co-expression networks in *metacells* constructed from the integrated snRNA-seq dataset and bulk-tissue RNA-seq data of

the human PFC from two distinct cohorts. We call this approach Single-nucleus Consensus WGCNA (scWGCNA; Extended Data Fig. 7-10; Supplementary Data 7), performed iteratively for each cell-type, where each edge in a co-expressed module is supported by both bulk-tissue RNA-seq data (this study and ROSMAP²²) and aggregated snRNA-seq data (this study and Mathys *et al.*).

AD-upregulated module OM4 demonstrated enrichment of GO terms related to MAPK/ERK signaling, which we also identified in our trajectory analysis, whereas AD-downregulated module OM5 was enriched in GO terms associated with ion homeostasis and cellular signaling (Supplementary Fig. 11). Moreover, we noted the enrichment of OM1 genes particularly in intermediate oligodendrocyte clusters from both snRNA-seq and snATAC-seq (Supplementary Fig. 11). Altogether, these modules highlight key biological processes disrupted in oligodendrocytes in AD.

Supplementary Discussion

In comparison to previously published single-cell transcriptomic studies of the human brain in AD, we profiled glia to a much greater extent, enabling us to capture a more complete picture of the cellular heterogeneity in astrocytes, microglia, and oligodendrocytes. Community detection followed by differential expression and differential accessibility analyses independently resolved heterogeneous cell states with respect to control and AD. Using an integrated multi-omic trajectory analysis, we found a significantly increased proportion of DAAs and stage 1 DAMs in late-stage AD. Previous studies of the 5XFAD model identified these gene expression signatures, and we have expanded this work through characterization of both epigenomic and transcriptomic signatures of disease-associated glia in human AD. Although a

recent study has suggested that snRNA-seq is limited in profiling activated microglia²⁴, we found in our cross-species meta-analysis that the DAM signature is species-dependent, as we examined nuclei from both mouse and human. Additionally, we demonstrated the strength of profiling open chromatin profiles and gene expression in single nuclei by interrogating gl-cCREs, TF motif variability, and gene expression along the same cell-type trajectory. Using this strategy, we observed dynamic regulatory behavior of SPI1, an AD risk gene and a known regulator of AD-associated genes in microglia, whose TF motif variability was upregulated in a subset of microglia. We reveal that SPI1 may be acting as a repressor in late-stage AD microglia; however, further research is required to confirm this.

Although we only examined late-stage AD samples, and considering that AD is a progressive disorder, the chromatin landscape of human AD has not been previously described; therefore, our study provides important insights into the epigenetics of AD despite the lack of early-stage AD samples. In addition, it is still debated whether or not postmortem tissue from cases with early-stage pathology truly represent cases of early AD. In our co-expression analysis, we leveraged external datasets in order to provide insight into systems-level changes in AD at both early and late stages.

Supplementary Methods

Immunohistochemistry

Fixed human postmortem tissue was sectioned on a vibratome (Leica Biosystems, Buffalo Grove, IL) at 30 µm. Sequential sections were collected and stored in PBS with 0.02% sodium azide until use. Following standard immunohistochemistry protocols, free-floating sections were washed with PBS prior to incubation with 3% H₂O₂ and 10% methanol in PBS for 30 min to quench endogenous

peroxidase activity. Following washes, sections were placed in sodium citrate buffer and heated at 80°C for 30 min. To prevent nonspecific antibody binding, sections were blocked with 3% BSA, 5% normal serum and 0.2% TritonX-100 in PBS for 1hr at room temperature. The sections were then incubated overnight at 4°C with anti-SREBP1 (1:200, Cat#NB100-2215, Novus Biologicals) prepared in PBS containing 2% BSA plus 5% normal serum. After consecutives washes, sections were incubated with the secondary antibody (Horse anti-rabbit, Cat#BA-1100-1.5, Vector Laboratories) prepared in PBS containing 2% BSA plus 5% normal serum for 1 hour at room temperature. This was followed by amplification of the signal using an avidin-biotin complex peroxidase kit and 3,3' diaminobenzidine substrate kit (Vector Laboratories, Burlingame, CA). Each tissue section was mounted on a glass slide and coverslipped with DPX mounting media. All tissue sections were processed within a single experiment to reduce variability, and the researcher was blinded to the two different groups. The entire tissue slice was scanned at 40x magnification with the Aperio Versa 200 slidescanner. The Aperio ImageScope (v12.4.3.7009) software was used to draw four random 600 x 600 micron boxes in the white matter for each tissue section, and the Aperio positive pixel count algorithm (version 9) was used to quantify the amount of specific staining in each annotation box. A researcher blind to all samples conducted the analysis. We used linear mixed effects model as in our other staining experiments.

Processing sequencing reads from droplet-based single-nucleus RNA and ATAC -seq

Chromatin accessibility was quantified in single nuclei by mapping snATAC-seq reads to the hg38 reference genome (GRCh38.p12) and calling peaks using cellranger-atac count (v1.1.0, 10X Genomics) on each of the 20 sequencing libraries. We constructed a merged peak by barcodes matrix for all snATAC-seq libraries using the cellranger-atac aggr function, which

normalizes accessibility signals by sampling fragments from GEM wells to ensure uniform distributions of enriched Tn5 cut sites across libraries. The resulting peaks by barcodes matrix was binarized for downstream analysis. Gene expression was quantified by mapping reads from each of the 18 snRNA-seq libraries to a pre-mRNA build of the same reference genome, ensuring that we account for unprocessed nuclear transcripts, using cellranger count (v3.0.2 10X Genomics). We constructed a merged gene by barcodes matrix for all snRNA-seq libraries using the cellranger aggr function, normalizing for read depth across libraries.

Quality control for filtering of low-quality nuclei

Low-quality nuclei were identified and removed from the snATAC-seq dataset based on the following criteria: fewer than 300 or greater than 10,000 fragments mapping to peak regions, fewer than 15% of reads mapping to peaks, greater than 1% of reads mapping to ENCODE blacklist regions, and nucleosome signal greater than 10. Nucleosome signal was computed using the NucleosomeSignal function from the R package Signac (v0.2.0)²⁵ with default arguments. Low quality nuclei were identified and removed from the snRNA-seq dataset based on the following criteria: fewer than 200 or greater than 10,000 genes detected, and greater than 10% of reads mapping to mitochondrial genes. Additional nuclei were removed from both datasets if they clustered into a group of fewer than 100 nuclei, in total removing 298 nuclei from the snRNA-seq dataset and 12,983 nuclei from the snATAC-seq dataset. We also ran three different doublet detection methods in our snRNA-seq dataset—DoubletDetection (v3.0)²⁶, Scrublet (v1.0)²⁷, and DoubletFinder (v2.0)²⁸; however, we found that there were conflicting results from the different methods (Supplementary Table 3). We found one snATAC-seq cluster of 2,198

cells showing gene activity of marker genes from neuronal and glial cell types and removed it from subsequent analysis as a likely doublet cluster.

Dimensionality reduction and initial clustering

The filtered chromatin accessibility matrix was processed in Signac using Latent Semantic Indexing (LSI), which applies Term Frequency Inverse Document Frequency (TF-IDF) normalization and Singular Value Decomposition (SVD) dimensionality reduction. Technical variation from sequencing batches was corrected in the LSI representation using a Mutual Nearest Neighbors (MNN) approach implemented in the align_cds function of the R package monocle3^{29–31} (v 0.2.0). The corrected data matrix was embedded in two spatial dimensions using UMAP³². An initial cell clustering was performed using the Leiden clustering algorithm³³ in monocle3, which returns cluster assignments for single cells as well as UMAP partition assignments, which can be thought of as ‘super-clusters’. The filtered gene expression matrix was log-normalized and centered, followed by batch correction and dimensionality reduction via integrated non-negative matrix factorization (iNMF) in the R package LIGER^{34,35} (v0.4.2) with parameters k=30 for the number matrix factors and lambda=5 as the dataset-specific factor penalization. The iNMF reduced dataset was used for UMAP embedding and initial cell clustering with the FindNeighbors and FindClusters functions in the R package Seurat^{36,37} (v3.1.2), and res=0.9 as the cluster resolution parameter.

Inference of gene activity from chromatin accessibility

Gene activity signals were computed for 19,854 genes by summing snATAC-seq fragments that intersect a gene’s body or promoter regions. This was accomplished by obtaining gene body

coordinates for all protein coding genes in the reference genome, extending these regions upstream by 2000 bp to account for the promoter, and intersecting with snATAC-seq fragments using the FeatureMatrix function from Signac.

Constructing multi-omic glial cell type pseudotime trajectories

For oligodendrocytes, microglia, and astrocytes we constructed multi-omic trajectories analysis using monocle3. First, we isolated an individual cell-type from the integrated multi-omic dataset and re-embedded using UMAP. A principal graph was then applied to the multi-omic cell-type specific embedding using the monocle3 function learn_graph with default parameters, giving each cell a trajectory assignment. We oriented the start of trajectory programmatically using the strategy suggested by the authors of monocle3 (see monocle3 documentation). We computed the proportion of nuclei from AD samples across 50 evenly sized bins ordered along the trajectory and computed Pearson correlation between the proportion of AD nuclei with the trajectory assignment. We performed a gene scoring analysis for literature-derived gene sets^{3-5,38} relevant to each cell-type, in which we used the Seurat AddModuleScore function to compute gene module scores for a particular gene list, then plotted the trajectory of the averaged gene module score in each of the 50 trajectory bins. The following gene lists were used for gene scoring analysis: Mature ODC signature (*KLK6, APOD, SLC5A11, PDE1A*); myelin forming ODC signature (*MAL, MOG, PLP1, OPALIN, SERINC5, CTPS1*); newly formed ODC signature (*TCF7L2, CASR, CEMIP2, ITPR2*); homeostatic MG signature (*HEXB, CST3, CX3CR1, CTSD, CSF1R, CTSS, SPARC, TMSB4X, P2RY12, CIQA, CIQB*); stage 1 *TREM2* independent DAM MG signature (*TYROBP, CTSB, APOE, B2M, FTH1*); stage 2 *TREM2* dependent DAM MG signature (*TREM2, AXL, CST7, CTSL, LPL, CD9, CSF1, ITGAX, CLEC7A, LILRB4, TIMP2*);

GFAP-low ASC signature (*LUZP2*, *SLC7A10*, *MFGE8*); GFAP-high ASC signature (*GFAP*, *ID3*, *AQP4*, *MYOC*, *ID1*, *FABP7*); Disease Associated Astrocyte (DAA) ASC signature (*GFAP*, *CSTB*, *VIM*, *OSMR*, *GSN*, *GGTA1P*). For the ODC trajectory, we reversed the programmatically assigned trajectory to align with the trajectory of newly formed ODC and mature ODC gene signatures. We used generalized linear models (GLMs) to identify trajectory differentially expressed genes (t-DEGs) using the fit_models function in monocle3.

