nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00894-z

Check for updates

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

Samuel Morabito[1,2,5], Emily Miyoshi[2,3,5], Neethu Michael[2,3,5], Saba Shahin[2,3], Alessandra Cadete Martini[3,4], Elizabeth Head[3,4], Justine Silva[3], Kelsey Leavy[3], Mari Perez-Rosendahl[3,4] and Vivek Swarup[2,3]✉

The gene-regulatory landscape of the brain is highly dynamic in health and disease, coordinating a menagerie of biological processes across distinct cell types. Here, we present a multi-omic single-nucleus study of 191,890 nuclei in late-stage Alzheimer's disease (AD), accessible through our web portal, profiling chromatin accessibility and gene expression in the same biological samples and uncovering vast cellular heterogeneity. We identified cell-type-specific, disease-associated candidate *cis*-regulatory elements and their candidate target genes, including an oligodendrocyte-associated regulatory module containing links to *APOE* and *CLU*. We describe *cis*-regulatory relationships in specific cell types at a subset of AD risk loci defined by genome-wide association studies, demonstrating the utility of this multi-omic single-nucleus approach. Trajectory analysis of glial populations identified disease-relevant transcription factors, such as SREBF1, and their regulatory targets. Finally, we introduce single-nucleus consensus weighted gene coexpression analysis, a coexpression network analysis strategy robust to sparse single-cell data, and perform a systems-level analysis of the AD transcriptome.

The human brain is composed of multiple heterogeneous subsets of cells; both neuronal and nonneuronal cells work in concert to perform simple and higher-order tasks. Recent studies have provided more precise molecular characterization and identification of neuronal and nonneuronal cell populations in the cognitively normal brain[1–4]. However, our understanding of heterogeneous cell populations within the diseased brain is still largely limited, hindering our understanding of the biological processes underlying disease. Neurodegenerative disorders, such as AD, are marked with massive neuronal loss, accompanied by gliosis, and the role of specific neuronal and glial cell populations in AD pathophysiology remains unclear. Several single-cell and single-nucleus RNA-sequencing (snRNA-seq) studies have been performed on both mouse and human tissue to study AD, revealing cell-type-specific transcriptional changes[5–9], but the regulators of these disease-associated cell subtypes have yet to be defined.

Moreover, a slew of genetic studies has been performed on AD, identifying multiple associated genetic risk variants[10–16]. Genome-wide association studies (GWAS) of complex diseases such as AD show that a substantial proportion of genetic risk from common variants partitions to distal regulatory elements, which are often cell-type-specific regions in disease-relevant tissues. While much work has gone into intersecting GWAS signals with functional genomics assays, including bulk-tissue RNA-seq and assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)[17], the resolution of such studies is noticeably limited by cell-type heterogeneity. A prerequisite for linking GWAS hits to cell types is a map that links distal regulatory elements with their target genes.

ATAC-seq profiles the open-chromatin regions within a tissue and has recently been adapted for single-cell resolution[18]. To date, single-cell chromatin accessibility techniques, such as single-nucleus ATAC-seq (snATAC-seq) have been seldom used in primary samples of diseased tissues, with only two published studies of single-cell chromatin accessibility in the cognitively normal human brain[19,20]. Therefore, we performed snATAC-seq and snRNA-seq in the same AD postmortem human brain tissue samples to define AD-associated gene-regulatory programs at the epigenomic and transcriptomic levels, providing a powerful lens into the cellular heterogeneity of the brain and allowing us to unravel new biological pathways underlying neurodegeneration in specific cell populations.

Here, we present a multi-omic analysis of 191,890 nuclei from postmortem human brain tissue of AD and cognitively healthy controls at the single-nucleus resolution, in which we directly integrated snRNA-seq and snATAC-seq datasets, thus providing a more complete understanding of the molecular changes in AD. We identified cell-type-specific candidate *cis*-regulatory elements (cCREs) based on chromatin accessibility and found disease-associated cell-subpopulation-specific transcriptomic changes. We identified transcription factors (TFs) that may regulate AD gene expression changes. Further, we applied pseudotime trajectory analysis on our integrated dataset to extensively characterize disease-associated glial cell states at the epigenomic and transcriptomic levels, expanding on previous work exploring gene expression in diverse glial subtypes. We integrated fine-mapped GWAS signals at selected AD risk loci with our snATAC-seq data to link AD risk signals to the specific cell types in which they are accessible and defined the *cis*-regulatory

[1]Mathematical, Computational and Systems Biology (MCSB) Program, University of California, Irvine, CA, USA. [2]Institute for Memory Impairments and Neurological Disorders (MIND), University of California, Irvine, CA, USA. [3]Department of Neurobiology and Behavior, University of California, Irvine, CA, USA. [4]Department of Pathology and Laboratory Medicine, University of California, Irvine, CA, USA. [5]These authors contributed equally: Samuel Morabito, Emily Miyoshi, Neethu Michael. ✉e-mail: vswarup@uci.edu

chromatin accessibility networks at these loci. Moreover, as network analysis has been effective at clarifying disease transcriptomic signatures in tissue-level RNA-seq data, we designed a coexpression network analysis pipeline, integrating single-cell and bulk-tissue RNA-seq datasets, that robustly identified AD-associated coexpression networks within each cell type. Altogether, we have clarified the gene-regulatory landscape of AD, highlighting the role of glia in AD pathophysiology and identifying several genes, namely *SREBF1* in oligodendrocytes, for further study in the context of AD. Finally, we provide an online interface for exploration of these datasets (https://swaruplab.bio.uci.edu/singlenucleiAD).

Results

Multi-omic analysis of the human prefrontal cortex. We performed both snATAC-seq (10x Genomics; $n=12$ late-stage AD; $n=8$ control) and snRNA-seq (10x Genomics v3; $n=11$ late-stage AD; $n=7$ control) on nuclei isolated from the prefrontal cortex (PFC) using postmortem human tissue from individuals with late-stage AD and age-matched cognitively healthy controls (74–90+ years old; Fig. 1a). We defined late-stage AD and controls based on both Braak and plaque staging (Supplementary Tables 1 and 2). We specifically aimed to generate both transcriptomic and epigenetic data from the same tissue sample (aliquots of samples from the same dissection; Methods) to minimize differences in cell-type composition between the two methods, thus allowing for meaningful downstream integrated analysis. After quality-control filtering, we retained a total of 130,418 nuclei for snATAC-seq and 61,472 nuclei for snRNA-seq (Supplementary Figs. 1 and 2, Supplementary Table 3 and Supplementary Note). To ensure the rigor of our study, we applied batch correction methods to the data from both assays, as library preparation limitations required multiple batches. For snATAC-seq, we used mutual nearest neighbors (MNN)[21] to correct the latent semantic indexing reduced chromatin accessibility matrix, and for snRNA-seq, we used integrative non-negative matrix factorization (iNMF)[22] to reduce dimensionality while simultaneously eliminating batch effects (Methods, Extended Data Fig. 1 and Supplementary Note). We applied uniform manifold approximation and projection (UMAP)[23] dimensionality reduction and Leiden clustering[24] to the batch-corrected epigenomic and transcriptomic datasets, identifying distinct cell-type clusters in snATAC-seq (35) and snRNA-seq (34; Fig. 1b,c). With snATAC-seq, we profiled all major cell types of the brain—excitatory neurons (24,076 nuclei; EX.a–e), inhibitory neurons (9,644 nuclei; INH.a–d), astrocytes (15,399 nuclei; ASC.a–f), microglia (12,232 nuclei; MG.a–e), oligodendrocytes (62,253 nuclei; ODC.a–m) and oligodendrocyte progenitor cells (4,869 nuclei; OPC.a)—annotated based on chromatin accessibility at the promoter regions of known marker genes (Fig. 1d and Extended Data Fig. 2). We used chromVAR[25] to compute TF motif variability in single nuclei by estimating the enrichment of TF binding motifs in accessible chromatin regions (Methods) and examined the enrichment of TF motifs by cell type with respect to diagnosis, identifying several TF motifs with increased enrichment with disease in astrocytes, excitatory neurons and microglia (Supplementary Fig. 3 and Supplementary Data 1). Moreover, we performed TF footprinting analysis to further clarify cell-type-specific TF regulation, highlighting the SOX9 TF footprint in oligodendrocytes. Interestingly, we noticed TF motif enrichment of oligodendrocyte-related TFs in excitatory neurons. Likewise, we detected similar cell types using snRNA-seq—excitatory neurons (6,369 nuclei; EX1–5), inhibitory neurons (5,962 nuclei; INH1–4), astrocytes (4,756 nuclei; ASC1–4), microglia (4,126 nuclei; MG1–3), oligodendrocytes (37,052 nuclei; ODC1–13) and oligodendrocyte progenitor cells (2,740 nuclei; OPC1–2)—classified by the gene expression of cell-type markers (Fig. 1e). In both assays, oligodendrocytes were the most commonly profiled cell type (Supplementary Fig. 3). Additionally, while many differentially expressed genes (DEGs) in each major cell type agreed with previous literature, we also found cluster-specific genes previously established as neuronal or glial subtype markers, such as *LINC00507* for L2–3 excitatory neurons (EX1)[4], *SV2C* for L3 interneurons (INH4)[1] and *CX3CR1* for homeostatic microglia (MG2)[26] (Supplementary Fig. 3 and 4 and Supplementary Data 1).

Because the epigenomic landscape is deeply intertwined with downstream gene expression signatures, we integrated our snATAC-seq and snRNA-seq datasets using Seurat's integration platform[27,28] (Methods, Fig. 1f, Extended Data Fig. 3 and Supplementary Fig. 3). Cell types that were independently classified using chromatin data or transcriptome data overwhelmingly grouped together in the integrated UMAP space (Fig. 1g and Supplementary Fig. 3). Using the same biological samples in snATAC-seq and snRNA-seq resulted in a high degree of overlap between nuclei from these two data modalities in the jointly constructed space. Additionally, we confirmed cell-type identities by gene activity and gene expression in a panel of canonical cell-type marker genes (Supplementary Fig. 3) and used Seurat's label-transfer algorithm to verify cell-type annotations in the snATAC-seq dataset using the snRNA-seq dataset as a reference (Supplementary Fig. 5).

Multi-omic characterization of cellular heterogeneity in Alzheimer's disease. In both snATAC-seq and snRNA-seq, we discovered multiple neuronal and glial subpopulations, and we annotated the subpopulations from snRNA-seq based on previously identified marker genes[1,4] (Fig. 2, Supplementary Figs. 6 and 7 and Supplementary Note). For our snATAC-seq clusters, we used Seurat's label-transfer algorithm to calculate cluster prediction scores allowing for supervised annotation of our cell clusters, in which we mapped EX.a to EX1 and ASC.b to ASC2, for example (Supplementary Figs. 6 and 7). We examined the composition of each cluster in the context of disease and found several that were significantly overrepresented or underrepresented in late-stage AD compared to control samples, in both data modalities (Fig. 2d–g and Methods). ASC3 ($GFAP^{high}/CHI3L^{+}$) significantly increased in proportion with disease (bootstrapped cluster proportion analysis using a two-sided Wilcoxon rank-sum test, false discovery rate (FDR) $=8.63\times10^{-5}$), whereas ASC4 ($GFAP^{low}/WIF1^{+}/ADAMTS17^{+}$) significantly decreased (FDR $=4.68\times10^{-7}$), consistent with a recent snRNA-seq study of the 5XFAD mouse model of AD[29]. We also found that the proportions of MG.a. and MG.b were increased in late-stage AD (FDR $=9.82\times10^{-7}$ and 8.88×10^{-10}, respectively), both of which mapped to the activated snRNA-seq cluster MG1 ($SPP1^{high}/CD163^{+}$), which was also increased with disease (FDR $=6.32\times10^{-7}$). Additionally, we found that immune oligodendrocyte cluster ODC13 was significantly increased in late-stage AD (FDR $=1.62\times10^{-4}$).

Further, we identified both differentially accessible chromatin regions and DEGs in late-stage AD for each cell cluster and found high cluster specificity for Gene Ontology (GO) term enrichment of distal and proximal differentially accessible chromatin regions, as well as DEGs (Supplementary Figs. 7–9, Supplementary Data 1–6 and Supplementary Note). For example, we identified *NEAT1* as upregulated in astrocytes and oligodendrocytes, in agreement with previous findings in the entorhinal cortex[7], and we confirmed AD upregulation of *NEAT1* with in situ hybridization (Extended Data Fig. 4). Altogether, we found cluster-specific epigenetic and transcriptomic changes in late-stage AD, which may underlie the dysregulation of distinct biological pathways in different cell subpopulations in neurodegeneration.

Cell-type-specific *cis*-gene regulation in late-stage Alzheimer's disease. Based on our experimental design using both snATAC-seq and snRNA-seq in the same samples, we reasoned that we could identify the target genes of cCREs in specific cell populations (Extended