Differential feature analysis

We performed a series of differential analyses on the basis of three key features: gene expression, chromatin accessibility, and TF motif variability. In all cases, we used the Seurat FindAllMarkers function with varying parameters to compare features between groups. For differential chromatin accessibility and differential TF motif variability, we used the ‘LR’ option to perform logistic regression analyses, and for differential gene expression we used the ‘MAST’ option, which applies the MAST algorithm (v1.12.0)³⁹, a hurdle model specifically designed for snRNA-seq data. We iteratively analyzed cell-types in a one vs all else manner in order to identify features that are specifically up- and down- regulated in each group compared to all other cells and used Bonferroni correction to adjust p-values. For analysis of sub-clusters, we performed a similar iterative one vs all DEG analysis only comparing clusters belonging to the same major cell-type. We identified differential features with disease using another set of differential feature analysis, using the Seurat FindMarkers function to compare cells within the same cell-type or cluster group based on the AD diagnosis of the patient of origin. EnrichR⁴⁰ was used to identify biological processes that are up- and down- regulated with disease in each cluster by using the top 200 DEGs by average log fold-change to query the GO Biological

Processes 2018 database. Similarly, we used GREAT⁴¹ to identify biological processes that are associated with differentially accessible chromatin regions that are up- and down- regulated with disease.

Modeling the pseudotemporal dynamics of gene expression and chromatin accessibility using a recurrent variational autoencoder (RVAE) neural network

We used the Python (v3.7.9) package RVAgene (v1.0)⁴² to model the dynamics of gene expression and chromatin accessibility along the glial cell trajectories using a PyTorch (v1.4.0) recurrent variational autoencoder (RVAE) implementation. First, we averaged the expression of t-DEGs at each of the 50 trajectory bins and scaled the values for each gene between -1 and 1 to construct the input matrix for RVAE analysis. For the encoder and decoder neural networks, we used a symmetrical architecture with two hidden layers and 48 nodes per layer, with two latent variable dimensions. Additionally, we used a mean squared error (MSE) loss function and trained the model using mini batches until the loss function reached convergence. After training the model, the full input matrix is passed back through the model to produce a reconstructed trajectory for each input gene and a feature embedding in two-dimensional latent space. A matrix of chromatin accessibility of gl-cCREs identified in each cell-type is then passed through the trained model to produce a reconstructed trajectory for each input feature and a feature embedding in the same two-dimensional latent space. For each feature (gene or gl-cCRE), we assign a trajectory rank based on the point along the trajectory that the feature reaches 75% of its maximum reconstructed value, where features with low ranks are highly expressed at the beginning of the trajectory and features with high ranks are highly expressed at the end of the trajectory. We computed Pearson correlations between reconstructed gene expression trajectories

and single-cell TF binding motif variability averaged in the same 50 trajectory bins and visualized these correlations on the two-dimensional latent feature embeddings. Similarly, we correlated reconstructed feature trajectories to the proportion of AD nuclei in each of the 50 trajectory bins and visualized the result on the latent feature embeddings.

Cross-species integrated analysis of microglia signatures

We used online-iNMF, a highly scalable and efficient implementation of iNMF in the LIGER R package (v 0.5.0)⁴³, to integrate snRNA-seq data from this study with three independent datasets (Mathys *et al.*, all samples; Zhou *et al.* human control samples and human AD samples with *TREM2* common variant; Zhou *et al.* WT and WT 5XFAD samples). To perform a cross-species analysis, we used homologous genes for iNMF integration. We performed microglia specific iNMF integration by taking microglia from each dataset based on cell-type assignments from the original studies and using k=15 and lambda=5 as iNMF parameters. UMAP and Leiden clustering were then used to identify 12 cross-species consensus clusters. DAM signature genes as defined in Keren-Shaul *et al.* were analyzed in the integrated microglia data by using the Seurat function AddModuleScore for the following gene lists: homeostatic microglia markers (*HEXB*, *CST3*, *CX3CR1*, *CTSD*, *CSF1R*, *CTSS*, *SPARC*, *TMSB4X*, *P2RY12*, *C1QA*, *C1QB*); stage 1 *TREM2* independent DAM markers (*TYROBP*, *CTSB*, *APOE*, *B2M*, *FTH1*); stage 2 *TREM2* dependent DAM markers (*TREM2*, *AXL*, *CST7*, *CTSL*, *LPL*, *CD9*, *CSF1*, *ITGAX*, *CLEC7A*, *LILRB4*, *TIMP2*). We performed differential gene expression analysis using the MAST algorithm between human and mouse cells in consensus cluster 6, only considering nuclei from AD patients and 5XFAD mice.

Estimating AD GWAS enrichment in the microglia pseudotime trajectory

We estimated the enrichment of AD GWAS polymorphisms from Jansen *et al.*¹⁷ using the R package gchromVAR (v0.3.2)⁴⁴, a modified version of the previously described chromVAR pipeline for identifying enriched TF binding motifs in single-cell chromatin accessibility data. We collected CAVIAR Bayes factor⁴⁵ GWAS fine-mapping posterior probabilities for polymorphisms at 30 causal blocks for AD¹⁷ from CausalDB⁴⁶, lifted coordinates to hg38 using the UCSC liftover (v377) tool, and formatted them for analysis with gchromVAR. gchromVAR deviation scores, which correspond to enrichment of the trait of interest in accessible chromatin regions, for AD were computed separately at distal chromatin regions and gene-proximal chromatin regions in microglia using the function computeWeightedDeviations. We then averaged deviation scores in each of the 50 microglia trajectory bins and computed linear regressions and Pearson correlations to quantify the change of GWAS enrichment throughout the trajectory.

Joint analysis of snRNA-seq with Mathys *et al* snRNA-seq

We re-analyzed snRNA-seq from the Mathys *et al.*¹¹ snRNA-seq study of AD in human samples. We followed the standard Seurat analysis procedure of log normalization, scaling, PCA, and UMAP, ensuring that original cell-type assignments grouped together in UMAP space. From the R package LIGER, iNMF was used to integrate our snRNA-seq dataset with the Mathys *et al.* dataset, using k=30 matrix factors and lambda=5.0 as the dataset specific factor penalization. A joint UMAP was then computed using the iNMF reduced merged dataset.

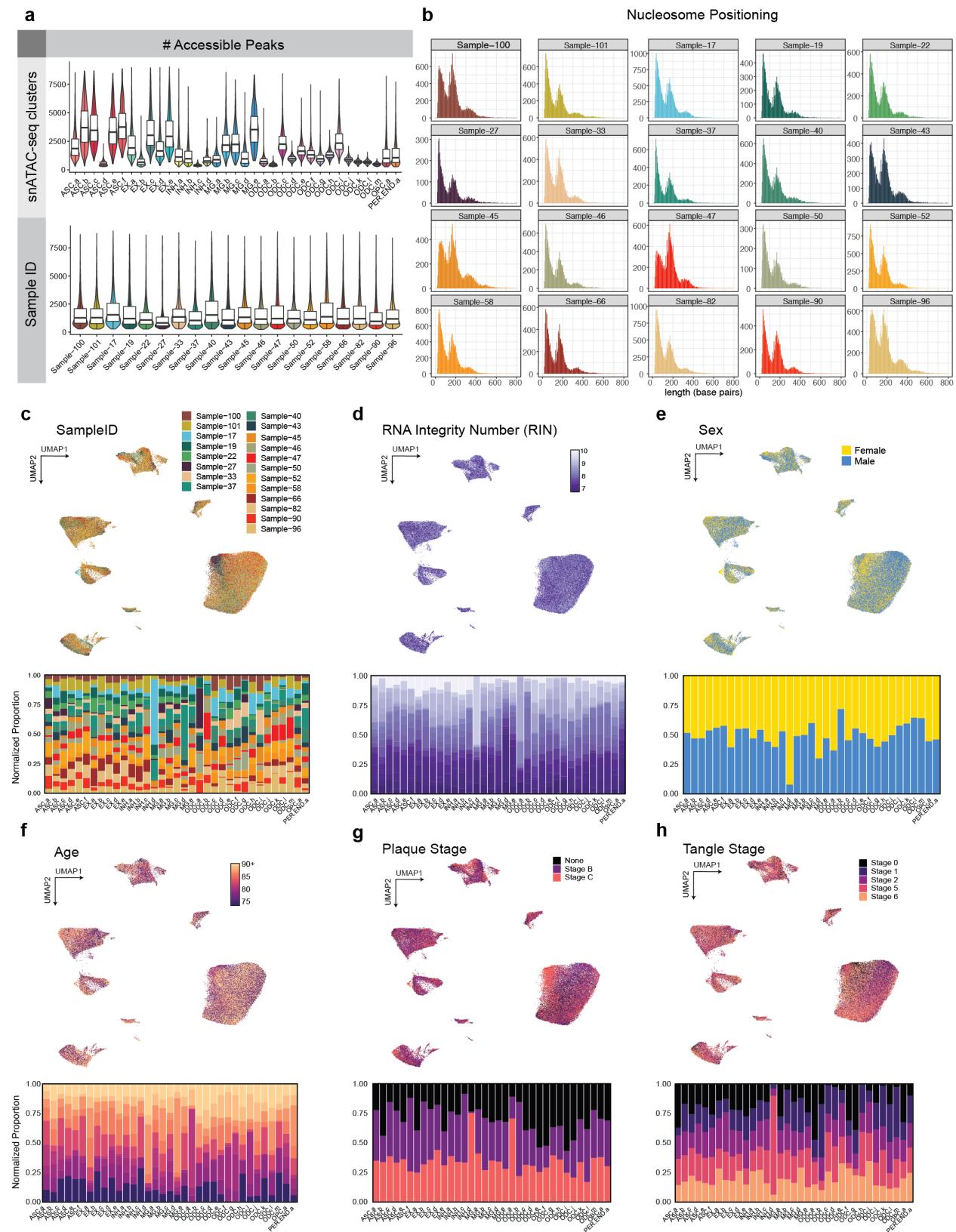
Processing sequencing reads from Bulk RNA-seq