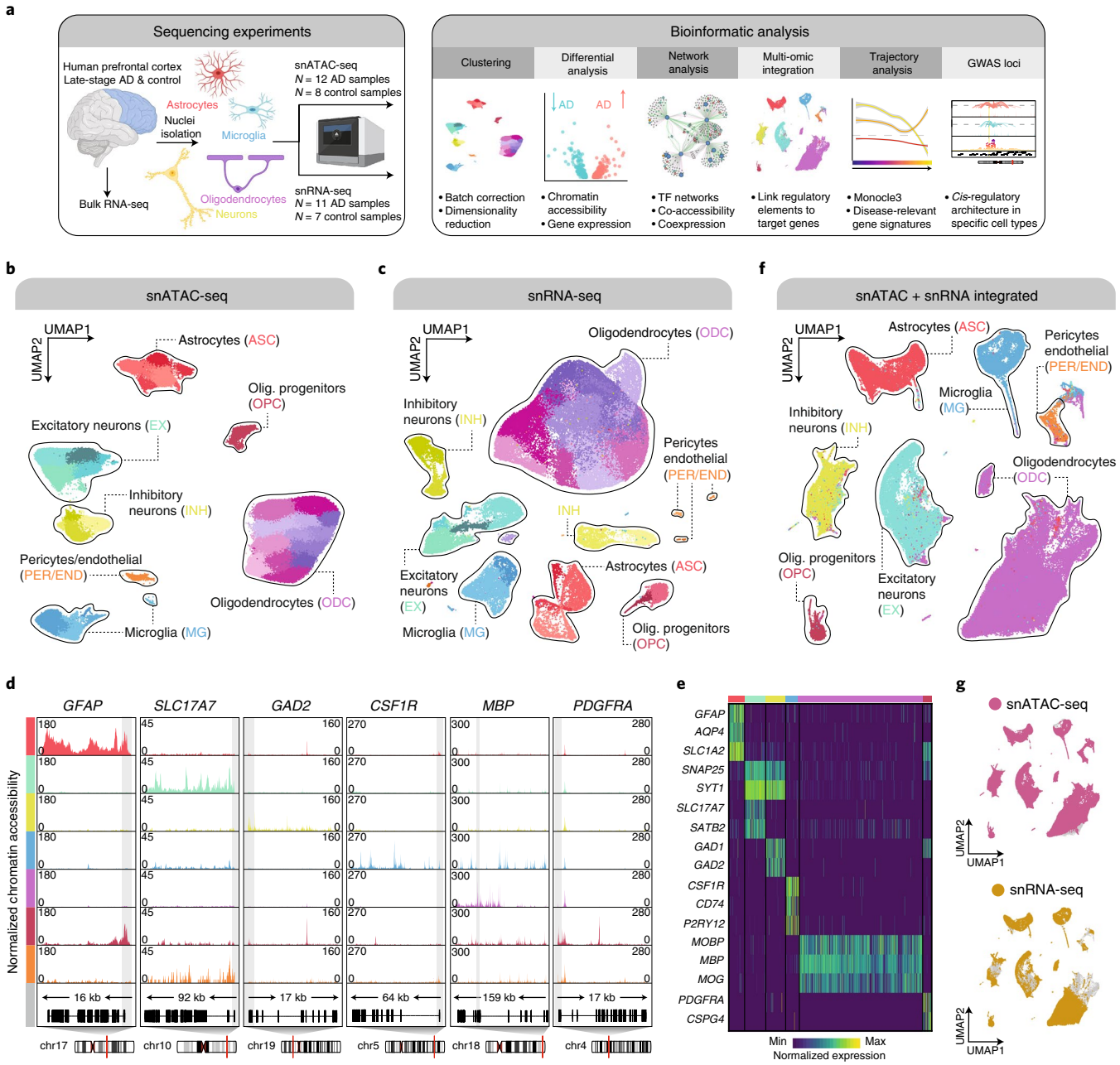


Fig. 1 | Single-nucleus ATAC-seq and single-nucleus RNA-seq for the study of cellular diversity in the diseased brain. a, Schematic representation of the samples used in this study, sequencing experiments and downstream bioinformatic analyses, created with BioRender.com. **b**,**c**, UMAP visualizations where dots correspond to individual nuclei for 130,418 nuclei profiled with snATAC-seq (**b**) and 61,472 nuclei profiled with snRNA-seq (**c**), colored by Leiden cluster assignment and cell type. ASC, astrocytes; EX, excitatory neurons; INH, inhibitory neurons; MG, microglia; ODC, oligodendrocytes; OPC, oligodendrocyte progenitor cells; PER/END, pericytes/endothelial cells. **d**, Pseudo-bulk chromatin accessibility profiles for each cell type at canonical cell-type marker genes. For each gene, 1 kb upstream and downstream are shown. The promoter/transcription start site (TSS) is highlighted in gray with gene model and chromosome position shown below. Chromosome coordinates are: *GFAP* chr17: 44904008–44919937; *SLC17A7* chr19: 49428401–49445360; *GAD2* chr10: 26213307–26305558; *CSF1R* chr5: 150052291–150116372; *MBP* chr18: 76977827–77136683; *PDGFRA* chr4: 54226097–54299247. **e**, Row-normalized single-nucleus gene expression heatmap of cell-type marker genes. **f**, UMAP plot of 186,167 nuclei from a jointly learned subspace of snATAC-seq and snRNA-seq, colored by cell-type assignment. **g**, Integrated UMAP as in **f**, colored by originating dataset. Smaller gray dots represent nuclei from the other data modality. A consistent coloring scheme for each cell type and cluster is used throughout the paper.

Data Fig. 5a and Methods). To this end, we sought to elucidate the *cis*-regulatory architecture of the PFC in late-stage AD by constructing *cis* co-accessibility networks[30] (CCANs) separately for late-stage AD and control samples in each cell type (Methods). To identify target genes of cCREs, we focused on the subset of co-accessible peaks where one of the peaks lies in a promoter element, yielding a set of cCREs and candidate target genes. For this set of co-accessible links, we correlated the expression of the candidate target gene to the chromatin accessibility of the cCRE, strengthening the evidence of a potential regulatory relationship beyond co-accessibility alone. Finally, we used NMF to analyze and cluster these gene-linked cCREs (gl-cCREs) based on their chromatin accessibility in each cell

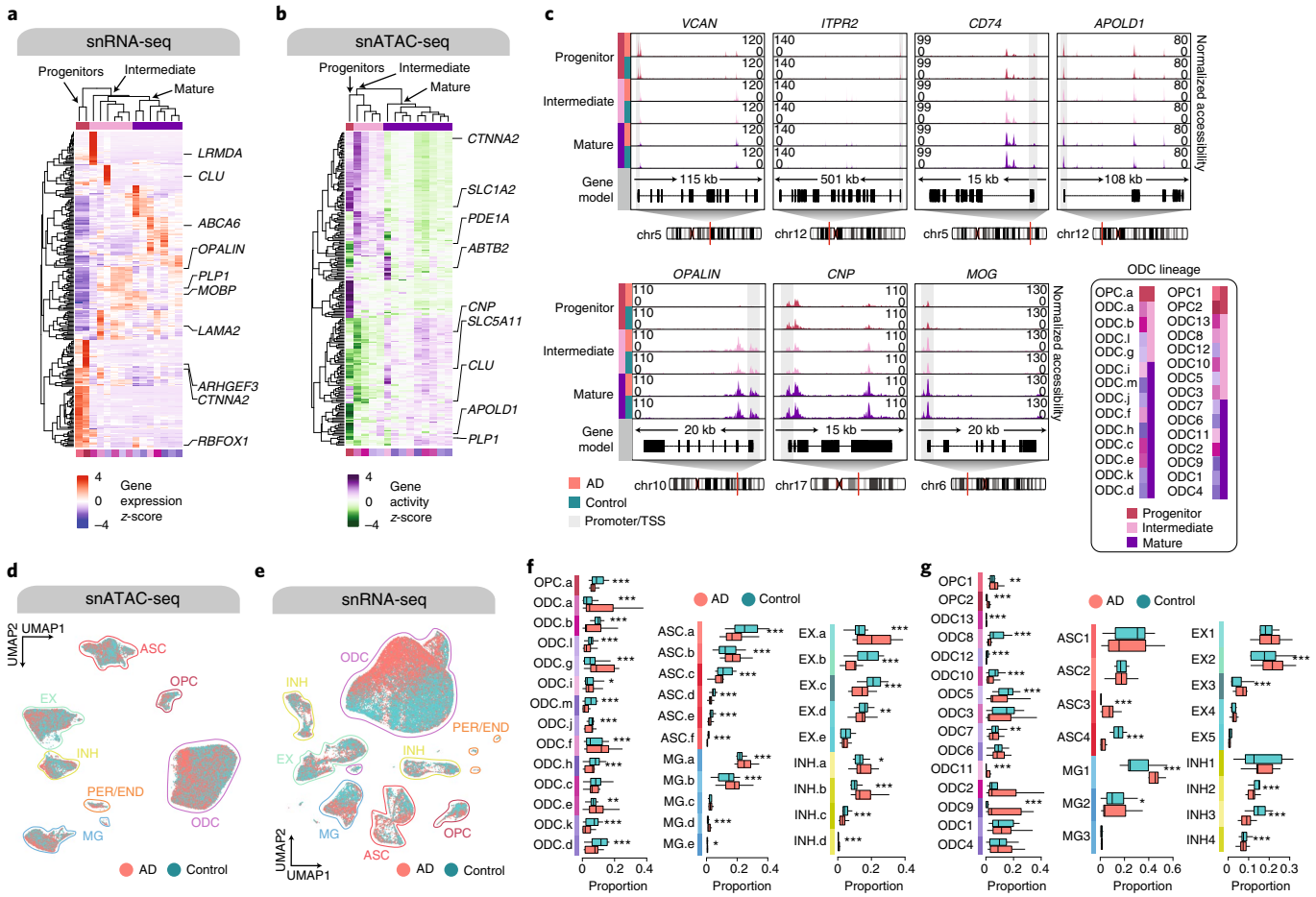


Fig. 2 | Epigenetically and transcriptionally distinct cell subpopulations in human AD prefrontal cortex. a,**b**, Hierarchically clustered heatmaps of row-normalized gene expression in snRNA-seq OPC and oligodendrocyte clusters (**a**) and gene activity in snATAC-seq OPC and oligodendrocyte clusters (**b**) for the top 25 upregulated DEGs (sorted by average log fold change) identified in each oligodendrocyte subpopulation. **c**, Pseudo-bulk chromatin accessibility coverage profiles for OPC (progenitor), intermediate oligodendrocyte and mature oligodendrocyte snATAC-seq clusters; assignments as in **b**. Promoter/TSS regions highlighted in gray with gene model and chromosome position shown below. Chromosome coordinates are: *VCAN* chr5: 83468465–83583303; *ITPR2* chr12: 26335515–26836198; *CD74* chr5: 150400637–150415929; *APOLD1* chr12: 12722917–12830975; *OPALIN* chr10: 96342216–96362365; *CNP* chr17: 41963741–41978731; *MOG* chr6: 29653981–29673372. **d**,**e**, snATAC-seq (**d**) and snRNA-seq (**e**) UMAPs as in Fig. 1, where nuclei are colored by AD diagnosis. Clusters annotated by cell type. **f**,**g**, Box-and-whisker plots showing the proportion of nuclei mapping to each cluster for each sample, split by control and late-stage AD samples for snATAC-seq (**f**) and snRNA-seq (**g**) clusters, with measures of significance from bootstrapped cluster composition analysis (Wilcoxon test, Methods; *** FDR ≤ 0.001, ** FDR ≤ 0.01, * 0.01 < FDR ≤ 0.05) and *n* as per Supplementary Tables 7–9. For box-and-whisker plots, the 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.

cluster. In sum, this approach results in a set of candidate enhancer elements (gl-cCREs) grouped into functional modules, as well as a set of cCRE-linked genes, for each major cell type in late-stage AD and control samples.

In total, using this approach, we identified 56,552 gl-cCREs and 11,440 cCRE-linked genes, with a median of 4 cCREs linked to each of these genes (Fig. 3a and Supplementary Tables 4 and 5). By examining the overlap between sets of cCRE-linked genes identified in each cell type, we observed a substantial number of genes with linked cCREs that were shared across multiple cell types, in addition to those that are cell-type specific (Fig. 3b). For several cell types, we found a significant overlap between the set of cCRE-linked genes and cell-type marker DEGs, as well as genes that are upregulated in AD within that cell type, highlighting a critical role of cCREs in disease-related gene expression changes (Fig. 3c). We also investigated the chromatin accessibility in each snATAC-seq cluster for these gl-cCREs and noted a high degree of cell-type and cluster specificity (Fig. 3d). The majority of the gl-cCREs mapped to intronic regions (58.35%; Fig. 3e). Moreover, by inspecting the NMF coefficient matrix (***H***), we were able to identify which cluster or cell type each NMF module corresponds to, and we annotated several modules that are specific to control or late-stage AD nuclei within a given cluster (Fig. 3f,g and Supplementary Note). Additionally, we found that some of the cCRE target genes that are common to more than one cell type are regulated by different cCREs in each cell type.

Cell-type-specific transcription factors in late-stage Alzheimer's disease. To complement our analysis of *cis*-regulatory elements, we sought to identify cell-type-specific *trans*-regulatory elements in late-stage AD. TFs tightly control cell fate in neurodevelopment and have been implicated in neurodegenerative processes. We examined the regulatory role of microglial TF SPI1 (also known as PU.1) and nuclear respiratory factor 1 (NRF1) in oligodendrocytes (Fig. 4a–f, Supplementary Fig. 10 and Supplementary Note). SPI1 motif variability in our snATAC-seq microglia clusters was significantly increased in only upregulated clusters MG.a and MG.b,

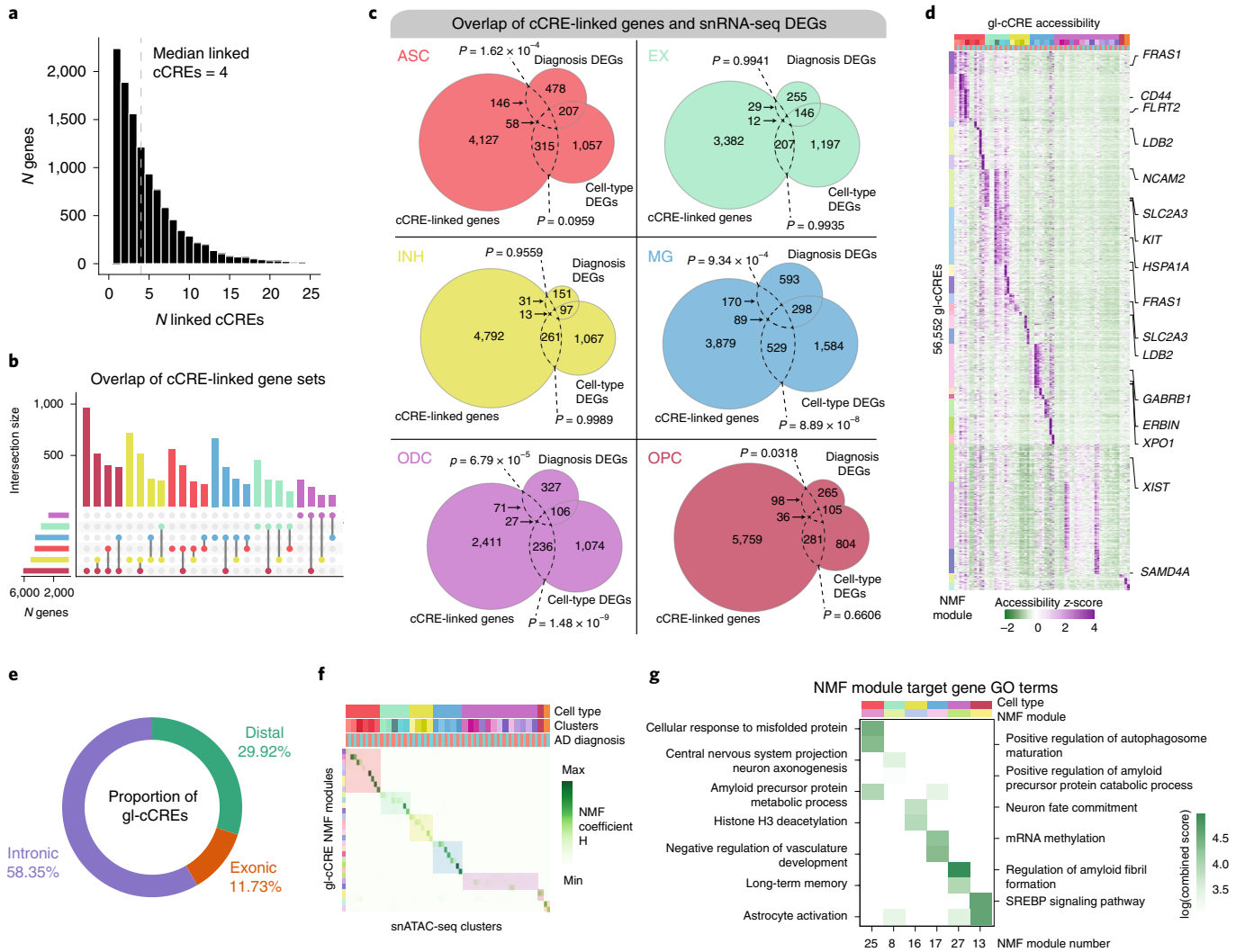


Fig. 3 | Linking *cis*-regulatory elements to downstream target genes in specific cell types. a, Histogram showing the number of genes that have 1 through 25 linked cCREs. **b**, Upset plot showing the size of overlaps between the sets of cCRE-linked genes identified in each cell type. The bar plot on the left shows the set size of cCRE-linked genes for each cell type, and the bar plot on the top shows the number of overlapping genes between two sets, or the number of unique genes in one set. **c**, Venn diagrams for each major cell type showing the overlaps between the set of cCRE-linked genes and genes upregulated in that cell type (cell-type DEGs) and genes upregulated in AD within this cell type (diagnosis DEGs). A one-sided Fisher's exact test was used for gene-set overlap significance (***$P \leq 0.001$, **$P \leq 0.01$, *$P < 0.05$). **d**, Heatmap showing row-normalized pseudo-bulk chromatin accessibility in each snATAC-seq cluster split by nuclei from control and late-stage AD samples. Rows (cCREs) are organized based on NMF module assignment. Annotations correspond to genes from differential gene expression analysis that were upregulated in AD in at least one cell type. **e**, Donut chart showing the percentage of gl-cCREs that mapped to intronic, exonic or distal regions. **f**, Heat map showing NMF coefficients in each snATAC-seq cluster split by nuclei from control and late-stage AD samples. **g**, Heatmap showing log-transformed enrichR combined scores for GO terms for gene sets of selected NMF modules.

but SPI1's targets were significantly downregulated in only MG1 (Fig. 4a,b and Supplementary Fig. 10). We also identified NRF1 as dysregulated in select oligodendrocyte clusters (Fig. 4d–f and Supplementary Fig. 10). These results indicate that SPI1 acts as a transcriptional repressor in late-stage AD, providing insight into how SPI1 contributes to AD pathophysiology. Additionally, NRF1 has previously been associated with mitochondrial function[31], and impaired mitochondrial function, mediated by NRF1 dysregulation, may contribute to neuronal dysfunction in late-stage AD through the disruption of myelination. TF analyses in neuronal populations and Fos-related antigen 2 (FOSL2) in astrocytes are shown in Extended Data Fig. 4 and Supplementary Fig. 10.

To gain further insight into TF-mediated gene regulation in late-stage AD, we constructed cell-type-specific TF regulatory networks. For a given TF, we identified candidate target genes as those whose promoters or linked cCREs are accessible and contain the TF's binding motif in the cell type of interest, and we repeated this for several select TFs, generating microglia-specific and oligodendrocyte-specific TF regulatory networks (Fig. 4g,h, Extended Data Fig. 5b and Supplementary Note). Within these networks, we identified multiple AD DEGs, in addition to genes located at known AD GWAS loci, regulated by SPI1 in microglia and NRF1 in oligodendrocytes.

Integrated trajectory analysis of disease-associated glia. To further uncover molecular mechanisms driving glial heterogeneity in AD, we performed pseudotime trajectory analysis using Monocle3 (refs. [32–34]) on the integrated snATAC-seq and snRNA-seq data in oligodendrocytes, microglia and astrocytes (Supplementary Note). Multi-omic trajectory analysis allows us to investigate the dynamics of gene

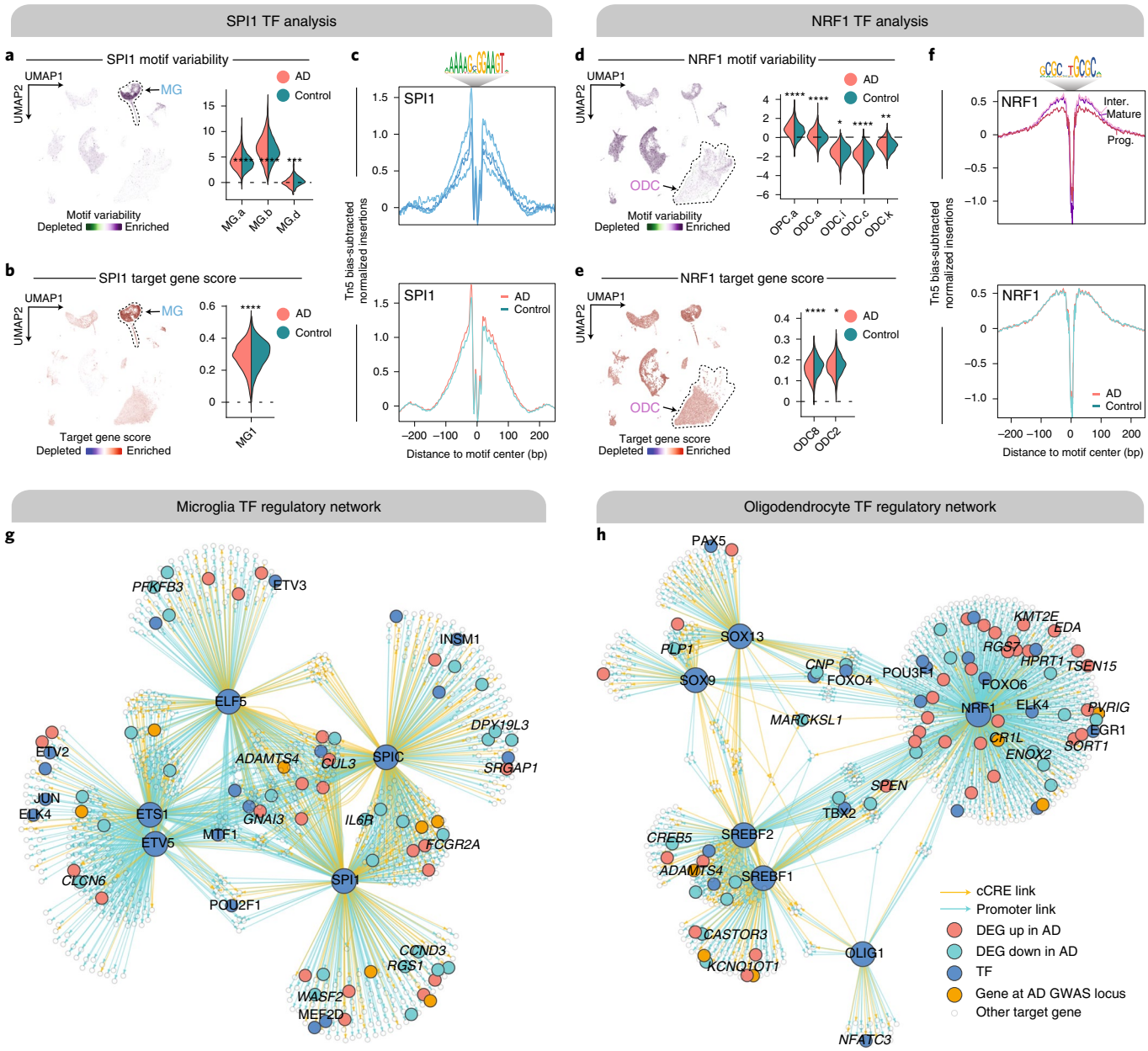


Fig. 4 | Cell-subpopulation-specific transcription factor regulation in late-stage AD. a, Left: snATAC-seq and snRNA-seq integrated UMAP colored by SPI1 motif variability with microglia outlined. Right: violin plots of SPI1 motif variability in snATAC-seq microglia clusters that had a significant difference between control and late-stage AD, split by diagnosis. **b**, Left: Integrated UMAP colored by SPI1 target gene score with microglia circled. Right: violin plots of SPI1 target gene score in snRNA-seq microglia clusters that had a significant difference between control and late-stage AD, split by diagnosis as in **a**. **c**, Tn5 bias-subtracted TF footprinting for SPI1 by snATAC-seq microglia cluster (top) and by AD diagnosis (bottom). TF binding motif shown as motif logo above. **d**, Left: integrated UMAP colored by NRF1 motif variability with oligodendrocytes outlined. Right: violin plots of NRF1 motif variability in snATAC-seq oligodendrocyte clusters that had a significant difference between control and late-stage AD, split by diagnosis as in **a**. **e**, Left: integrated UMAP colored by NRF1 target gene score with oligodendrocyte outlined. Right: violin plots of NRF1 target gene score in snRNA-seq oligodendrocyte clusters that had a significant difference between control and late-stage AD, split by diagnosis as in **a**. **f**, Tn5 bias-subtracted TF footprinting for NRF1 by snATAC-seq oligodendrocyte cluster (top) and by AD diagnosis (bottom) as in **c**. **g**,**h**, TF regulatory networks showing the predicted candidate target genes for the following TFs: ELF5, ETS1, ETV5, SPIC and SPI1 in microglia (**g**); SOX9, SOX13, SREBF1, SREBF2, OLIG1 and NRF1 in oligodendrocytes (**h**). For violin plots, two-sided Wilcoxon test was used to compare control versus AD samples; not significant, $P > 0.05$; *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, ****$P \leq 0.0001$.

expression, chromatin accessibility and TF motif variability throughout a continuum of cell-state transitions. We modeled gene expression and chromatin accessibility dynamics using a recurrent variational autoencoder (RVAE)[35]. Briefly, a RVAE is an encoder–decoder neural network framework that uses long short-term memory units to effectively model temporal biological data, yielding a two-dimensional (2D) latent representation of the input features, as well as a de-noised reconstructed version of the original input (Supplementary Note). For each cell type, we identified genes that were differentially expressed along the trajectory (t-DEGs; Supplementary Data 7) and used these genes as features to train the RVAE until the loss function converged (Supplementary Note and Extended Data Fig. 6).

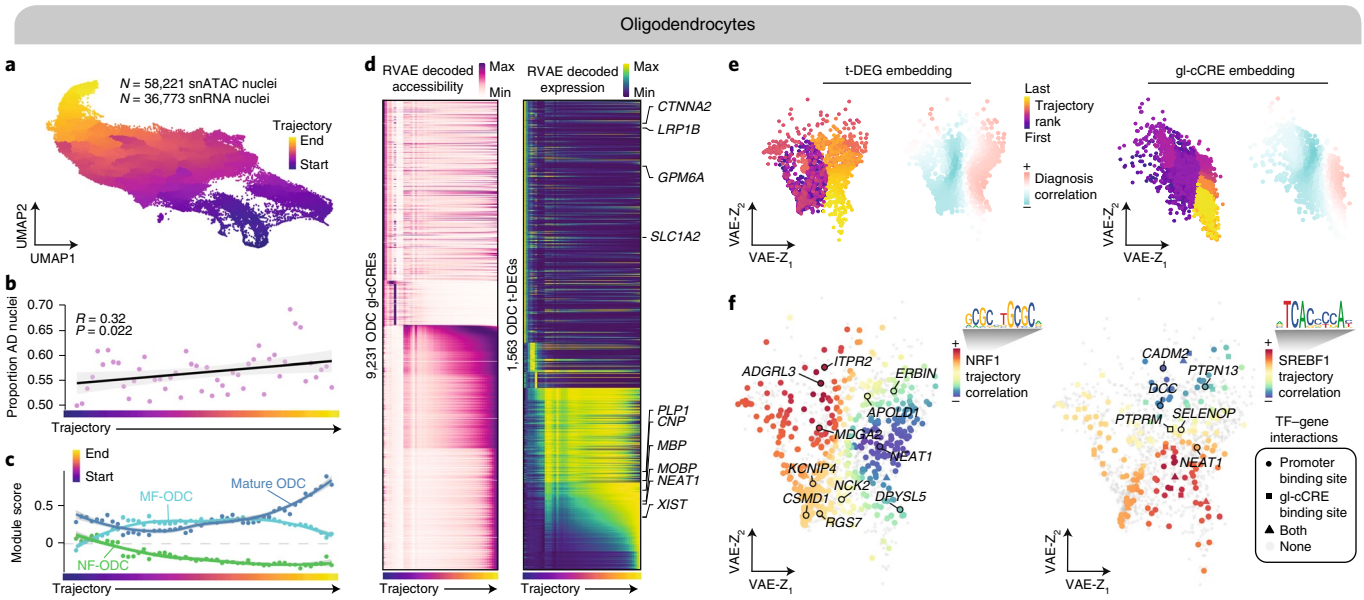


Fig. 5 | Multi-omic oligodendrocyte trajectory analysis. a, UMAP dimensionality reduction of oligodendrocytes from the integrated snATAC-seq ($n = 58,221$ nuclei) and snRNA-seq ($n = 36,773$ nuclei) analysis. Each cell is colored by its pseudotime trajectory assignment. **b**, Scatterplot showing the proportion of oligodendrocyte nuclei from AD samples at 50 evenly sized bins across the trajectory. The black line shows a linear regression, and the gray outline represents the 95% CI. Pearson correlation coefficient and *P* value from two-sided test are shown. **c**, Scatterplot of module scores for NF-ODC, MF-ODC and mature ODC gene signatures[36,37] (see Supplementary Note for full gene lists) averaged for nuclei in each of the 50 trajectory bins. Solid colored lines represent loess regressions for each signature, and the gray outlines represent 95% CIs. **d**, Left: heatmap of chromatin accessibility at 9,231 oligodendrocyte gl-cCREs reconstructed using RVAE. Right: heatmap of gene expression for 1,563 oligodendrocyte t-DEGs reconstructed using RVAE. Annotated genes are DEGs in oligodendrocytes, with respect to other cell types, or AD upregulated genes in oligodendrocytes. **e**, 2D latent space learned by RVAE modeling of oligodendrocyte t-DEGs (left) and gl-cCREs (right), where each dot represents one gene. Left: genes colored by trajectory rank, the point in the trajectory where the gene reaches 75% of max expression. Right: genes colored by correlation of RVAE reconstructed expression with AD diagnosis proportion as in **b**. **f**, Oligodendrocyte t-DEG latent space colored by correlation of reconstructed gene expression to NRF1 (left) and SREBF1 (right) motif variability. The shape of each point represents the regulatory relationship between the TF and each gene, while genes without regulatory evidence are shown as small gray dots. Annotated genes are AD upregulated genes in oligodendrocytes (AD DEGs). TF binding motifs are shown as motif logos.

Oligodendrocyte trajectory reveals SREBF1 dysregulation. We constructed an integrated oligodendrocyte trajectory using 58,221 nuclei from snATAC-seq and 36,773 nuclei from snRNA-seq (Fig. 5a), noting that the proportion of nuclei from late-stage AD samples appeared to increase along the trajectory (Fig. 5b; Pearson correlation $R = 0.32$, P value $= 0.022$). To clarify the functional state of oligodendrocytes associated with late-stage AD, we examined the gene expression signatures[36,37] of newly formed oligodendrocytes (NF-ODCs), myelin-forming oligodendrocytes (MF-ODCs) and mature oligodendrocytes (mature ODCs; Fig. 5c; see Supplementary Note for gene signature lists). Interestingly, we found that the mature ODC gene expression signature increased at the end of the trajectory, whereas the MF-ODC gene signature decreased. In addition, the NF-ODC gene signature decreased throughout the trajectory, altogether suggesting that the oligodendrocyte pseudotime trajectory recapitulates oligodendrocyte maturation. Chromatin accessibility of 9,231 oligodendrocyte gl-cCREs and gene expression of 1,563 oligodendrocyte t-DEGs reconstructed with a RVAE showcases the vast amount of chromatin remodeling and transcriptional reprogramming that may underlie oligodendrocyte maturation (Fig. 5d).