We performed QC on raw fastq reads, sequenced at median read-depth of 50-million paired-end reads/sample, using FastQC. Reads were aligned to the reference genome (GRCh38.p13) using STAR aligner (v2.6.1)⁴⁷ and aligned BAM file QC metrics were obtained using PicardTools (v2.21.3) (<http://broadinstitute.github.io/picard/>). Gene expression was quantified as transcripts per million (TPM) using Salmon (v1.3.0)⁴⁸. We used linear regression model, similar to our previous work²³, to remove the effects of biological covariates like age, gender, post-mortem interval (PMI), and technical variables like batch (isolation and processing batches), RNA integrity number (RIN) and sequencing biases using PicardTools metrics. The final model used was implemented in R is as follows:

```
lm(expression ~ diagnosis + age + gender + PMI + batch + RIN + Seq.PC1 +  
Seq.PC2+Seq.PC3)
```

where Seq.PC1, Seq.PC2 and Seq.PC3 are sequencing PCs obtained from aggregating sequencing metrics obtained from Picard tools.

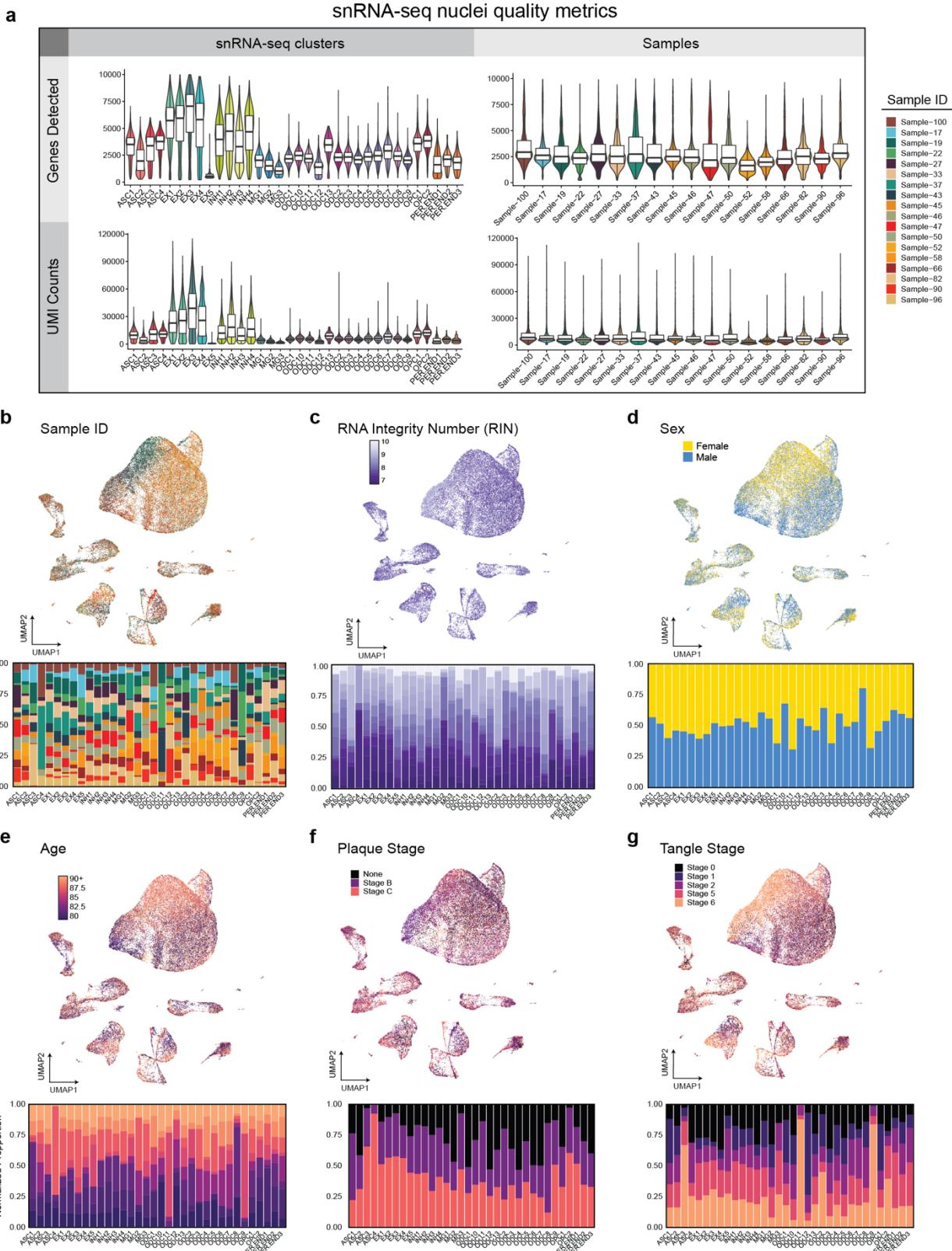
Supplementary Figures



Supplementary Figure 1: Data quality and categorical variables in snATAC-seq

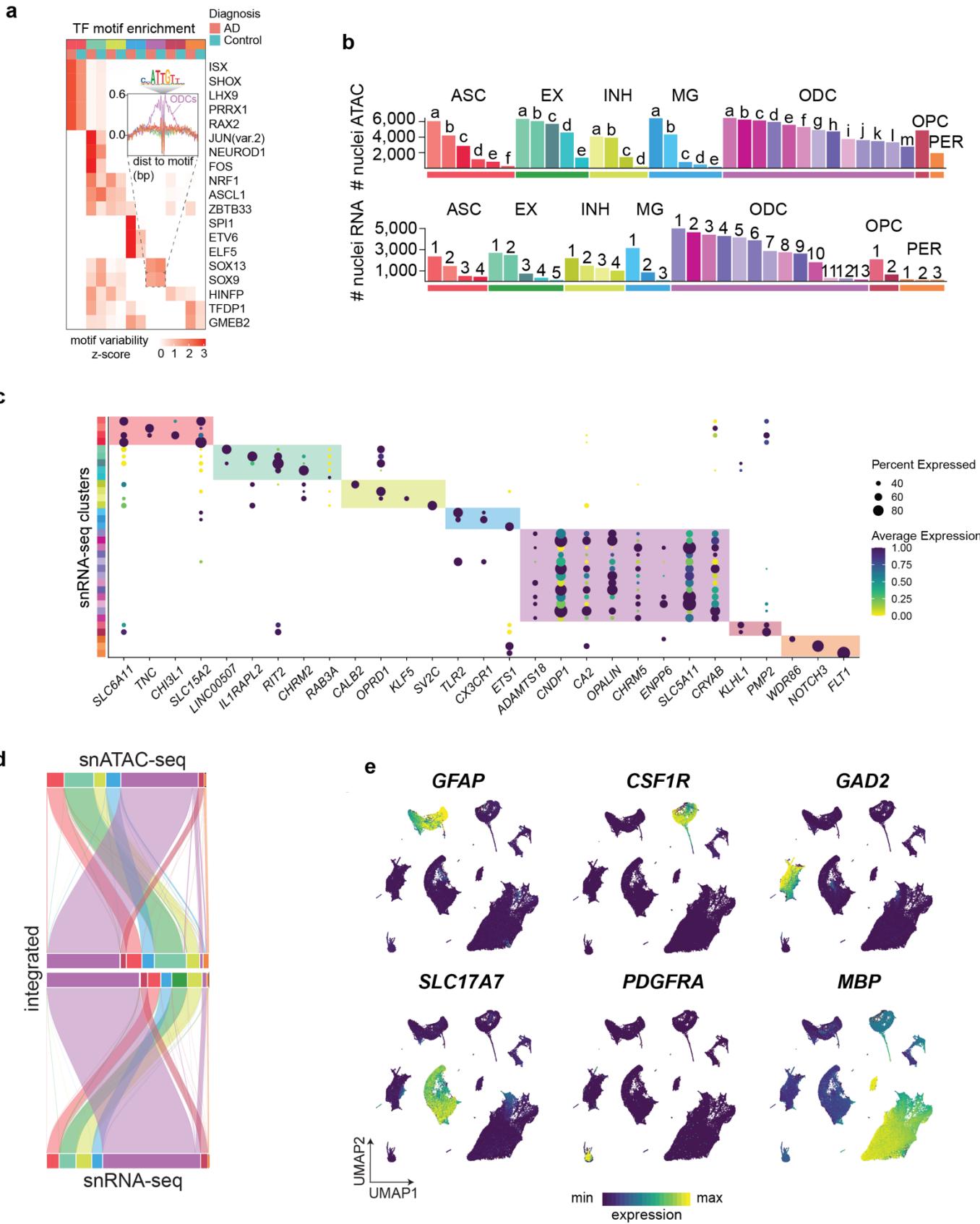
a, Distributions of the number of accessible peaks in each snATAC-seq cluster (top) and in each sample (bottom). Box boundaries and line correspond to the interquartile range (IQR) and median respectively. Whiskers extend to the lowest or highest data points that are no further than 1.5 times the IQR from the box boundaries. **b**, Histograms of nucleosome positioning in each biological sample. **c, d, e, f, g, h**, UMAP plots colored by categorical variables (top) and the normalized proportion of each categorical variable in each snATAC-seq cluster (bottom). Categorical variables: Sample ID (**c**), RNA integrity (**d**), sex (**e**), age at death (**f**), plaque stage (**g**), tangle stage (**h**).