Additionally, the latent feature space ($\boldsymbol{Z}$) learned by the RVAE provides further biological insight into the pseudotime trajectory and gene regulation in disease (Fig. 5e). Here, each dot represents a single feature (gene or chromatin region), and they are organized in 2D space based on their pseudotemporal dynamics learned by the RVAE. We ranked each feature based on the point in the trajectory that it reached 75% of its maximum value, which we termed as the feature's 'trajectory rank'. We then correlated the reconstructed feature trajectories, as in Fig. 5d, to the proportion of late-stage AD nuclei, as in Fig. 5b, to see which features consistently changed with AD. For both genes (t-DEGs) and chromatin regions (gl-cCREs), the latent space clearly groups features together that are positively or negatively correlated with the proportion of late-stage AD nuclei and groups features together with similar trajectory ranks, demonstrating the power of this RVAE model for the analysis and interpretation of multi-omic pseudotemporal dynamics.

We showcase two key TFs in oligodendrocytes: NRF1 and sterol regulatory element binding transcription factor 1 (SREBF1). SREBF1 is critical in regulating the expression of genes involved in cholesterol and fatty acid homeostasis[38], and it is proposed that Aβ inhibits SREBF1 activation[39]. We found that NRF1 motif variability is upregulated in oligodendrocytes in late-stage AD (Bonferroni-adjusted P value $= 5.13 \times 10^{-20}$; Fig. 4d), and SREBF1 motif variability is downregulated with disease in oligodendrocytes (Bonferroni-adjusted P value $= 2.67 \times 10^{-191}$; Extended Data Fig. 4). We correlated TF motif variability trajectories (Extended Data Fig. 6) with the reconstructed t-DEG expression trajectories and visualized the correlation between the TF and each gene within the 2D latent space, identifying candidate target genes activated or repressed by TF binding events (positive or negative trajectory correlation, respectively; Fig. 5f and Supplementary Note). We found that NRF1 was negatively correlated with target genes at the end of the trajectory, while SREBF1 was positively correlated with target genes at both the beginning and the end of the trajectory, indicating that SREBF1 acts as a transcriptional activator throughout the trajectory.

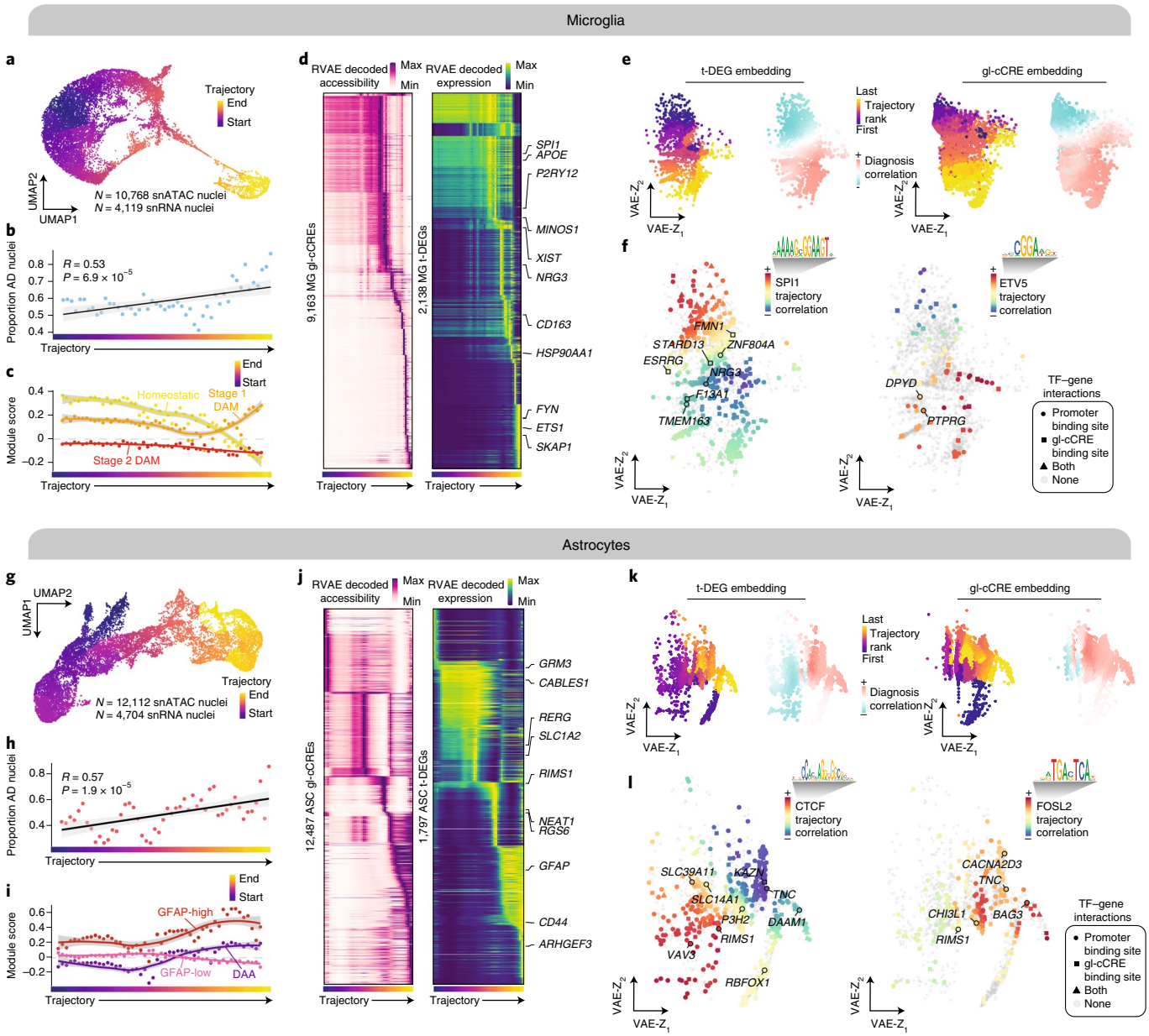


Fig. 6 | Multi-omic microglia and astrocyte trajectory analyses. a, UMAP dimensionality reduction of microglia from the integrated snATAC-seq ($n = 10{,}768$ nuclei) and snRNA-seq ($n = 4{,}119$ nuclei) analysis. **b**, Scatterplot of the proportion of AD microglia nuclei as in Fig. 5b. **c**, Scatterplot of module scores as in Fig. 5c for gene signatures from Keren-Shaul et al.[40]: homeostatic microglia, stage 1 DAM and stage 2 DAM (see Supplementary Note for full gene lists). **d**, Heatmaps of RVAE reconstructed chromatin accessibility and gene expression as in Fig. 5d, for 9,163 microglia gl-cCREs (left) and 2,138 microglia t-DEGs (right). **e**, 2D latent space learned by RVAE modeling of microglia t-DEGs (left) and gl-cCREs (right), as in Fig. 5e. **f**, Microglia t-DEG latent space colored by correlation of gene expression to SPI1 (left) and ETV5 (right) motif variability, as in Fig. 5f. **g**, UMAP dimensionality reduction of astrocytes from the integrated snATAC-seq ($n = 12{,}112$ nuclei) and snRNA-seq ($n = 4{,}704$ nuclei) analysis. **h**, Scatterplot of the proportion of AD astrocyte nuclei as in **b**. **i**, Scatterplot of module scores as in **c** for gene signatures from Habib et al.[29]: GFAP-low, GFAP-high and DAAs (see Supplementary Note for full gene lists). **j**, Heatmaps of RVAE reconstructed chromatin accessibility and gene expression as in **d** for 12,487 astrocyte gl-cCREs (left) and 1,797 astrocyte t-DEGs (right). **k**, 2D latent space learned by RVAE modeling of astrocyte t-DEGs (left) and gl-cCREs (right), as in **e**. **l**, Astrocyte t-DEG latent space colored by correlation of gene expression to CTCF (left) and ETV5 (right) motif variability, as in **f**.

Microglia trajectory to define disease-associated microglia. Using the same analytical approach as that used for our oligodendrocyte trajectory analysis, we constructed an integrated microglia trajectory using 10,768 nuclei from snATAC-seq and 4,119 nuclei from snRNA-seq (Fig. 6a). The proportion of nuclei from late-stage AD samples significantly increased throughout the microglia trajectory (Fig. 6b; Pearson correlation $R = 0.53$, P value $= 6.9 \times 10^{-5}$). We next sought to investigate gene signatures of disease-associated microglia (DAM), which were introduced in a single-cell transcriptomic study[40] of 5XFAD mice and are highly debated in the field of AD genomics. DAM are described as AD-associated phagocytic microglia that are sequentially activated in *TREM2*-independent and *TREM2*-dependent stages (stage 1 and stage 2, respectively). We found that the integrated microglia trajectory follows a decrease in the homeostatic signature, an increase in the stage 1 DAM signature and a distinct global depletion of the stage 2 *TREM2*-dependent

DAM signature (Fig. 6c; see Supplementary Note for gene signature lists), suggesting that this microglia trajectory describes the transcriptional and epigenetic changes during the transition from a homeostatic to disease-associated cell state.

To further dissect the microglia trajectory, we modeled the chromatin accessibility and gene expression dynamics of 9,163 microglia gl-cCREs and 2,138 microglia t-DEGs, respectively, using RVAE (Fig. 6d,e). We highlight two ETS family TFs, SPI1 and ETS variant 5 (ETV5), both of which showed upregulated motif variability in late-stage AD (Bonferroni-adjusted P values 1.19×10^{-20} and 6.68×10^{-19}, respectively), and their candidate target genes along the trajectory (Fig. 6f and Supplementary Note). We observed that the SPI1 motif trajectory was negatively correlated with genes at the end of the trajectory, supporting our previous findings that SPI1 acts as a repressor in late-stage AD.

Disease-associated astrocytes in human Alzheimer's disease. We also constructed an integrated astrocyte trajectory using 12,112 nuclei from snATAC-seq and 4,704 nuclei from snRNA-seq (Fig. 6g), and we again found that the proportion of late-stage AD nuclei significantly increased throughout the trajectory (Fig. 6h; Pearson correlation $R=0.57$, P value $=1.9\times10^{-5}$). In a similar fashion to our analysis of the DAM signature in the microglia trajectory, we investigated the gene signature of disease-associated astrocytes (DAAs), described in a recent snRNA-seq study of the hippocampus in 5XFAD mice[29] as an AD-specific *GFAP*^{high} astrocyte subpopulation that is distinct from another *GFAP*^{high} astrocyte subpopulation found in aged wild-type and 5XFAD mice (*GFAP*-high). Based on DAA gene signature analysis, we reasoned that this trajectory follows a trend from a *GFAP*-low state to *GFAP*-high and DAA-like states (Fig. 6i; see Supplementary Note for gene signature lists).

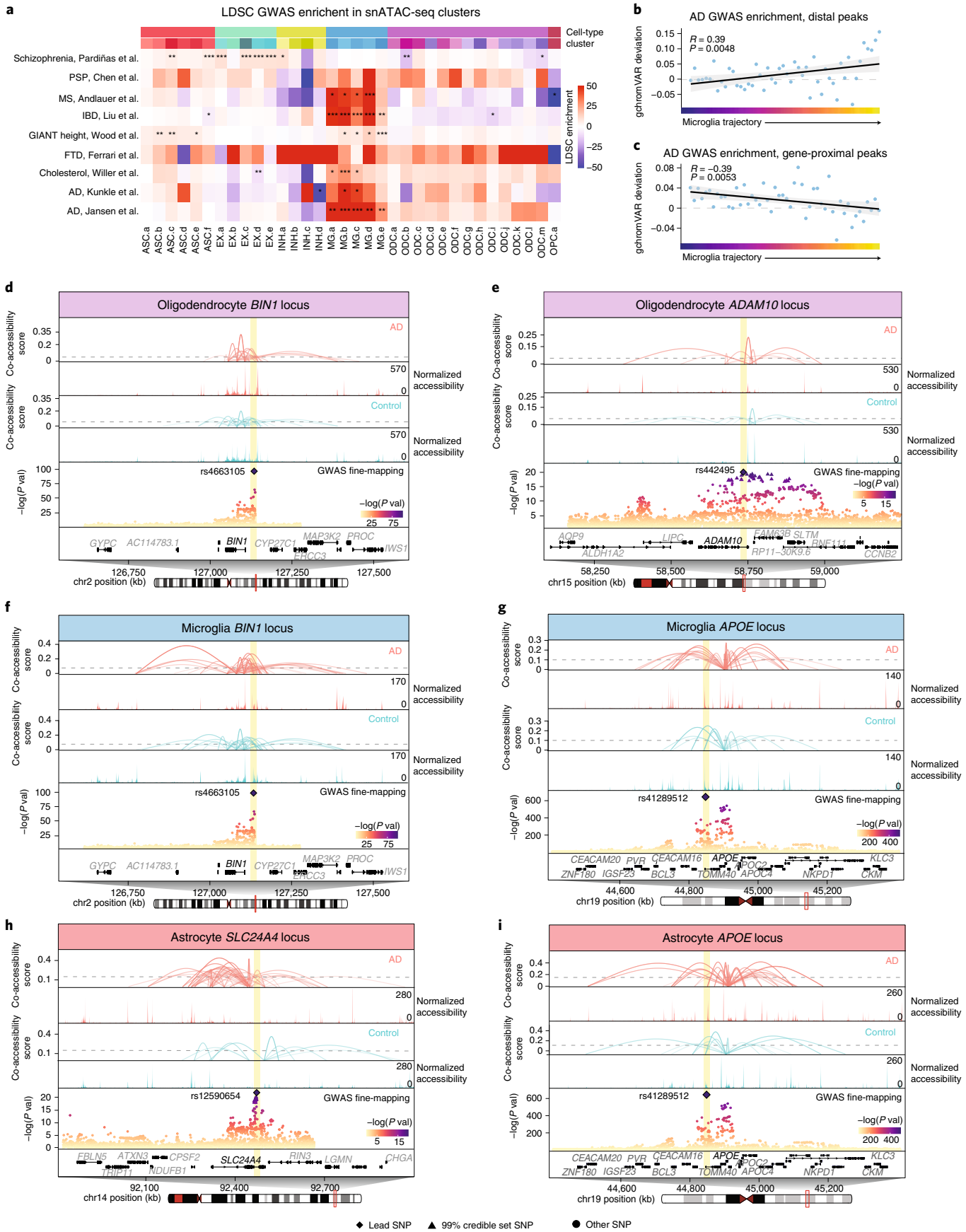
RVAE modeling of 12,487 astrocyte gl-cCREs and 1,797 astrocyte t-DEGs revealed rich gene-regulatory dynamics across the trajectory (Fig. 6j,k). We investigated the relationship between astrocyte t-DEGs and two TFs: CCCTC-binding factor (CTCF) and FOSL2, whose motif variability we have found to be downregulated and upregulated in late-stage AD, respectively (Bonferroni-adjusted P values 6.45×10^{-17} and 5.65×10^{-99}, respectively). CTCF is known as a master chromatin regulator[41,42], and we observed that the CTCF motif variability trajectory was anti-correlated with the DAA and *GFAP*-high signatures (end of the trajectory; Extended Data Fig. 6) and positively correlated with t-DEGs in the *GFAP*-low phase of the trajectory (Fig. 6l). Alternatively, we found a positive correlation between the motif variability trajectory of FOSL2 with the *GFAP*-high and DAA gene signatures and a positive correlation with genes at the end of the trajectory (Fig. 6l and Supplementary Note). These findings suggest that FOSL2 may be an activator of the DAA signature, whereas CTCF may promote a more homeostatic or non-diseased astrocyte state. By relating gene expression with TF motif enrichment, TF binding site accessibility and using the temporal information learned by the RVAE, we begin to unravel the role of TFs in regulating cell states, such as DAAs.

Cell-type-specific *cis*-regulation at Alzheimer's disease genetic risk loci. To further our understanding of AD genetic risk signals, we performed cell-type-specific linkage-disequilibrium score regression[43] (LDSC) analysis in our snATAC-seq clusters using GWAS summary statistics in AD[10,11] and other relevant traits[44–50] (Methods, Supplementary Table 6 and Supplementary Note). Microglia clusters MG.b and MG.c showed a significant enrichment (FDR < 0.05) for AD GWAS SNPs from the study by Kunkle et al., and all five microglia clusters showed a significant enrichment (MG.a and MG.e FDR < 0.005; MG.b, MG.c and MG.d FDR < 0.0005) for GWAS SNPs from the study by Jansen et al., which included familial AD-by-proxy samples in addition to data from patients with AD (Fig. 7a). The results of this GWAS heritability analysis supports previous findings in non-diseased humans[20] and mice[51] snATAC-seq data. We further investigated AD risk signals in microglia using gchromVAR[52] to compute the enrichment of fine-mapped AD-associated polymorphisms from Jansen et al. along the microglia pseudotime trajectory and observed a significant increase (Pearson correlation, P value = 0.0048; Methods and Fig. 7b,c) in the gchromVAR deviation score in distal peaks throughout the microglia trajectory, in stark contrast with a significant decrease in the deviation score for the analogous gene-proximal peak analysis (Pearson correlation, P value = 0.0053), highlighting AD-associated SNPs at distal enhancers in DAM. By overlaying the co-accessibility map with chromatin accessibility signal and GWAS statistics along the genomic axis, we unraveled the potential *cis*-regulatory relationships disrupted by causal disease variants in GWAS genes, such as *BIN1*, *ADAM10*, *APOE* and *SLC24A4* (Fig. 7d–i). We found that the *APOE* locus, which harbors the main determinants of AD heritability and is one of the best studied AD risk loci, has *cis*-regulatory chromatin networks altered in disease in microglia and astrocytes, highlighting cCREs that are prime candidates for further study using genome-editing technologies.

Single-cell coexpression networks using scWGNCA. To recontextualize snRNA-seq data in a systems-level framework, we sought to develop a gene coexpression network analysis approach for single-cell data based on weighted gene coexpression analysis (WGCNA)[53,54], a powerful analytical approach for identifying disease-associated gene modules[55,56] originally designed for bulk gene expression data. Our revised approach uses aggregated expression profiles in place of potentially sparse single cells, where 'metacells' are constructed from specific cell populations by computing the mean expression from 50 neighboring cells using k-nearest neighbors (Methods, Extended Data Fig. 7 and Supplementary Note). We reprocessed published AD snRNA-seq data from Mathys et al.[5] and used iNMF to integrate these with our snRNA-seq data (Methods and Extended Data Fig. 8). Additionally, we performed bulk RNA-seq in early-stage and late-stage AD samples, as well as pathological controls, and curated additional AD bulk-tissue RNA-seq samples from the Religious Orders Study and Memory and Aging Project (ROSMAP) study[57]. Finally, we used consensus WGCNA[58],

Fig. 7 | Cell-type-specific regulatory landscapes of GWAS loci in the AD brain. a, Heatmap showing LDSC enrichment of GWAS traits and disorders in snATAC-seq clusters. P values were derived from LDSC enrichment tests, and FDR-corrected P values are overlaid on the heatmap (*FDR < 0.05, **FDR < 0.005, ***FDR < 0.0005). PSP, progressive supranuclear palsy; MS, multiple sclerosis; IBD, inflammatory bowel disease; FTD, frontotemporal dementia. **b**,**c**, Scatterplots showing gchromVAR enrichments along the microglia pseudotime trajectory in distal peaks (**b**) and gene-proximal peaks (**c**) averaged for nuclei in each of the 50 trajectory bins. The black line shows a linear regression, and the gray outline represents the 95% CI. Pearson correlation coefficients and P values are shown. **d–i**, *Cis*-regulatory architecture at the following GWAS loci and cell types: *BIN1* (**d**) and *ADAM10* (**e**) in oligodendrocytes; *BIN1* (**f**) and *APOE* (**g**) in microglia; *SLC24A4* (**h**) and *APOE* (**i**) in astrocytes. Co-accessible links for late-stage AD and control samples are shown separately, with the line height and opacity corresponding to the co-accessibility score; links with a score below the gray dashed line were removed for visualization purposes. Genomic coverage plots for AD and control samples are shown separately. AD GWAS statistics from Jansen et al. for SNPs at each locus are shown. Lead SNPs are shown as diamonds, and SNPs in 99% credible set are shown as triangles. Chromosome ideogram indicates genomic coordinates in a 500-kb radius centered at each GWAS gene. Chromosome coordinates are: *BIN1* chr2: 127047027–127110355; *ADAM10* chr15: 58587807–58752978; *APOE* chr19: 44902754–44910393; *SLC24A4* chr14: 92319581–92502483.

a meta-analytical approach, to jointly form coexpression 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 Figs. 1, 7, 9 and 10 and Supplementary Data 7), performed iteratively for each cell type, where each edge



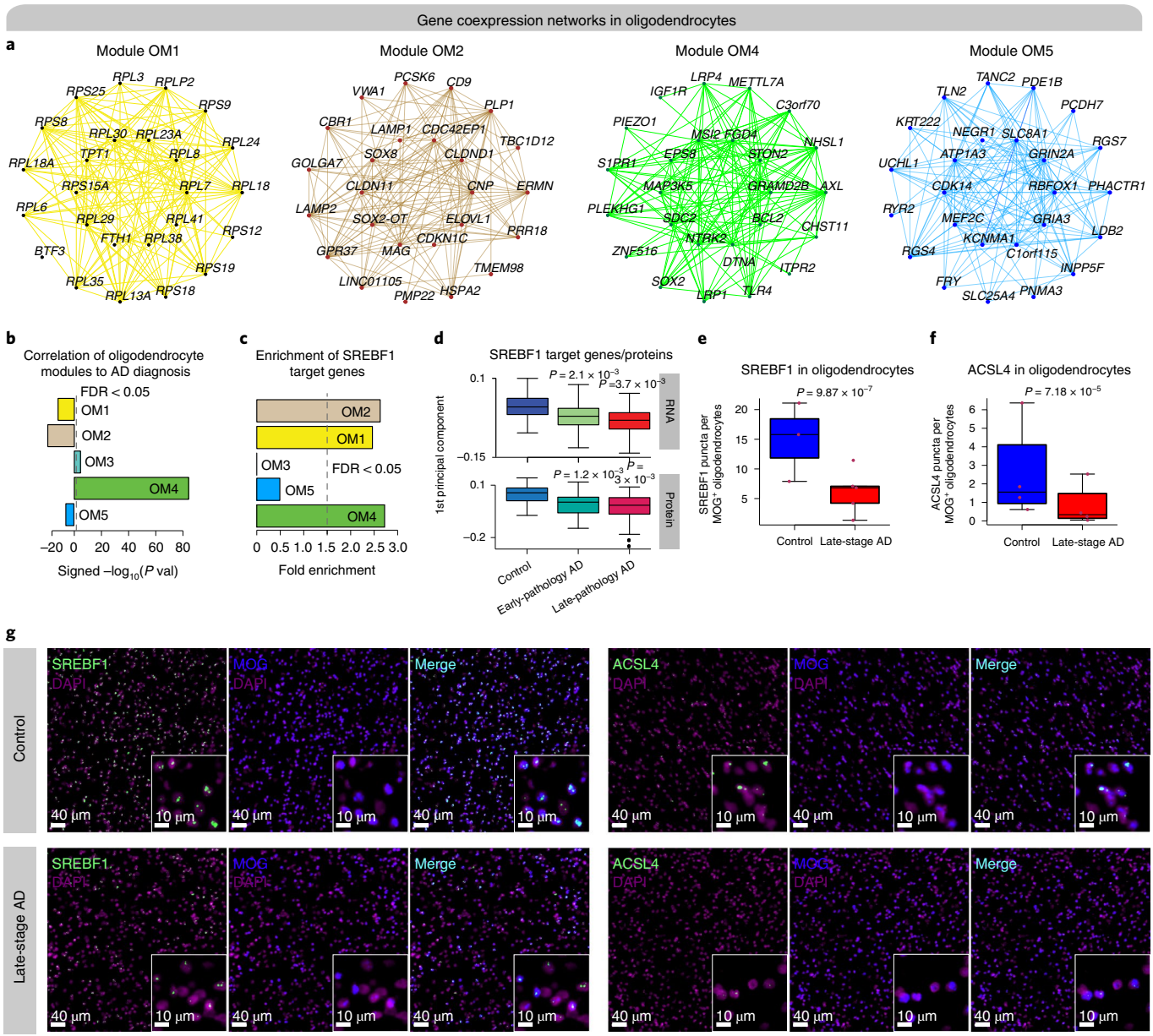


Fig. 8 | Robust gene expression modules revealed using integrated bulk and single-cell coexpression network analysis. a, Coexpression plots for modules OM1, OM2, OM4 and OM5. **b**, Signed correlation oligodendrocyte coexpression modules with AD diagnosis. **c**, Enrichment of SREBF1 target genes in oligodendrocyte coexpression modules. **d**, Boxplots showing RNA (top) and protein expression[59] (bottom; $n = 98$ controls, 76 early pathology and 101 late pathology) of the targets of SREBF1 with AD pathological staging. Two-sided Wilcoxon test. **e**, Boxplots showing quantification of *SREBF1* puncta per *MOG*$^+$ oligodendrocyte; $n = 3$ cognitively healthy controls, 5 late-stage AD. Data are represented as the mean of four equally sized regions per sample. Linear mixed-effects model. **f**, Boxplots showing quantification of *ACSL4* puncta per *MOG*$^+$ oligodendrocyte; $n = 4$ cognitively healthy controls, 4 late-stage AD. Data are represented as the mean of four equally sized regions per sample. Linear mixed-effects model. **g**, Representative RNA FISH (RNAscope) images from postmortem human brain tissue for combined *SREBF1* and *MOG* staining as in **e** (left) and *ACSL4* and *MOG* staining as in **f** (right) with DAPI nuclear counterstain. For box-and-whisker plots, the box boundaries and line correspond to the 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.

in a coexpressed module is supported by both bulk-tissue RNA-seq data (this study and ROSMAP[57]) and aggregated snRNA-seq data (this study and Mathys et al.).

We specifically highlight our scWGCNA analysis for oligodendrocytes; we found four coexpression modules significantly correlated with AD diagnosis—OM1, OM2, OM4 and OM5 (Fig. 8a,b and Supplementary Data 7). For example, hub genes of the AD-downregulated module OM1 encoded ribosomal subunits (*RPS15A*, *RPL30* and *RPL23A*, for example), consistent with its enrichment of GO terms related to protein synthesis and sorting (Supplementary Fig. 11). OM2 gene members *MAG*, *CNP* and *PLP1* are known to be involved in myelination, and unsurprisingly, we found OM2 downregulated with disease.

Additionally, we examined the downstream regulatory targets of SREBF1 in the context of coexpression networks (Methods). Notably, we found that three of the oligodendrocyte modules were significantly enriched for targets of SREBF1, indicating the importance of SREBF1 in regulating gene expression in these modules (Fig. 8c).

Using multi-scale data from bulk-tissue RNA-seq, high-throughput proteomics[59] and SREBF1 chromatin immunoprecipitation (ChIP) with sequencing (ChIP–seq; ENCODE), we defined a protein–protein interaction (PPI) network of SREBF1 target genes (Extended Data Fig. 7). Additionally, we found module eigengene expression of SREBF1 targets downregulated in early-stage and late-stage AD samples at the level of proteins[59] and RNA (Fig. 8d), corroborated by downregulation of SREBF1 motif variability in snATAC-seq data (Extended Data Fig. 4). We also validated the downregulation of SREBF1 in late-stage AD through RNA in situ hybridization and immunohistochemistry and found a decrease in *ACSL4* expression, one of SREBF1's targets identified in ENCODE ChIP–seq data, in late-stage AD (Fig. 8e–g and Extended Data Fig. 7). Overall, our coexpression network analysis approach facilitates the identification of cell-type-specific disease biology, and we have highlighted TF SREBF1, largely unstudied in the context of AD, in oligodendrocytes to demonstrate our approach's ability to yield new disease insights.

Discussion

Our integrated multi-omic analysis of late-stage AD provides a unique lens into the continuum of cellular heterogeneity underlying disease pathogenesis. Pinpointing causal mechanisms of complex diseases requires a rigorous understanding of cell-population-specific gene-regulatory systems at both the epigenomic and transcriptomic levels. While single-cell chromatin accessibility can provide important insights into disease, it is a challenging data modality to work with due to its inherent sparsity. We circumvented the issue of sparsity by integrating single-nucleus open-chromatin and single-nucleus transcriptomes from the same samples, in addition to using aggregation methods for pseudo-bulk accessibility profiling and co-accessibility analysis. Taking these considerations into account, our multi-omic analysis enabled us to analyze cell-type-specific epigenomic dysregulation in neurodegeneration and expands on previous work to decipher the transcriptomes of single nuclei in human AD.

A major contribution of our study is that we identified cell-type-specific gl-cCREs, which may mediate gene-regulatory changes in late-stage AD, along with TFs that may bind to these gl-cCREs within the given cell type. While cCREs can be identified with epigenetic data alone, our analysis is substantiated by integrating single-nucleus transcriptomic data, as we link the gene expression of candidate target genes with cCRE chromatin accessibility. Previous studies of AD have not explored *cis*-gene regulation at a cell-type or cell-subpopulation level. We have highlighted both *cis*-gene and *trans*-gene regulation disrupted in late-stage AD, providing potential targets for further study into AD, such as NRF1 in oligodendrocytes and FOSL2 in astrocytes and their corresponding gl-cCREs. Further, we examined *cis*-regulatory interactions in our multi-omic dataset to elucidate cell-type and disease-specific patterns of genes implicated in inherited AD risk by GWAS, which are of particular interest as candidate therapeutic targets. For a subset of AD GWAS loci, we compared *cis*-regulatory networks between AD and control cell populations to identify interactions that are uniquely found in disease. Thus, this study serves as a resource for the broader AD community to explore cell-type and cell-state-specific regulatory landscapes of genes and genomic regions that may be of particular interest, such as AD GWAS loci.