snRNA-seq nuclei quality metrics



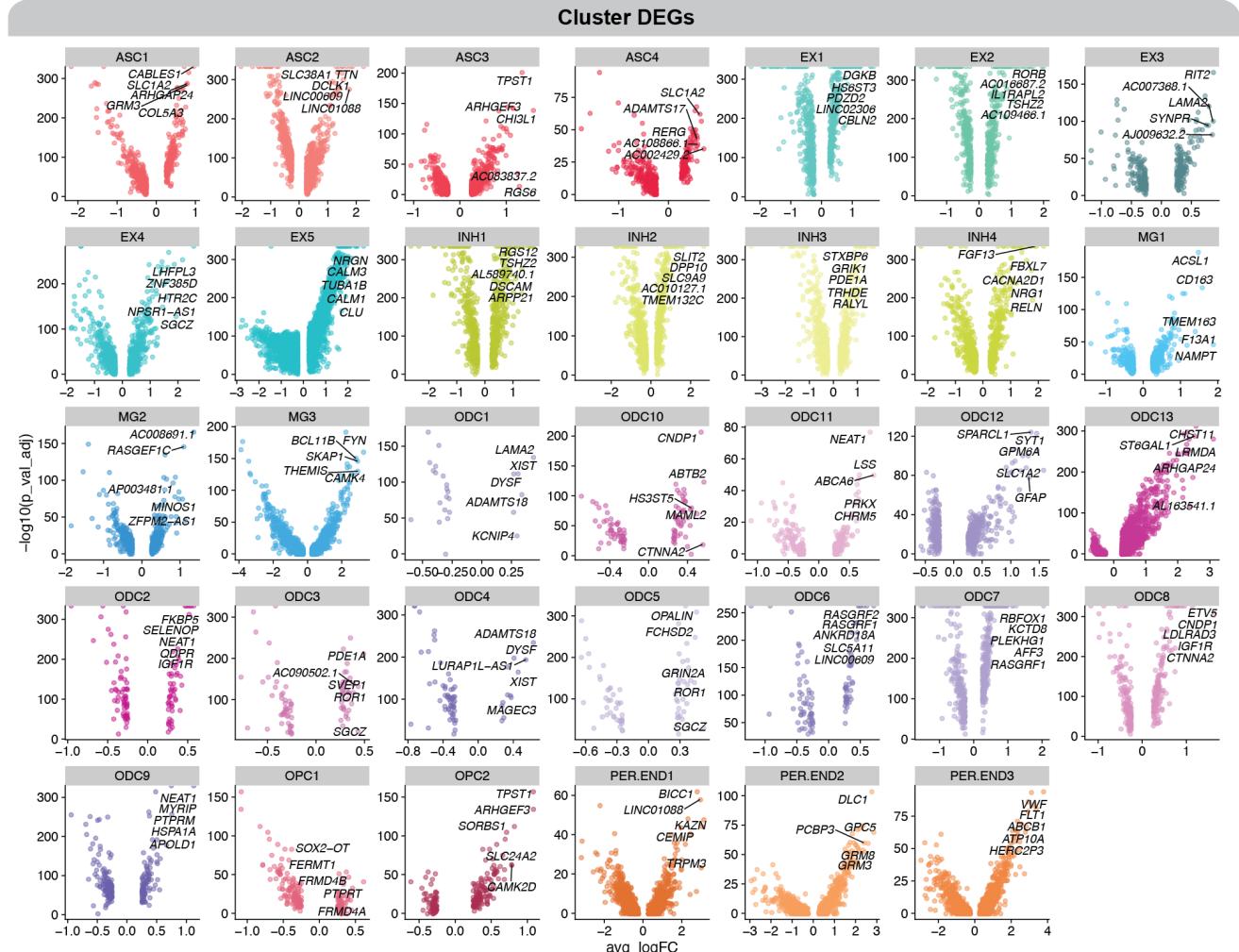
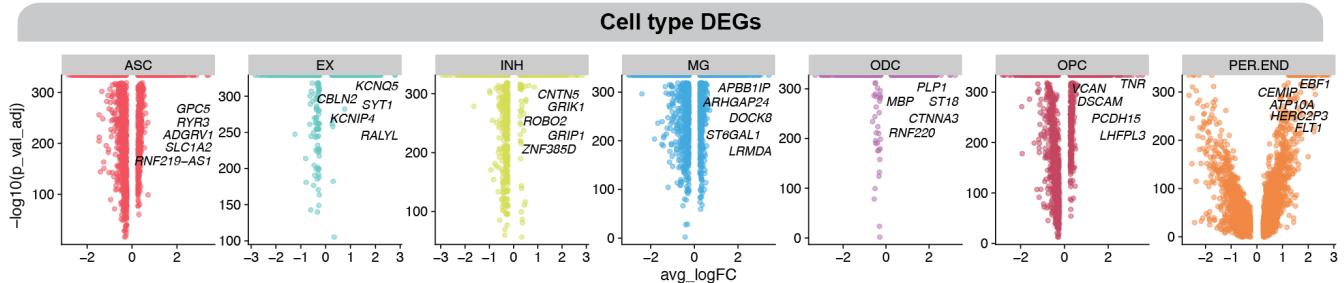
Supplementary Figure 2: Data quality and categorical variables in snRNA-seq

a, Distributions of the number of genes detected (top) and UMI counts (bottom) in each snRNA-seq cluster (left) and in each sample (right). Box boundaries and line correspond to the interquartile range (IQR) and median respectively. Whiskers extend to the lowest or highest data points that are no further than 1.5 times the IQR from the box boundaries. **b, c, d, e, f, g**, UMAP plots colored by categorical variables (top) and the normalized proportion of each categorical variable in each snRNA-seq cluster (bottom). Categorical variables: Sample ID (**b**), RNA integrity (**c**), sex (**d**), age at death (**e**), plaque stage (**f**), tangle stage (**g**).



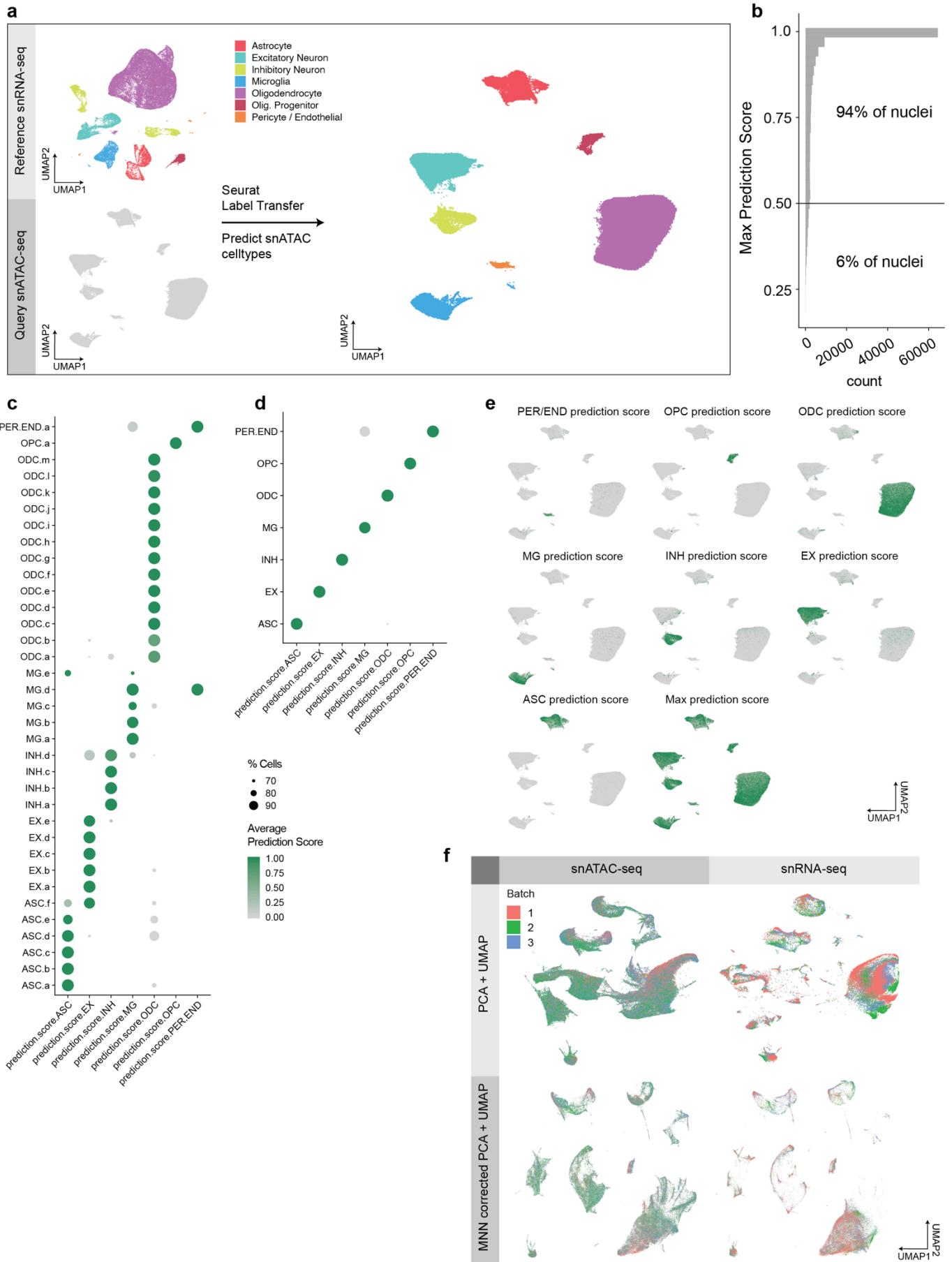
Supplementary Figure 3: Multi-omic cell type characterization

a, Heatmap of row-normalized TF motif variability by cell-type and diagnosis. Cell-type colors are the same as in Fig. 1b. Inner panel shows Tn5 bias subtracted TF footprinting for SOX9 for each major cell-type. **b**, Bar plots showing the number of nuclei in each cell-type cluster in snATAC-seq (top) and in snRNA-seq (bottom). **c**, Dot plot of differentially expressed genes in snRNA-seq clusters, colored by scaled gene expression. Dot size represents the percentage of nuclei in each cluster expressing each gene. **d**, Alluvial plot showing mappings between original cell-type assignments and their corresponding UMAP partitions in the joint representation. **e**, Integrated UMAP as in Fig. 1f colored by gene expression (snRNA-seq) and imputed gene activity (snATAC-seq) of canonical cell-type marker genes.

a**b**

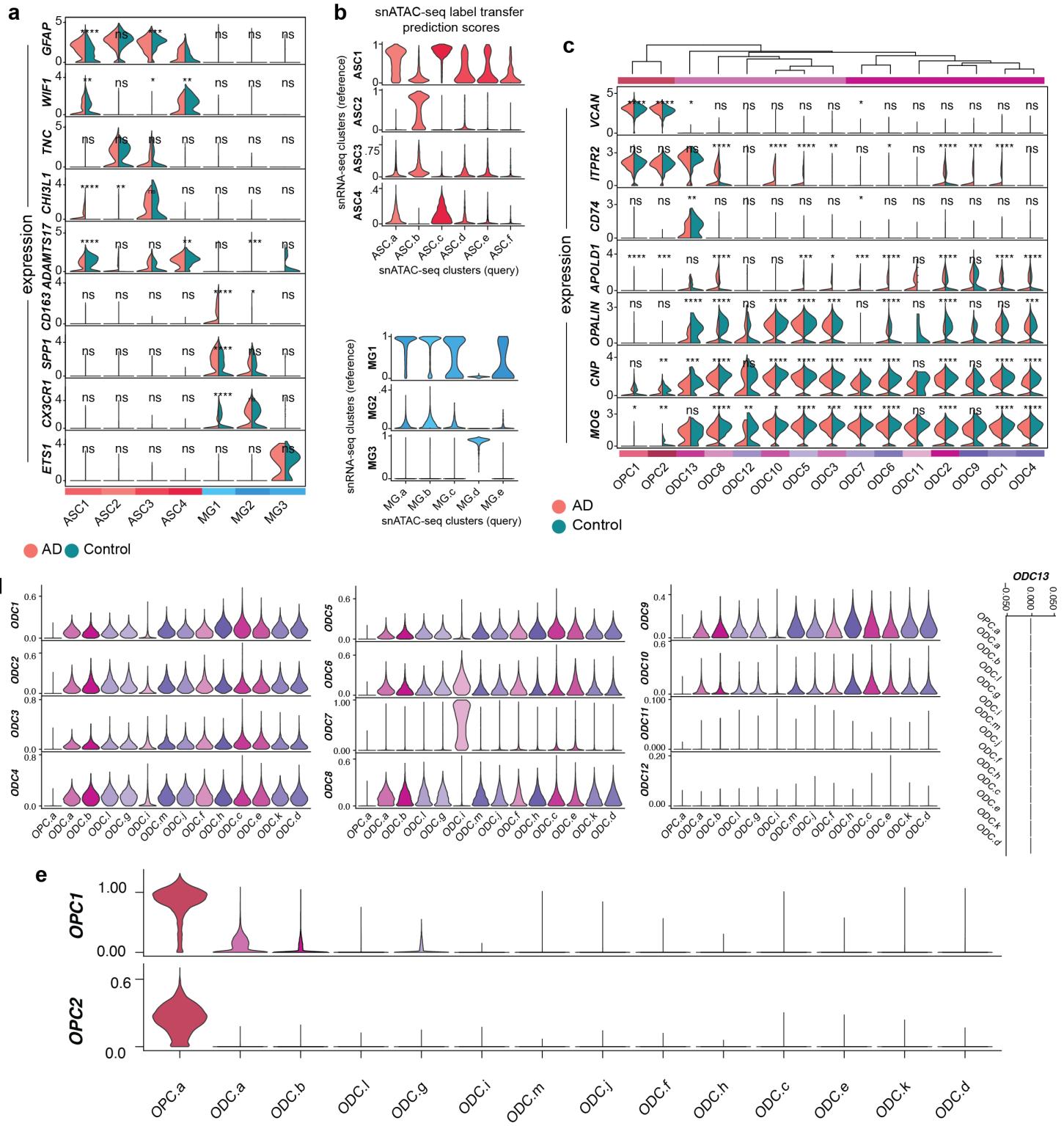
Supplementary Figure 4: Differentially expressed genes (DEGs) by cell population in snRNA-seq

a, b, Volcano plots showing the results of differential gene expression analysis in each snRNA-seq cluster (**a**) and cell-type (**b**), comparing between clusters or major cell-type. The top 5 genes by average log fold change are annotated. Cluster DEGs were identified by iteratively comparing expression in one cluster to all other clusters within the same major cell-type, and cell-type DEGs were identified by iteratively comparing expression in each cell-type to all other cell-types.



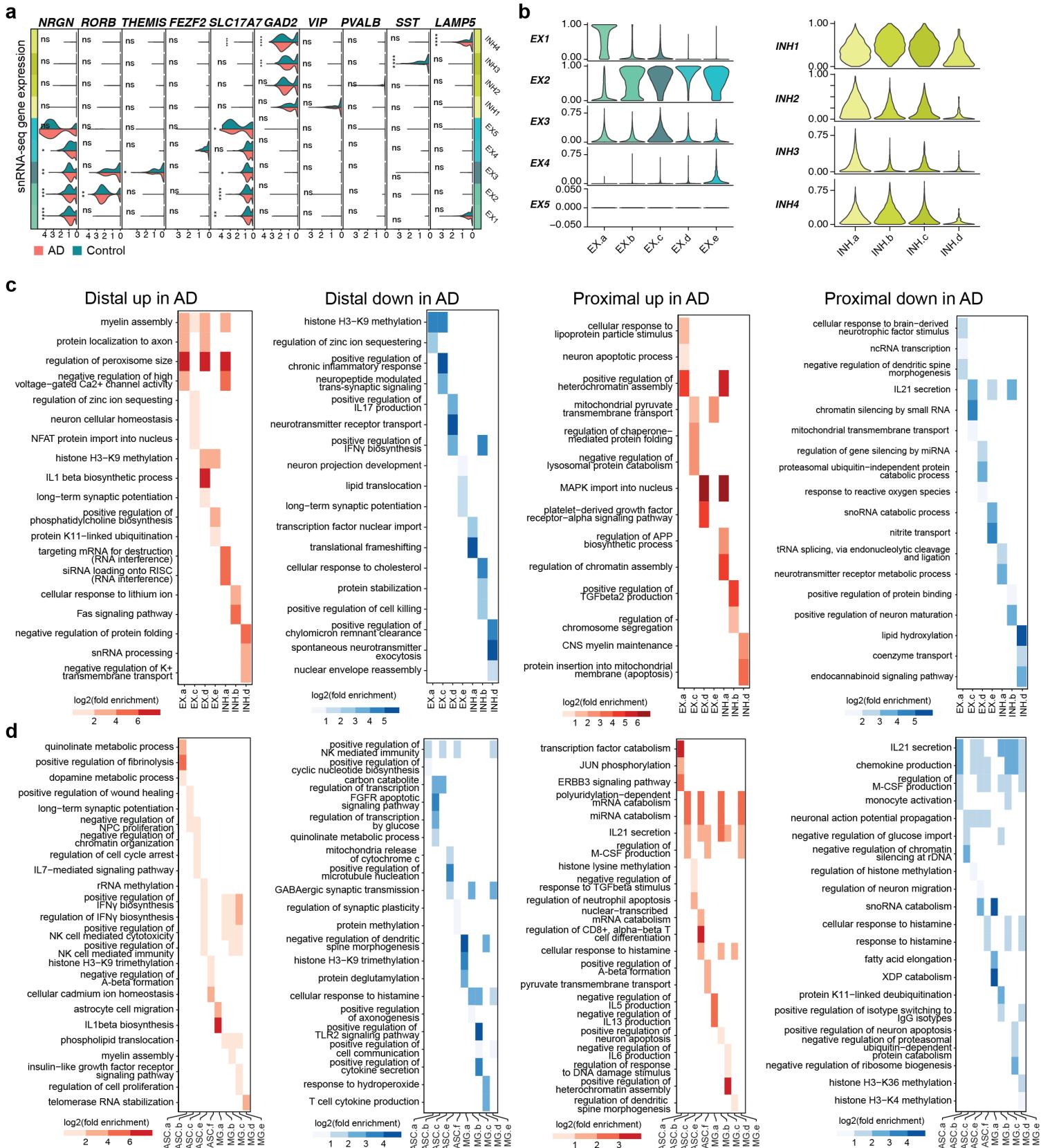
Supplementary Figure 5: Integration of snATAC-seq and snRNA-seq data

a, Schematic representation of Seurat label transfer procedure used to predict cell-types of nuclei from snATAC-seq using annotated snRNA-seq data. **b**, Histogram of the maximum cell-type prediction score in each snATAC-seq nucleus. **c**, **d**, Dot plot of cell-type prediction scores in snATAC-seq clusters (**c**) and snATAC-seq cell-type assignments (**d**). **e**, snATAC-seq UMAP plots colored by cell-type prediction scores. **f**, Integrated multi-omic UMAPs before (top) and after MNN batch correction (bottom).