Moreover, independent and joint analyses of the transcriptome and chromatin profiles of oligodendrocytes revealed disrupted gene regulation and biological pathways in AD (Supplementary Note). We described a multi-omic oligodendrocyte trajectory and evaluated gene expression signatures in the transition from newly formed to mature oligodendrocytes, observing that the trajectory seemed to follow oligodendrocyte maturation. Notably, we analyzed the trajectory dynamics of SREBF1, a TF involved in the regulation of cholesterol and lipid metabolism that has been shown to be involved in Aβ-related processes[39]. We found that SREBF1 motif variability was decreased in late-stage AD, indicating that fewer SREBF1 binding sites are accessible in disease, and *SREBF1* gene expression is also downregulated in AD oligodendrocytes. Trajectory analysis revealed that SREBF1 motif variability is positively correlated with t-DEGs throughout the trajectory, suggesting that it acts as a transcriptional activator in oligodendrocytes.

Coexpression network analysis methods like WGCNA have been widely used for discovery of disease-associated gene modules in bulk gene expression data[60,61]; however, these approaches are rarely used in single-cell transcriptomics, with some exceptions[62] due to challenges in network construction from noisy data. Here we introduced scWGCNA, a method for interrogation of cell-population-specific coexpression networks that leverages aggregated metacells to combat the sparsity of single-cell gene expression. Using scWGCNA, we found gene coexpression networks in human AD by jointly analyzing our snRNA-seq and bulk RNA-seq data with additional snRNA-seq and bulk RNA-seq samples from the ROSMAP cohort[5,57]. This meta-analytical approach ensured the robustness of our network analysis and allowed us to evaluate the resulting gene modules in early-stage AD (Supplementary Note). Notably, scWGCNA identified three oligodendrocyte modules that were enriched for target genes of SREBF1 and showed that the gene and protein expression of these targets were decreased in late-stage AD. With our coexpression and trajectory analysis of SREBF1 in oligodendrocytes, SREBF1 is clearly a gene to prioritize for follow-up studies as a candidate target for AD therapeutics, demonstrating the utility of our approach in identifying new gene targets for disease.

While the causative molecular mechanisms of sporadic AD remain unknown, our work offers new insights that assist in unraveling the nature of gene regulation in AD, especially regarding genomic loci with well-described heritable disease risk. Additional work is needed to spatially resolve the complexity of gene expression and epigenomics in AD and neurodegeneration in general. The data presented here are a valuable resource for understanding regulatory relationships in the diseased brain, and our analysis framework serves as a blueprint for making discoveries in complex traits using single-cell multi-omic data. Finally, our intuitive web portal for exploring single nuclei in the human brain allows for the accessibility of our results to anyone with an internet-equipped device.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00894-z.

Received: 26 June 2020; Accepted: 2 June 2021;
Published online: 08 July 2021

References

1. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
2. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
3. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
4. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
5. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
6. 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).

7. Grubman, A. et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* **22**, 2087–2097 (2019).
8. Leng, K. et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* **24**, 276–287 (2021).
9. Del-Aguila, J. L. et al. A single-nuclei RNA-sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimers. Res. Ther.* https://doi.org/10.1186/s13195-019-0524-x (2019).
10. 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).
11. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
12. Carrasquillo, M. M. et al. Genetic variation in *PCDH11X* is associated with susceptibility to late-onset Alzheimer's disease. *Nat. Genet.* **41**, 192–198 (2009).
13. Harold, D. et al. Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. *Nat. Genet.* **41**, 1088–1093 (2009).
14. Hollingworth, P. et al. Common variants at *ABCA7*, *MS4A6A/MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease. *Nat. Genet.* **43**, 429–436 (2011).
15. Hibar, D. P. et al. Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
16. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
17. Nott, A. et al. Brain cell-type-specific enhancer–promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
18. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
19. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
20. Corces, M. R. et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).
21. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
22. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887 (2019).
23. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
24. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
25. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. ChromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
26. Masuda, T. et al. Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* **566**, 388–392 (2019).
27. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
28. 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).
29. Habib, N. et al. Disease-associated astrocytes in Alzheimer's disease and aging. *Nat. Neurosci.* https://doi.org/10.1038/s41593-020-0624-8 (2020).
30. Pliner, H. A. et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871 (2018).
31. Satoh, J. I., Kawana, N. & Yamamoto, Y. Pathway analysis of ChIP–seq-based NRF1 target genes suggests a logical hypothesis of their involvement in the pathogenesis of neurodegenerative diseases. *Gene Regul. Syst. Bio.* **2013**, 139–152 (2013).
32. 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).
33. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
34. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
35. Mitra, R. & Maclean, A. L. RVAgene: generative modeling of gene expression time-series data. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.10.375436 (2020).
36. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).
37. Jäkel, S. et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
38. Shimano, H. & Sato, R. SREBP-regulated lipid metabolism: convergent physiology-divergent pathophysiology. *Nat. Rev. Endocrinol.* **13**, 710–730 (2017).
39. Mohamed, A., Viveiros, A., Williams, K. & De Chaves, E. P. Aβ inhibits SREBP-2 activation through Akt inhibition. *J. Lipid Res.* **59**, 1–13 (2018).
40. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290 (2017).
41. Pugacheva, E. M. et al. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl Acad. Sci. USA* **117**, 2020–2031 (2020).
42. Kim, S., Yu, N. K. & Kaang, B. K. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp. Mol. Med.* **47**, e166 (2015).
43. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
44. Ferrari, R. et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* **13**, 686–699 (2014).
45. Chen, J. A. et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* **13**, 41 (2018).
46. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
47. Andlauer, T. F. M. et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci. Adv.* **2**, 1–12 (2016).
48. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
49. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
50. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1285 (2013).
51. Cusanovich, D. A., Hill, A. J., Disteche, C. M., Trapnell, C. & Shendure, J. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324 (2018).
52. Ulirsch, J. C. et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019).
53. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
54. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 1128 (2005).
55. Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
56. Rexach, J. E. et al. Tau pathology drives dementia risk-associated gene networks toward chronic inflammatory states and immunosuppression. *Cell Rep.* **33**, 108398 (2020).
57. 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).
58. Morabito, S., Miyoshi, E., Michael, N. & Swarup, V. Integrative genomics approach identifies conserved transcriptomic networks in Alzheimer's disease. *Hum. Mol. Genet.* **29**, 2899–2919 (2020).
59. Swarup, V. et al. Identification of conserved proteomic networks in neurodegenerative dementia. *Cell Rep.* **31**, 107807 (2020).
60. Swarup, V. et al. Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat. Med.* **25**, 152–164 (2019).
61. Allen, M. et al. Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. *Alzheimer's Dement.* **14**, 352–366 (2018).
62. Wu, Y. E., Pan, L., Zuo, Y., Li, X. & Hong, W. Detecting activated cell populations using single-cell RNA-seq. *Neuron* **96**, 313–329 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Human samples. Human PFC brain samples were obtained from the University of California Irvine Institute for Memory Impairments and Neurological Disorders (UCI MIND) Alzheimer's Disease Research Center (ADRC) tissue repository and under UCI's Institutional Review Board. Postmortem human brain tissue from the ROSMAP study was obtained under the Institutional Review Board of Rush University Medical Center. Informed consent was obtained for all human participants. Samples were dissected, homogenized on a dry-ice pre-chilled isolating platform and aliquoted for snRNA-seq and snATAC-seq. For details on the human samples used in this study (AD, $n=6$ males and 6 females; controls, $n=5$ males and 3 females; all 74–90+ years old), please see Supplementary Tables 1 and 2. ROSMAP RNA-seq data and details can be found on https://www.synapse.org/ under synapse ID syn3219045.

Bulk RNA-seq. Total RNA was isolated from human PFC using Mini Nucleospin RNA kit (740955.250, Takara Bio). RNA integrity was assessed using a 2100 Bioanalyzer (Agilent). Total RNA was quantified using Qubit RNA HS assay kit (Q32852, Invitrogen). Around 500 ng total RNA was used to prepare the cDNA library using SMARTer Stranded Total RNA Sample Prep kit-HI Mammalian (634874, Takara Bio). cDNA library concentration was calculated using Qubit dsDNA HS assay kit (Q32851, Invitrogen). Library quality was assessed using either high-sensitivity DNA assay kit (5067-4626) on a 2100 Bioanalyzer or D5000 HS kit (5067-5589 and 5067-5588) on a 4200 Tapestation (Agilent). Libraries were multiplexed with 96 and 95 samples in two lanes on an Illumina NovaSeq S4 for 100-bp paired-end reads.

Single-nucleus RNA-seq. Single-nucleus suspensions were isolated from ~50 mg frozen human PFC. Samples were homogenized in Nuclei EZ Lysis buffer (NUC101-1KT, Sigma-Aldrich) and incubated for 5 min. Samples were passed through a 70-µm filter and incubated in additional lysis buffer for 5 min and centrifuged at 500*g* for 5 min at 4 °C before two washes in Nuclei Wash and Resuspension buffer (1× PBS, 1% BSA and 0.2 U μl^{-1} RNase inhibitor). Nuclei were FACS sorted with DAPI (NucBlue Fixed Cell ReadyProbe Reagent, R37606, Thermo Fisher) before running on the 10x Chromium Single Cell 3′ v3 platform. cDNA library quantification and quality were assessed as per bulk RNA-seq. Libraries were sequenced using the Illumina NovaSeq 6000 S4 platform at the New York Genome Center, using 100-bp paired-end sequencing.

Single-nucleus ATAC-seq. Single-nucleus suspensions were isolated from ~50 mg frozen human PFC according to the 10x Genomics Nuclei Isolation from Mouse Brain Tissue protocol (CG000212, Rev A) with an additional sucrose purification step. Before resuspending our nuclei in Diluted Nuclei Buffer, we removed cellular debris by preparing a sucrose gradient (Nuclei PURE Prep Nuclei Isolation Kit, NUC201-1KT, Sigma). Nuclei were spun at 13,000*g* for 45 min at 4 °C and then washed once and filtered before running on the 10x Chromium Single Cell ATAC platform. Library quantification and quality checking were performed according to the manufacturer's recommendations. Libraries were sequenced using the Illumina NovaSeq 6000 S4 platform at the UCI Genomics core facility, using 100-bp paired-end sequencing.

RNAscope (FISH). Fresh-frozen human postmortem tissue was sectioned at 20 µm on a cryostat at −20 °C. Slides were stored in airtight containers at −80 °C until use. Immediately after removing from −80 °C, slides were dried for 20 min at room temperature and then fixed in 4% paraformaldehyde/PBS for 15 min at 4 °C. Slides were then washed in RNase-free PBS for 5 min at room temperature three times. For single-labeling experiments, slides were incubated in PBS with an LED light for 96 h at 4 °C to quench autofluorescence[63], and for dual-labeling experiments, autofluorescence was quenched with TrueBlack (Biotium) for 30 s before coverslipping. Slides were processed following the RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (ACD) instructions for fresh-frozen tissue, except protease IV incubation was performed for 15 min. Probes used were NEAT1 (411531), PLP1 (499271), CNP (509131-C2), SREBF1 (469871), ACSL4 (408301), MOG (543181-C2) and AQP4 (482441-C2). Fluorophores used were TSA Plus Cy5 (1:200 dilution; Perkin Elmer) and Opal 570 (1:200 dilution, Perkin Elmer) to avoid autofluorescence. Images were taken on a ZEISS Axio Scan.Z1 at ×20 magnification. Four regions per section were analyzed using QuPath. We used a trained object classifier to identify MOG^+ or $AQP4^+$ nuclei, except for ASCL4/MOG dual staining, which required manual assignment of MOG^+ nuclei due to high background. Subcellular detection was used to count RNA punctae. We used a linear mixed-effects model to account for random effects (age and sex) and fixed effects (multiple regions from the same individual).

Immunofluorescence. Fixed and cryoprotected human postmortem tissue was sectioned at 40 µm using a cryotome (Leica). For 6E10, Iba-1, MAP2 and GFAP, brain sections were treated with 90% formic acid for 4 min. For PDGFRA and Olig2, sections in sodium citrate buffer were heated at 80 °C in a bead bath for 30 min. Sections were then washed before blocking (PBS with 5% goat or donkey normal serum, respective to the antibodies, and 0.2% Triton X-100) for 1 h at room temperature. Primary antibodies were incubated at 4 °C overnight (6E10, 1:1,000 dilution, 803001, BioLegend; Iba-1, 1:1,000 dilution, 019-19741, Wako; MAP2, 1:500 dilution, ab32454, Abcam; GFAP, 1:500 dilution, G3893, Sigma; PDGFRA, 1:50 dilution, AF-307, R&D Systems; Olig2, 1:200 dilution, ab109186, Abcam). Secondary antibodies (goat anti-mouse 555, A-21422; goat anti-rabbit 488, A11034; goat anti-rabbit 488, A11034; goat anti-mouse 555, A-21422; donkey anti-goat 488, A-11055; donkey anti-rabbit 555, A31572; all from Thermo Fisher) were diluted (1:200) and incubated for 1 h. Slides were treated with 0.3% Sudan Black in 70% ethanol for 4 min to reduce autofluorescence and imaged on a confocal microscope (Leica). Images from three randomly selected areas were used for volume analysis of amyloid plaques using IMARIS. We used linear mixed-effects model as previously stated.

Annotation of major cell types. Major cell-type annotations were assigned to UMAP partitions and initial clusters in snATAC-seq and snRNA-seq datasets, respectively, through manual inspection of canonical marker gene signals. 'Pseudo-bulk' chromatin accessibility coverage profiles of gene-body and upstream promoter regions were visualized using the Signac[64] (v0.2.0) function CoveragePlot, while gene expression signals were visualized using Seurat[27,28] (v3.1.2). snRNA-seq cell-type assignments were further validated by integration with the Mathys et al.[5] dataset.

Integrated analysis of snRNA-seq and snATAC-seq data. A unified dataset of both chromatin accessibility and gene expression was constructed using Seurat's integration framework. Canonical correlation analysis was used to generate a shared dimensionality reduction of the 'query' snATAC-seq gene activity and the 'reference' snRNA-seq gene expression. MNNs were then identified in this shared space, effectively identifying pairs of corresponding cells that anchor the two datasets together. To confirm major cell-type annotations in snATAC-seq cell populations, we used Seurat's label-transfer algorithm, which leverages these anchors to predict cell types in snATAC-seq data, with cell-type annotations in snRNA-seq cells as the reference and chromatin accessibility with latent semantic indexing reduction as the weights. We achieved a maximum prediction score ≥ 0.5 in 94% of cells, demonstrating high correspondence between the two data modalities. Next, we used these shared anchors to impute gene expression signals in snATAC-seq data. Following imputation, we merged gene expression in cells from the snRNA-seq dataset with snATAC-seq cells that had a maximum prediction score ≥ 0.5. The merged dataset was then centered, dimensionally reduced with principal-component analysis using 30 dimensions, batch corrected with MNNs (monocle3, v0.2.0) and embedded with UMAP. Clusters and UMAP partitions were identified using Leiden clustering (monocle3). We visualized correspondence of major cell types from their dataset of origin to their joint UMAP partitions using ggalluvial (v0.11.1).