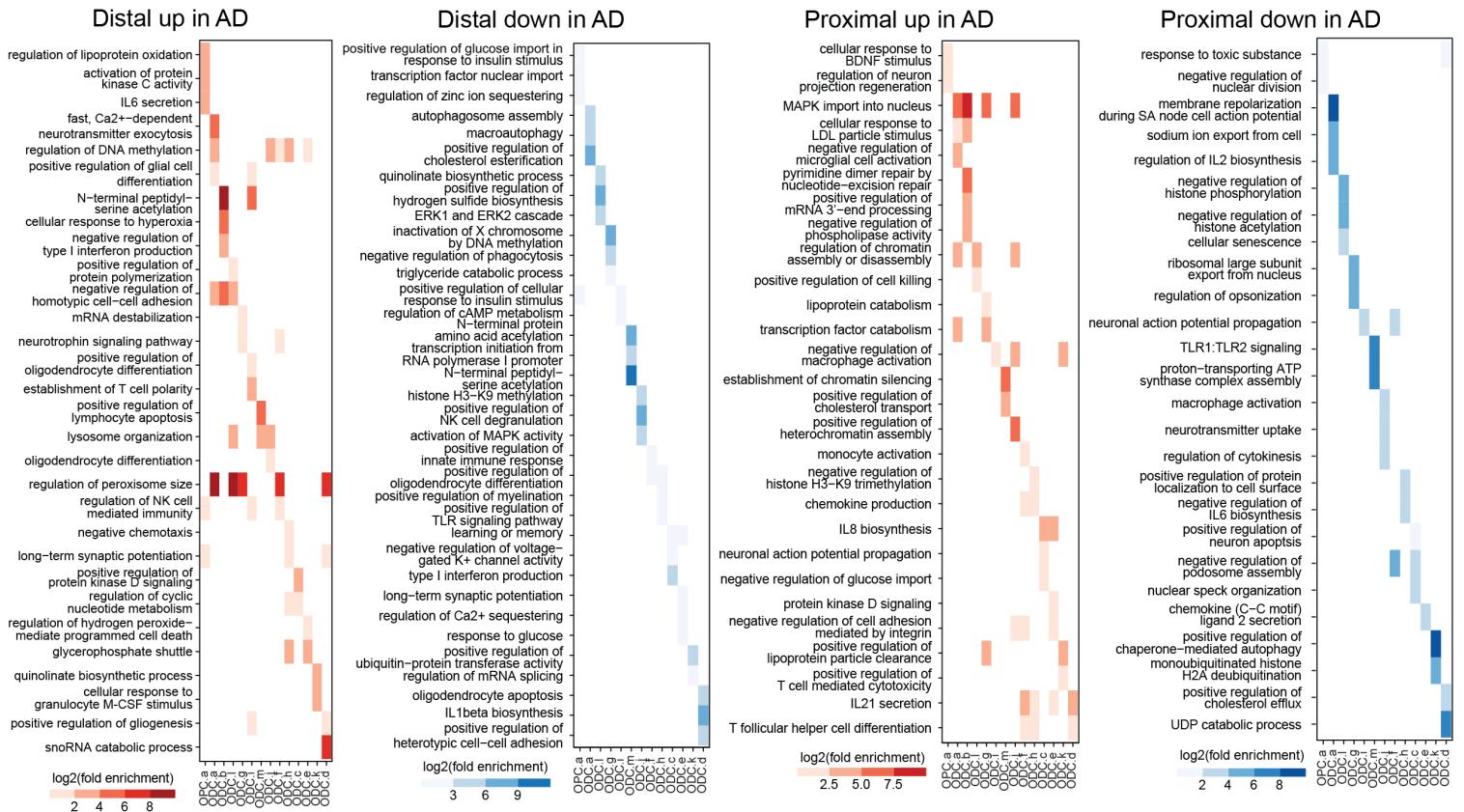
Supplementary Figure 6: Glial subpopulations

a, Violin plots of normalized gene expression for astrocyte and microglia cluster marker genes, split by diagnosis. Two-sided Wilcoxon test **** p <= 0.0001, *** p <= 0.001, ** p <= 0.01, * p <= 0.05. **b**, Violin plots of snRNA-seq cluster prediction scores for snATAC-seq astrocyte (top) and microglia (bottom) clusters. **c**, Violin plots of normalized gene expression for select genes in OPC and oligodendrocyte snRNA-seq clusters, split by diagnosis as in **a**. Clusters are ordered by the same hierarchy as in Fig. 2a. **d**, Violin plots of Seurat label transfer prediction scores for snATAC-seq for oligodendrocyte clusters, as in **b**. **b**, Violin plots of Seurat label transfer prediction scores for oligodendrocyte progenitor cell clusters, as in **b**.



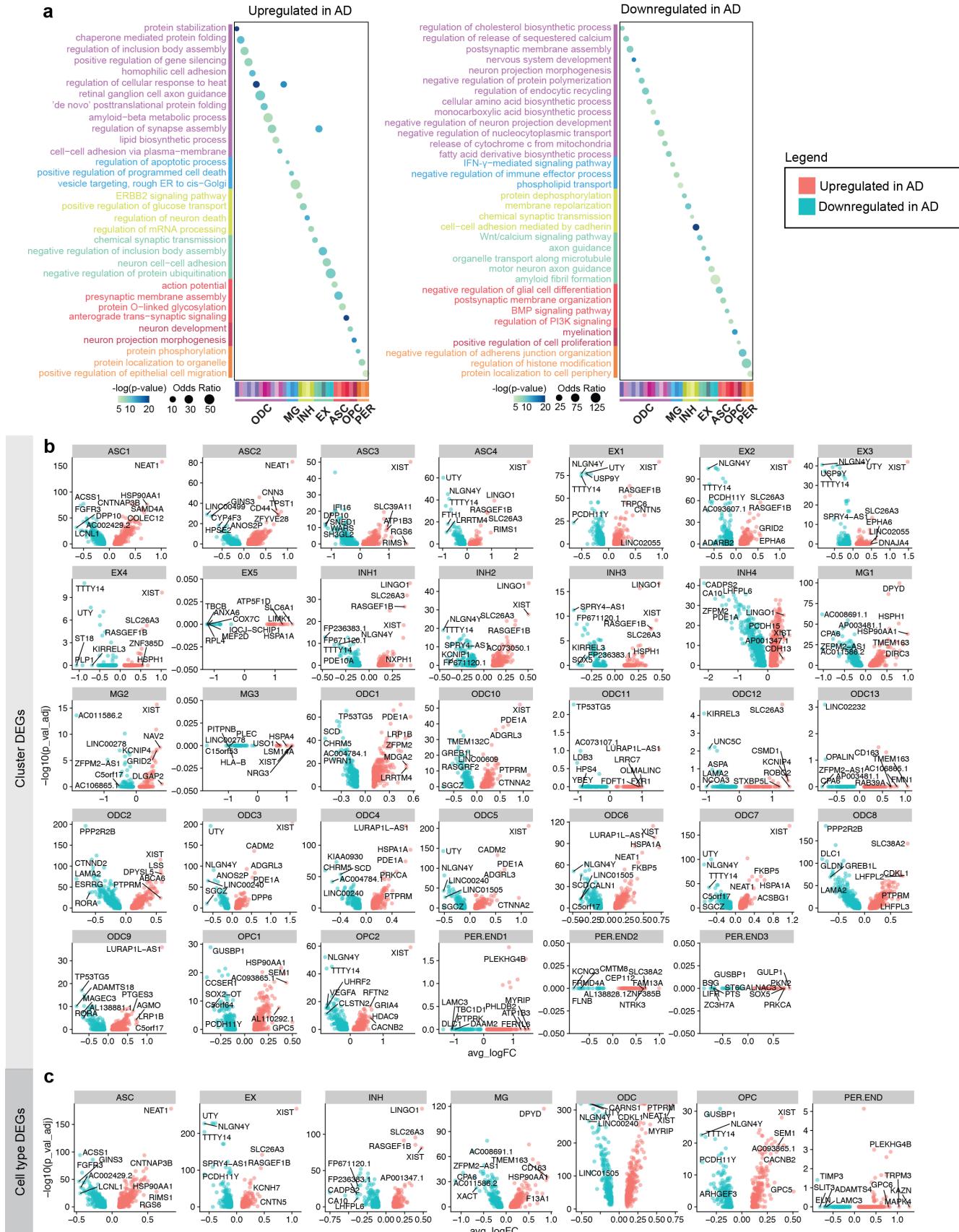
Supplementary Figure 7: Neuron subpopulations and differentially accessible regions (DARs) in snATAC-seq

a, Violin plots of normalized gene expression for excitatory and inhibitory neuron cluster marker genes, split by diagnosis. Two-sided Wilcoxon test ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$. **b**, Violin plots of Seurat label transfer prediction scores for snATAC-seq excitatory (left) and inhibitory neuron (right) clusters. **c, d**, GO term enrichment of distal and proximal peaks upregulated or downregulated in AD in neuronal clusters (**c**) and astrocyte and microglia clusters (**d**).



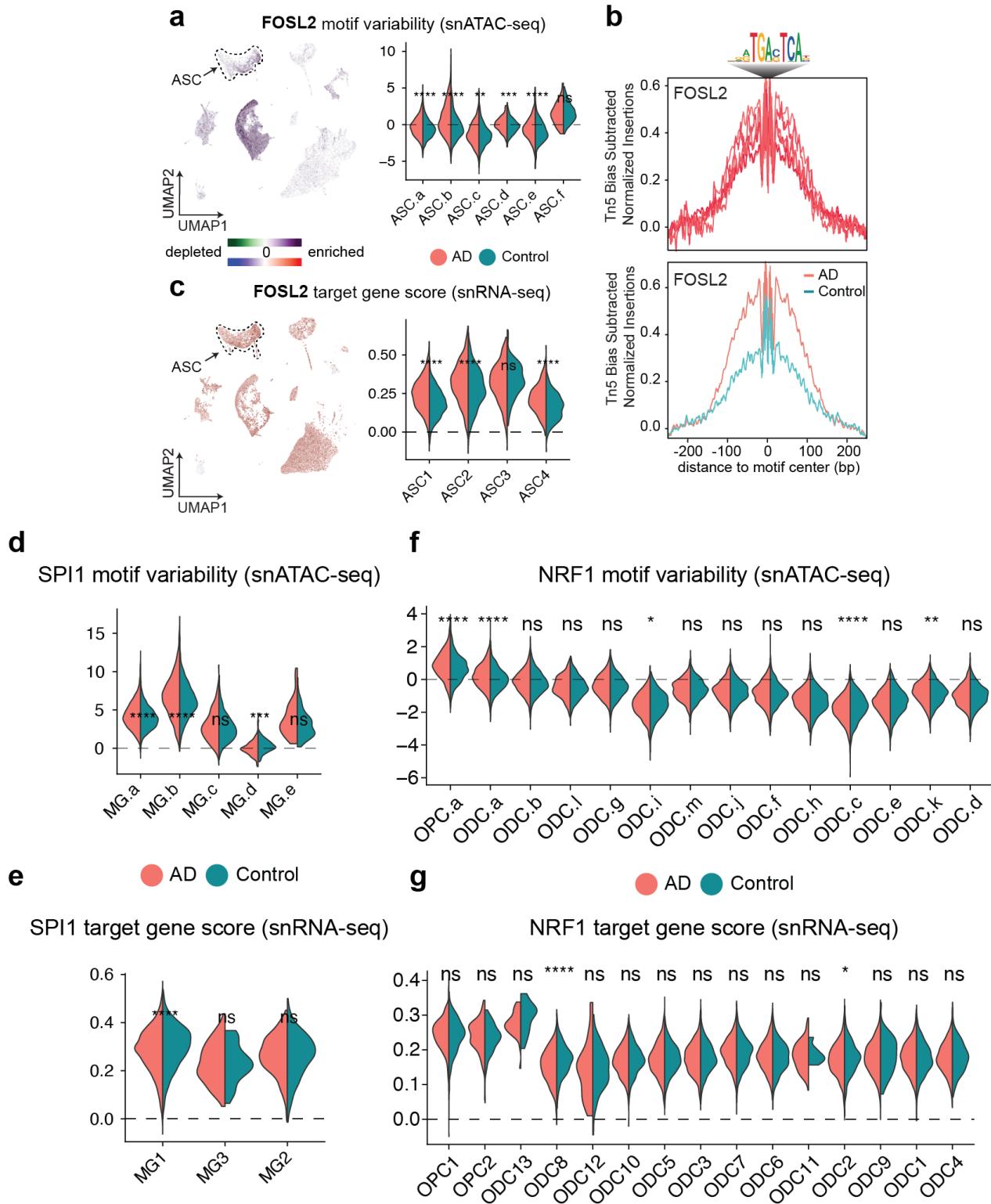
Supplementary Figure 8: Oligodendrocyte differentially accessible regions (DARs)

GO term enrichment of distal and proximal peaks upregulated or downregulated in AD in oligodendrocyte clusters.



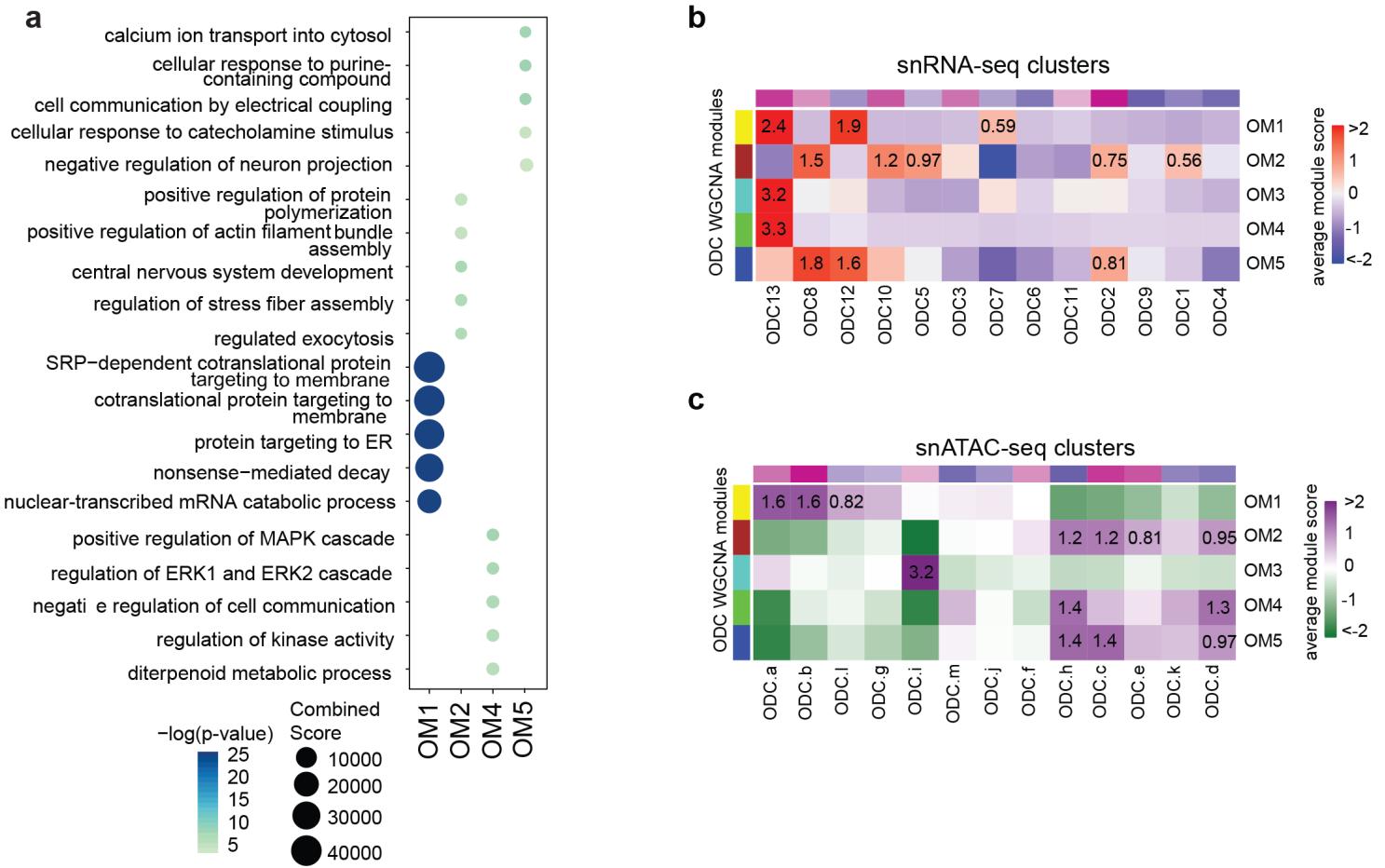
Supplementary Figure 9: Differentially expressed genes (DEGs) by AD diagnosis in snRNA-seq

a, Enriched GO terms for the top 200 significant genes upregulated (left) and downregulated (right) in AD in each snRNA-seq cluster. **b, c**, Volcano plots of DEGs computed by comparing gene expression between nuclei from control and late-stage AD samples in each snRNA-seq cluster (**b**) and cell-type (**c**). The top 5 up- and down-regulated genes by average log fold change are annotated.



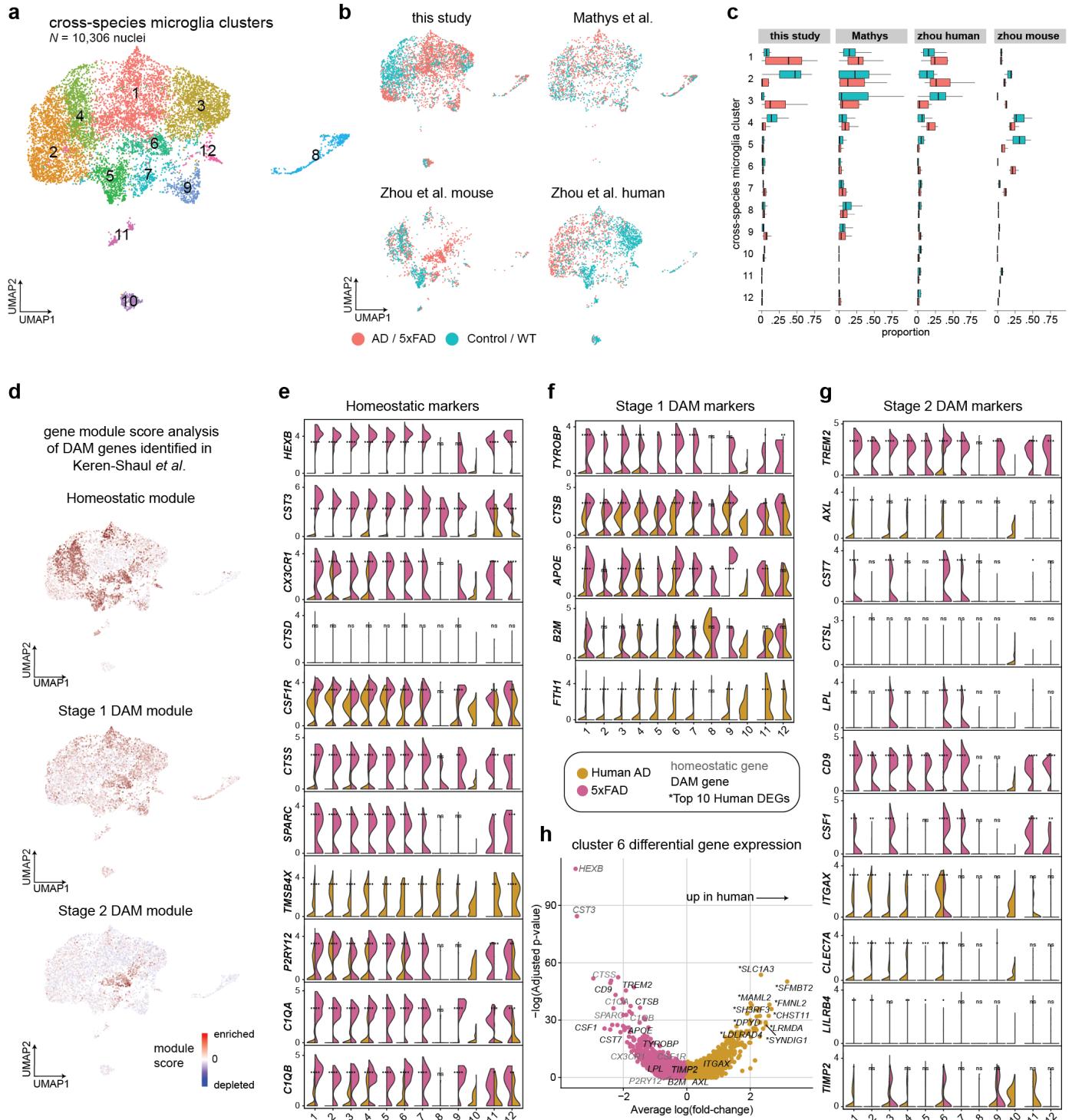
Supplementary Figure 10: Glial TF analyses