Cell-type-specific dimensionality reduction and cluster analysis. Cell-type-specific analyses were performed for snATAC-seq and snRNA-seq by subsetting each major cell type from the fully processed dataset followed by re-embedding with UMAP. Subpopulations of each cell type used for all downstream analysis were then identified using Leiden clustering (monocle3). Clusters smaller than 100 cells were removed as outliers. We then used the addReproduciblePeakSet function from the R package ArchR (v1.0.0)[65] with default parameters to call accessible chromatin peaks using MACS2 (v2.2.7.1) in each cell-type subcluster. For snRNA-seq and snATAC-seq clusters, we performed a bootstrapped cluster composition analysis to robustly assess the composition of each cluster with respect to AD diagnosis. Over 25 iterations, 20% of cells were sampled from the whole dataset, and the proportion of cells from AD or control samples was computed for each cluster. A two-sided Wilcoxon rank-sum test was used to compare the proportion of AD and control samples in each cluster using the wilcox.test R (v3.6.1) function with default parameters and Benjamini–Hochberg correction for multiple testing.

Annotation of cell subpopulations. snRNA-seq subpopulations for astrocytes, microglia, neurons and oligodendrocyte progenitors were annotated in a similar way to the major cell types, using canonical marker gene signals as well as DEGs. snATAC-seq subpopulations for astrocytes, microglia, neurons and oligodendrocyte progenitors were annotated using Seurat label-transfer prediction scores with the snRNA-seq clusters as a reference annotation. We annotated the snRNA-seq oligodendrocytes by hierarchically clustering oligodendrocyte and oligodendrocyte progenitor clusters based on the gene expression matrix of the top 25 DEGs (by average log fold change) from each oligodendrocyte subpopulation, grouping oligodendrocytes into major lineage classes such as progenitor, intermediate and mature. We used the same approach to annotate the snATAC-seq oligodendrocytes, hierarchically clustering the gene activity matrix using the same DEGs. The R package ComplexHeatmap (v2.7.6.1010)[66] was used for hierarchical clustering and visualization of these gene expression and gene activity matrices.

Single-nucleus transcription factor binding motif analysis. Single-nucleus TF motif enrichment was computed for a set of 452 TFs from the JASPAR 2018 database[67] using the Signac wrapper for chromVAR (v1.12.0)[25]. The motif accessibility matrix was first computed, describing the number of peaks that

contain each TF motif for all cells. chromVAR then uses this motif accessibility matrix to compute deviation *z*-scores for each motif by comparing the number of peaks containing the motif to the expected number of fragments in a background set that accounts for confounding technical factors such as GC content bias, PCR amplification and variable Tn5 tagmentation. To further analyze specific TFs of interest, we used the getFootprints function in ArchR to perform TF footprinting analysis in pseudo-bulk aggregates of single nuclei in the same cluster or cell type, splitting nuclei from control or late-stage AD samples where appropriate.

Chromatin *cis* co-accessibility network analysis. The correlation structure of chromatin accessibility data was analyzed using the R package cicero[30] (v1.3.4.7). Briefly, cicero quantifies 'co-accessibility' between pairs of genomic regions in a population of cells by correlating accessibility signals aggregated from several cells at a time, penalizing for the distance between regions using a graphical LASSO with a maximum interaction constraint of 500 kb. Importantly, before correlation and regularization, a bootstrap approach was used to generate metacells by aggregating 50 cells at a time using *k*-nearest neighbors, circumventing the sparsity of single-cell chromatin data. Finally, CCANs were identified through community detection. We applied this procedure in each major cell type as well as splitting each cell type into control and AD cells for CCAN analysis.

Analysis of gene-linked candidate *cis*-regulatory elements. We sought to further contextualize co-accessible chromatin regions by linking them to likely target genes using an accessibility–expression correlation strategy stratified by major cell type and disease status of each sample. First, we identified pairs of co-accessible peaks where one of the peaks overlaps a gene's promoter, which serves as a candidate target gene for that particular cCRE. We then computed the Pearson correlation between the expression of the candidate target gene from snRNA-seq with the chromatin accessibility of the linked cCRE from snATAC-seq, where expression and accessibility values have been averaged for all cells within a given cell population. This correlation analysis was performed iteratively across all promoter-cCRE co-accessible links identified separately in each major cell with regard to AD diagnosis status. Retaining links with the Pearson correlation coefficient in the 95th percentile and a P value ≤ 0.01, we defined gl-cCREs as genomic regions with a significant correlation to at least one target gene, and we defined cCRE-linked genes as genes with a significant correlation to at least one cCRE. We used NMF (v0.23.0) as implemented in the R NMF package[68] using $k = 30$ matrix factors on the gl-cCRE accessibility matrix averaged by each snATAC-seq cluster split by cells from control and AD samples, yielding 30 gl-cCRE modules. The NMF basis matrix ($\boldsymbol{W}$) was used to assign each gl-cCRE to its top associated module, and the NMF coefficient matrix ($\boldsymbol{H}$) was used to determine which cell cluster that each module was most associated to. To identify biological processes associated with these gl-cCRE modules, we used the enrichR[69,70] (v 3.0) package to query enriched GO terms for the set of target genes in each gl-cCRE module in the GO Biological Processes 2018, GO Cellular Component 2018 and GO Molecular Function 2018 databases.

TF regulatory network construction. Using snATAC-seq and snRNA-seq data in one cell type, we identified candidate TF regulatory target genes and used this information to construct cell-type-specific TF regulatory networks. We used the same set of TF binding motifs as those used in our single-cell TF motif enrichment analysis (JASPAR 2018 motifs). For a given TF, we defined candidate target genes as those with an accessible promoter containing the TF binding motif, or an accessible gl-cCRE linked to the target gene's promoter, allowing us to distinguish between TFs that regulate genes through promoter or enhancer binding events. We used this information to construct a directed TF regulatory network using the R package igraph (v1.2.6), where each vertex represents a TF or target gene, and each edge represents a promoter or linked-cCRE binding event, overlaying additional information onto the network such as DEG or AD GWAS gene status.

Estimating GWAS enrichment using cluster-specific accessible chromatin regions. To estimate heritability of a variety of complex traits, we used LDSC (v1.0.1)[43]. GWAS summary statistics were input to LDSC, which then computes the enrichment of heritability for an annotated set of SNPs conditioned on a baseline model to account for genomic features that influence heritability, and jointly modeling multiple annotations together. Sets of cluster-specific peaks were constructed by extending peaks upstream and downstream by 2,000 bp, identifying peaks that are accessible in 1% of all cells within each cluster and removing all peaks that are accessible in more than one other cell type. Cluster-specific peaks were formatted for LDSC using the make_annotation.py script, and LD scores were computed for each set using the ldsc.py script. Publicly available GWAS summary statistics were collected for AD[10,11], schizophrenia[46], frontotemporal dementia[44], progressive supranuclear palsy[45], multiple sclerosis[47], inflammatory bowel disease[48], height[49] and cholesterol[50]. Next, summary statistics were converted to hg38 coordinates using the UCSC liftover tool (v377) and formatted for LDSC using the munge_sumstats Python script. We followed the recommended guidelines for cell-type-specific partitioned heritability analysis, using HapMap3 SNPs and their provided hg38 baseline model (v2.2). The ldsc.py script was then used to compute cluster-specific enrichments of GWAS heritability, with Benjamini–Hochberg multiple-testing correction applied to enrichment P values.

Single-nucleus consensus weighted gene coexpression network analysis. We developed a coexpression network analysis approach to analyze single-cell data by integrating snRNA-seq and bulk-tissue RNA-seq datasets and called this approach single-nucleus consensus weighted gene coexpression network analysis (scWGCNA). scWGCNA is based on WGCNA, implemented using the WGCNA R package (v1.69)[53,54]. For scWGCNA, we used multiple transcriptomic datasets comprising our snRNA-seq data, snRNA-seq data from Mathys et al., bulk-tissue RNA-seq data from our UCI cohort and bulk-tissue RNA-seq data from the ROSMAP cohort[57]. First, we integrated our snRNA-seq dataset and the snRNA-seq dataset from Mathys et al. using the iNMF approach, and then constructed metacells in a fashion similar to that used in our CCAN analysis of chromatin accessibility data, in which we applied a bootstrapped aggregation process to single-nucleus transcriptomes. During metacell computation, we only pooled cells within the same cell type, and within the same AD diagnosis stage, to retain these metadata for scWGCNA. We then used a signed consensus WGCNA approach[58] within a given cell type, by calculating component-wise values for topological overlap for each dataset. First, bi-weighted mid-correlations were calculated for all pairs of genes, and then a signed similarity matrix was created. In the signed network, the similarity between genes reflects the sign of the correlation of their expression profiles. The signed similarity matrix was then raised to power β, varying between cell types, to emphasize strong correlations and reduce the emphasis of weak correlations on an exponential scale. The resulting adjacency matrix was then transformed into a topological overlap matrix. Modules were defined using specific module-cutting parameters, including a minimum module size of 100 genes, a deepSplit score of 4 and a threshold of correlation of 0.2. Modules with a correlation of greater than 0.8 were merged together. We used the first principal component of the module, called the module eigengene, to correlate with diagnosis and other variables. Hub genes were defined using intra-modular connectivity (kME) parameters of the WGCNA package. Gene-set enrichment analysis was done using EnrichR.

Analysis of regulatory targets of SREBP. We downloaded a dataset of ENCODE ChIP–seq validated TF target genes from EnrichR, containing regulatory targets of SREBP. Fisher's enrichment tests were performed with the R function fisher. test to test whether oligodendrocyte modules were significantly overrepresented with SREBP target genes, inferring which modules are regulated by SREBP. Module eigengenes were computed for the set of SREBP target genes, and the RNA expression as well as protein expression data from Inweb[71] and Biogrid[72] was also used to analyze SREBP targets during AD progression. A PPI network of SREBP target genes was constructed using SREBF1 ChIP–seq data from ENCODE and visualized using the STRING database[73], restricting the edges to known PPIs. In addition to bulk RNA-seq data, we used a proteomics dataset from our group's previous study[59] of 685 samples representing AD, asymptomatic AD and controls from the human PFC to interrogate the levels of SREBF1 target genes and target proteins in AD.

Statistics. All statistical methods and tests used in this paper are described in the figure legends, Methods, Supplementary Note or main text as appropriate.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All the multi-omics raw and processed data are available at https://www.synapse.org/#!Synapse:syn22079621/. Raw sequencing data have been deposited into the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE174367. Additionally, these data can be accessed through our web app at http://swaruplab.bio.uci.edu/singlenucleiAD. Source data are provided with this paper.

Code availability

The custom code used for this paper is available on GitHub[74] (https://doi.org/10.5281/zenodo.4681643) at https://github.com/swaruplab/Single-nuclei-epigenomic-and-transcriptomic-landscape-in-Alzheimer-disease/.

References

63. Sun, Y., Ip, P. & Chakrabartty, A. Simple elimination of background fluorescence in formalin-fixed human brain tissue for immunofluorescence microscopy. *J. Vis. Exp.* **127**, 56188 (2017).
64. Stuart, T., Srivastava, A., Lareau, C. & Satija, R. Multimodal single-cell chromatin analysis with Signac. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.09.373613 (2020).
65. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

66. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
67. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
68. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
69. Kuleshov, M. V. et al. Enrichr: a comprehensive gene-set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
70. Chen, E. Y. et al. Enrichr: interactive and collaborative *HTML5* gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
71. Lage, K. et al. A human phenome–interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).
72. Stark, C. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
73. Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
74. Morabito, S. & Swarup Lab. swaruplabUCI/single-nuclei-epigenomic-and-transcriptomic-landscape-in-Alzheimer-disease: publication. Zenodo https://doi.org/10.5281/zenodo.4681643 (2021).

Acknowledgements

Funding for this work was provided by UCI and UCI MIND start-up funds, a National Institute on Aging grant 1RF1AG071683, Adelson Medical Research Foundation funds and an American Federation of Aging Research young investigator award to V.S. The funders had no role in the study design, data collection and analysis, decision to publish or in the preparation of the manuscript. We thank the New York Genome Center for sequencing our bulk and single-nucleus RNA-seq libraries and UCI's Genomic High Throughput Facility for providing their facilities and sequencing our single-nucleus ATAC-seq libraries. In addition, we thank J. Atwood and UCI's Institute for Immunology Flow Cytometry Core for assisting us in FACS sorting. We also thank K. Green and F. LaFerla for generously providing their imaging facilities and M. Arreola, Y.-Q. Xue and I. Cobos for their technical advice. This work utilized the infrastructure for high-performance and high-throughput computing, research data storage and analysis, and scientific software tool integration built, operated and updated by the Research Cyberinfrastructure Center (RCIC) at the University of California, Irvine (UCI). For ROSMAP, study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by National Institute on Aging grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984 and U01AG46152, the Illinois Department of Public Health and the Translational Genomics Research Institute. The results presented here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (https://doi.org/10.7303/syn2580853).

Author contributions

S.M., E.M., N.M. and V.S. planned and directed all experiments and analyses and wrote the manuscript with assistance from all authors. N.M., E.M. and V.S. performed all dissections and sequencing experiments, and S.M., E.M. and V.S. performed all bioinformatic analyses. A.C.M., E.H., J.S., K.L., M.P.-R., E.M., N.M. and V.S. reviewed clinical information. E.M. performed RNAscope validation, and N.M., S.S. and A.C.M. performed immunofluorescence/immunohistochemistry.

Competing interests

The authors declare no competing interests.