a, Left: snATAC-seq and snRNA-seq integrated UMAP colored by FOSL2 motif variability with astrocytes circled. Right: Violin plots of FOSL2 motif variability in snATAC-seq astrocyte clusters, split by diagnosis. **b**, Tn5 bias subtracted TF footprinting for FOSL2 by snATAC-seq astrocyte cluster (top) and by AD diagnosis (bottom). TF binding motif shown as motif logo above, and Tn5 bias insertions shown below. **c**, Left: Integrated UMAP colored by FOSL2 target gene score with astrocytes circled. Right: Violin plots of FOSL2 target gene score in snRNA-seq astrocyte clusters, split by diagnosis as in **a**. **d**, Violin plots of SPI1 motif variability in snATAC-seq microglia clusters, split by diagnosis as in **a**. **e**, Violin plots of SPI1 target gene score in snRNA-seq microglia clusters, split by diagnosis as in **a**. **f**, Violin plots of NRF1 motif variability in snATAC-seq oligodendrocyte clusters, split by diagnosis as in **a**. **g**, Violin plots of NRF1 target gene score in snRNA-seq oligodendrocyte clusters, split by diagnosis as in **a**. For violin plots, two-sided Wilcoxon test was used to compare control versus AD, ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.



Supplementary Figure 11: Oligodendrocyte scWGCNA analyses

a, GO term enrichment of oligodendrocyte co-expression modules. **b, c**, Heatmap showing row-normalized Seurat module scores of oligodendrocyte co-expression modules in snRNA-seq (**a**) and snATAC-seq (**b**) oligodendrocyte clusters.



Supplementary Figure 12: Integrated cross-species meta-analysis of microglia

a, UMAP dimensionality reduction on iNMF integration of microglia from four datasets: this study, Mathys *et al.*, Zhou *et al.* human samples, and Zhou *et al.* mouse samples. Each cell is colored by consensus microglia Leiden clustering assignment, and cluster numbers 1-12 are annotated based on cluster centroids. **b**, UMAP plots as in **a**, split by dataset of origin and colored by AD diagnostic status for human samples and transgenic status for murine samples. **c**, Box and whisker plots showing distributions of the proportion of each sample composing each cluster, split by AD diagnostic status for human samples and transgenic status for murine samples, n as in Supplementary Data 7. Box boundaries and line correspond to the interquartile range (IQR) and median respectively. Whiskers extend to the lowest or highest data points that are no further than 1.5 times the IQR from the box boundaries. **d**, UMAP feature plots colored by single-cell module scores for microglia gene signatures from Keren-Shaul *et al.* (homeostatic, stage 1 *TREM2*-independent DAM, stage 2 *TREM2*-dependent DAM). **e**, **f**, **g**, Violin plots showing the distribution of normalized gene expression values in AD and 5XFAD nuclei from microglia consensus cluster for homeostatic microglia genes (**e**), stage 1 DAM genes (**f**), and stage 2 DAM genes (**g**). Statistical significance is computed using a Wilcoxon rank sum test (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$). **h**, Volcano plot showing the results of differential gene expression analysis between nuclei from human AD and 5XFAD mice in microglia cluster 6. Annotated genes are from Keren-Shaul *et al.* gene signatures (gray, black), or the top 10 DEGs upregulated in human nuclei by average log(fold-change), denoted in black with a *.

References

1. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
2. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science (80-.).* **352**, 1586–1590 (2016).
3. Habib, N. *et al.* Disease-associated astrocytes in Alzheimer’s disease and aging. *Nat. Neurosci.* 1–6 (2020). doi:10.1038/s41593-020-0624-8
4. Marques, S. *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science (80-.).* **352**, 1326–1329 (2016).
5. Jäkel, S. *et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
6. Gallo, F. T., Katche, C., Morici, J. F., Medina, J. H. & Weisstaub, N. V. Immediate early genes, memory and psychiatric disorders: Focus on c-Fos, Egr1 and Arc. *Front. Behav. Neurosci.* **12**, 1–16 (2018).
7. Rothaug, M., Becker-Pauly, C. & Rose-John, S. The role of interleukin-6 signaling in nervous tissue. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 1218–1227 (2016).
8. Solovyev, N. Selenoprotein P and its potential role in Alzheimer’s disease. *Hormones* **19**, 73–79 (2020).
9. Suh, J. *et al.* Loss of Ataxin-1 Potentiates Alzheimer’s Pathogenesis by Elevating Cerebral BACE1 Transcription. *Cell* **178**, 1159-1175.e17 (2019).
10. Crapser, J. D. *et al.* Microglia facilitate loss of perineuronal nets in the Alzheimer’s disease brain. *EBioMedicine* **58**, (2020).
11. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer’s disease. *Nature* **570**,

- 332–337 (2019).
12. Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* **26**, 131–142 (2020).
 13. Wiese, S., Karus, M. & Faissner, A. Astrocytes as a source for extracellular matrix molecules and cytokines. *Front. Pharmacol.* **3** JUN, 1–13 (2012).
 14. De Rossi, P. *et al.* Predominant expression of Alzheimer's disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts. *Mol. Neurodegener.* **11**, 1–21 (2016).
 15. Crotti, A. *et al.* BIN1 favors the spreading of Tau via extracellular vesicles. *Sci. Rep.* **9**, 1–20 (2019).
 16. McAvoy, K. M. *et al.* Cell-autonomous and non-cell autonomous effects of neuronal BIN1 loss in vivo. *PLoS One* **14**, 1–28 (2019).
 17. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
 18. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
 19. Wu, Y. E., Pan, L., Zuo, Y., Li, X. & Hong, W. Detecting Activated Cell Populations Using Single-Cell RNA-Seq. *Neuron* **96**, 313-329.e6 (2017).
 20. Kashima, Y. *et al.* Combinatory use of distinct single-cell RNA-seq analytical platforms reveals the heterogeneous transcriptome response. *Sci. Rep.* **8**, 1–16 (2018).
 21. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell

- Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).
22. Mostafavi, S. *et al.* A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819 (2018).
23. Morabito, S., Miyoshi, E., Michael, N. & Swarup, V. Integrative genomics approach identifies conserved transcriptomic networks in Alzheimer's disease. *Hum. Mol. Genet.* **00**, 1–21 (2020).
24. Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep.* **32**, 108189 (2020).
25. Stuart, T., Srivastava, A., Lareau, C. & Satija, R. Multimodal single-cell chromatin analysis with Signac. *bioRxiv* 2020.11.09.373613 (2020).
26. Gayoso, A. & Shor, J. JonathanShor/DoubletDetection: doubletdetection v3.0. (2020). doi:10.5281/ZENODO.4359992
27. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doubts in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).
28. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329-337.e4 (2019).
29. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
30. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
31. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat.*

- Methods* **14**, 979–982 (2017).
32. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arxiv* (2018).
 33. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).
 34. Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873–1887.e17 (2019).
 35. Liu, J. *et al.* Jointly defining cell types from multiple single-cell datasets using LIGER. *Nat. Protoc.* **15**, 3632–3662 (2020).
 36. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
 37. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
 38. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer’s Disease. *Cell* **169**, 1276–1290.e17 (2017).
 39. Finak, G. *et al.* MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 1–13 (2015).
 40. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
 41. Mclean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions HHS Public Access Author manuscript. *Nat Biotechnol* **28**, 495–501 (2010).

42. Mitra, R. & Maclean, A. L. RVAgene : Generative modeling of gene expression time series data. *bioRxiv* 1–17 (2020).
43. Gao, C. & Welch, J. D. Iterative refinement of cellular identity from single-cell data using online learning. *bioRxiv* (2020). doi:10.1101/978-3-030-45257-5_24
44. Ulirsch, J. C. *et al.* Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019).
45. Chen, W. *et al.* Fine mapping causal variants with an approximate bayesian method using marginal test statistics. *Genetics* **200**, 719–736 (2015).
46. Wang, J. *et al.* CAUSALdb: A database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic Acids Res.* **48**, D807–D816 (2020).
47. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
48. Patro, R. *et al.* Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* **14**, 417–419 (2017).