Additional information

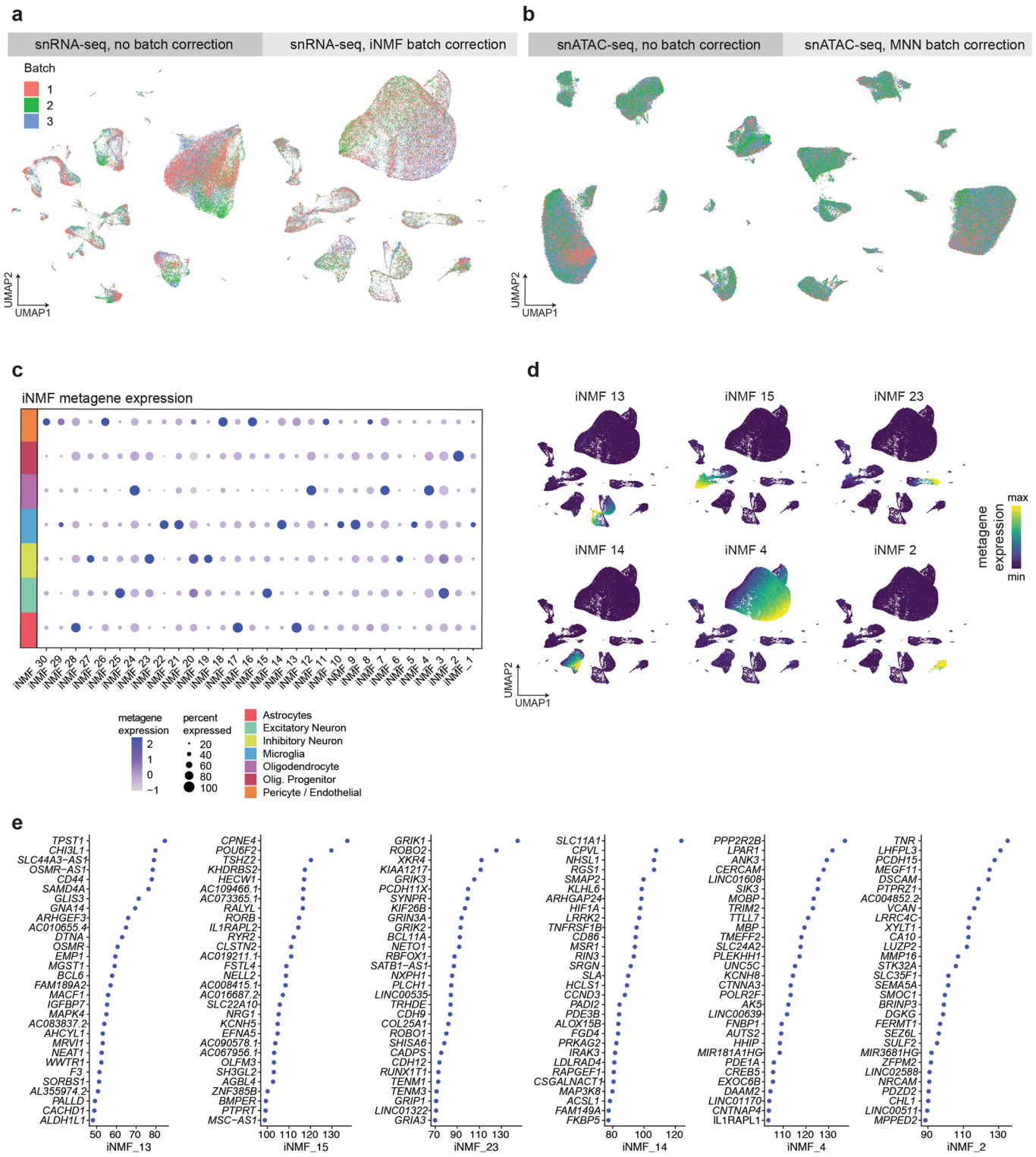
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00894-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00894-z.

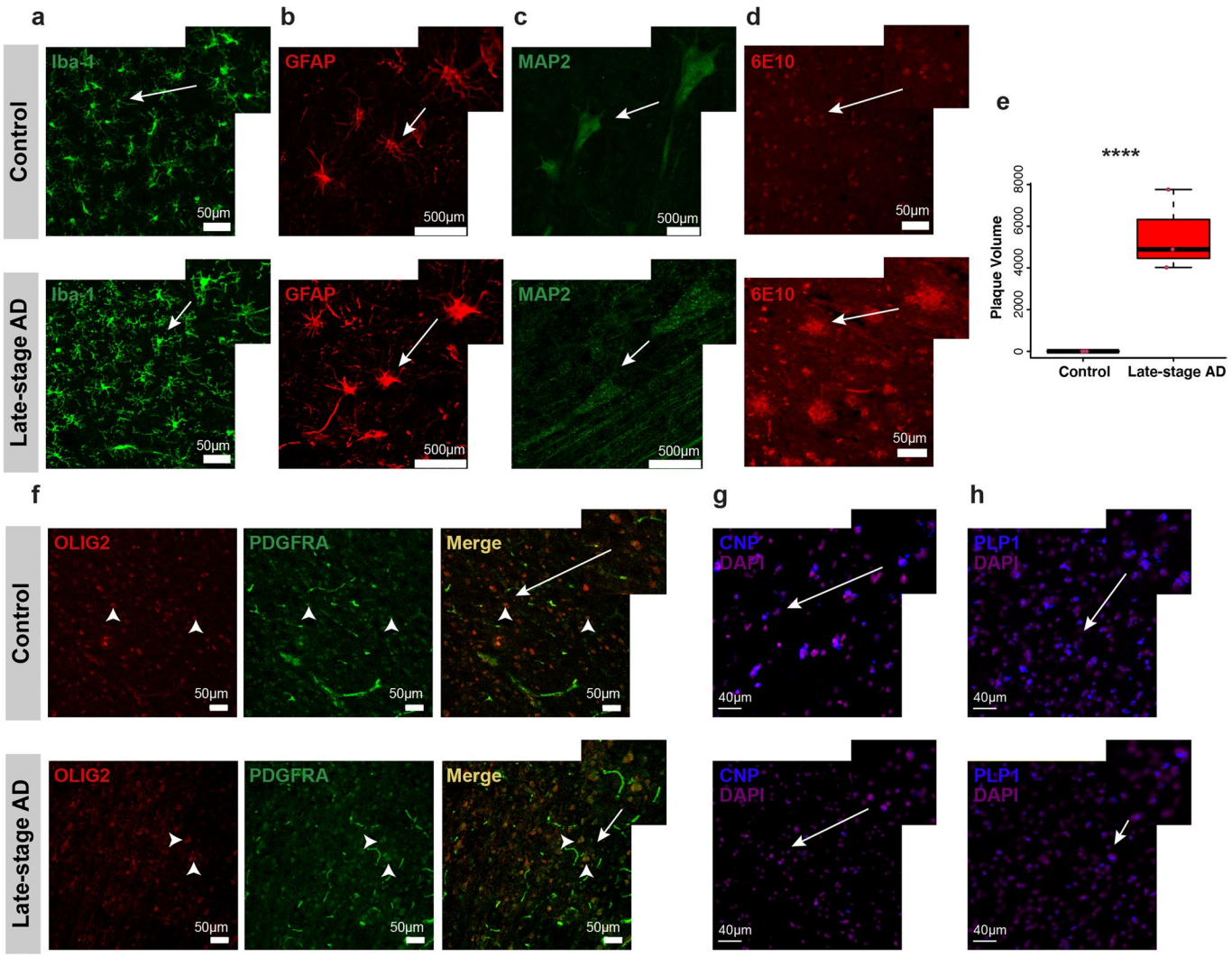
Correspondence and requests for materials should be addressed to V.S.

Peer review information *Nature Genetics* thanks Goncalo Castelo-Branco, Alexandra Grubman and Yin Shen for their contribution to the peer review of this work.

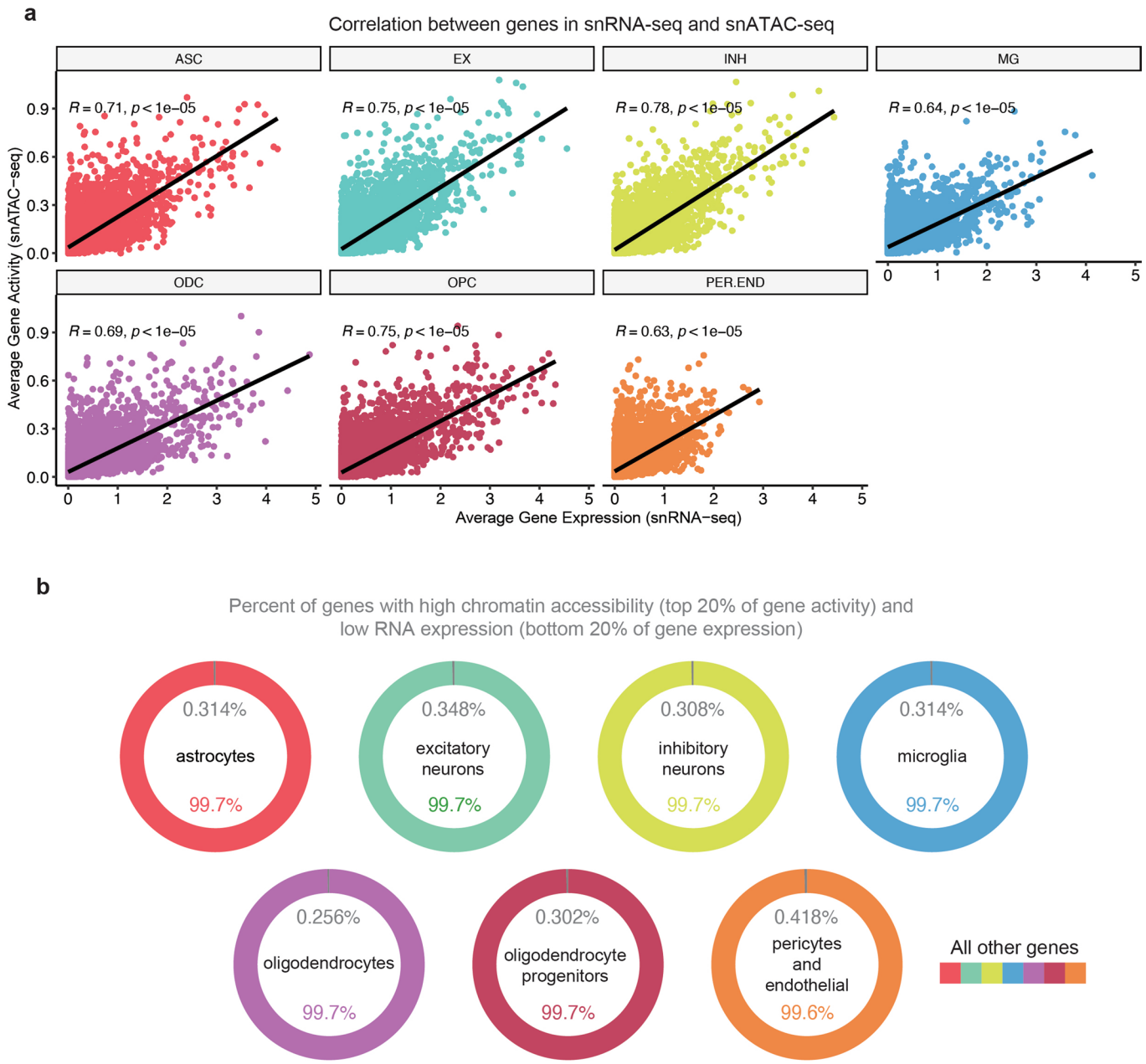
Reprints and permissions information is available at www.nature.com/reprints.



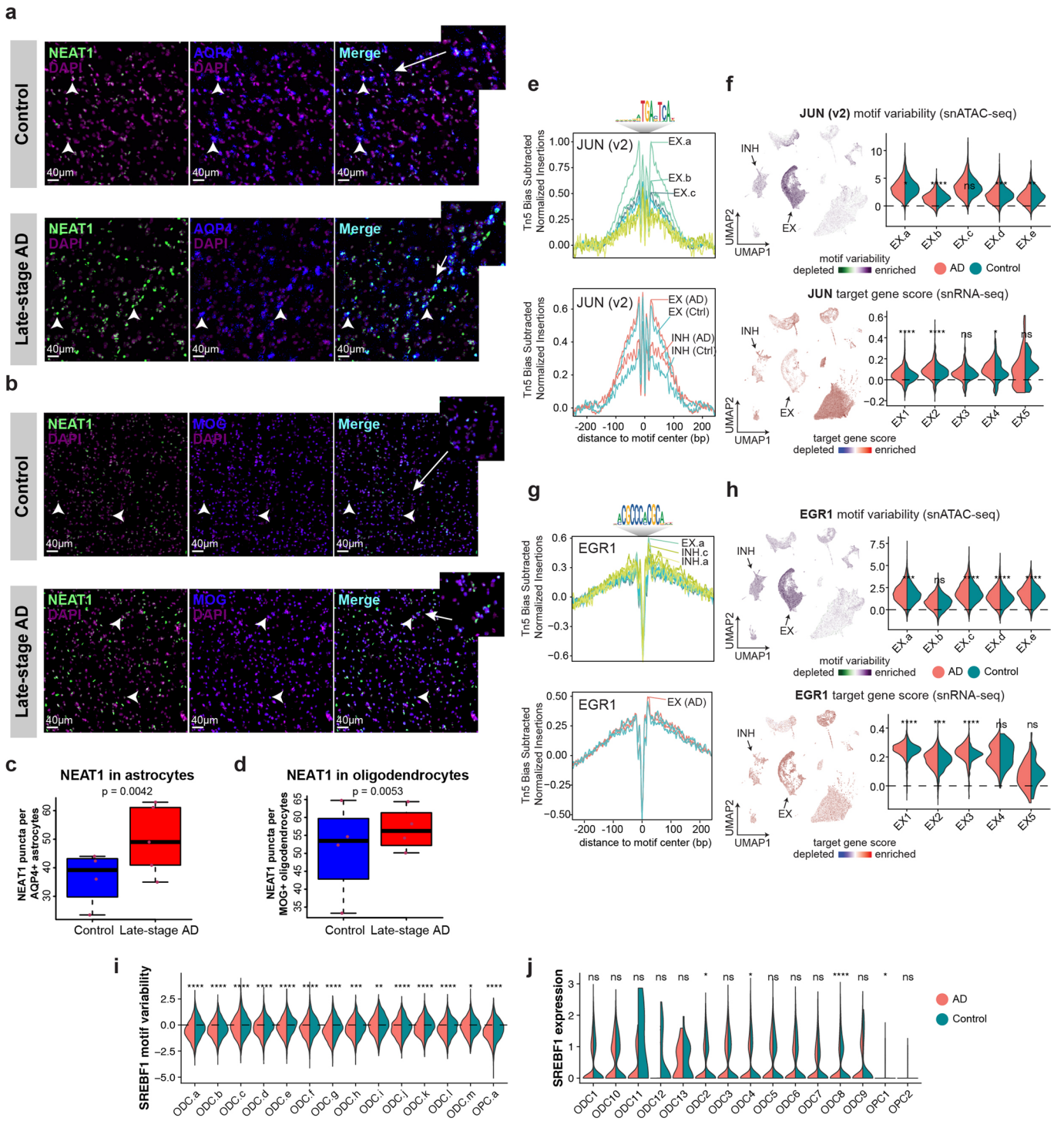
Extended Data Fig. 1 | Batch correction of snATAC-seq, snRNA-seq, and merged datasets. a, snRNA-seq UMAPs before (left) and after iNMF batch correction (right), colored by sequencing batch. **b**, snATAC-seq UMAPs before (left) and after MNN batch correction (right), colored by sequencing batch. **c**, Dot plot of iNMF metagene expression in each snRNA-seq cluster. **d**, snRNA-seq UMAPs colored by metagene expression of selected iNMF metagenes. **e**, Dot plots showing the iNMF loading for the top 30 genes for the same metagenes in **d**.



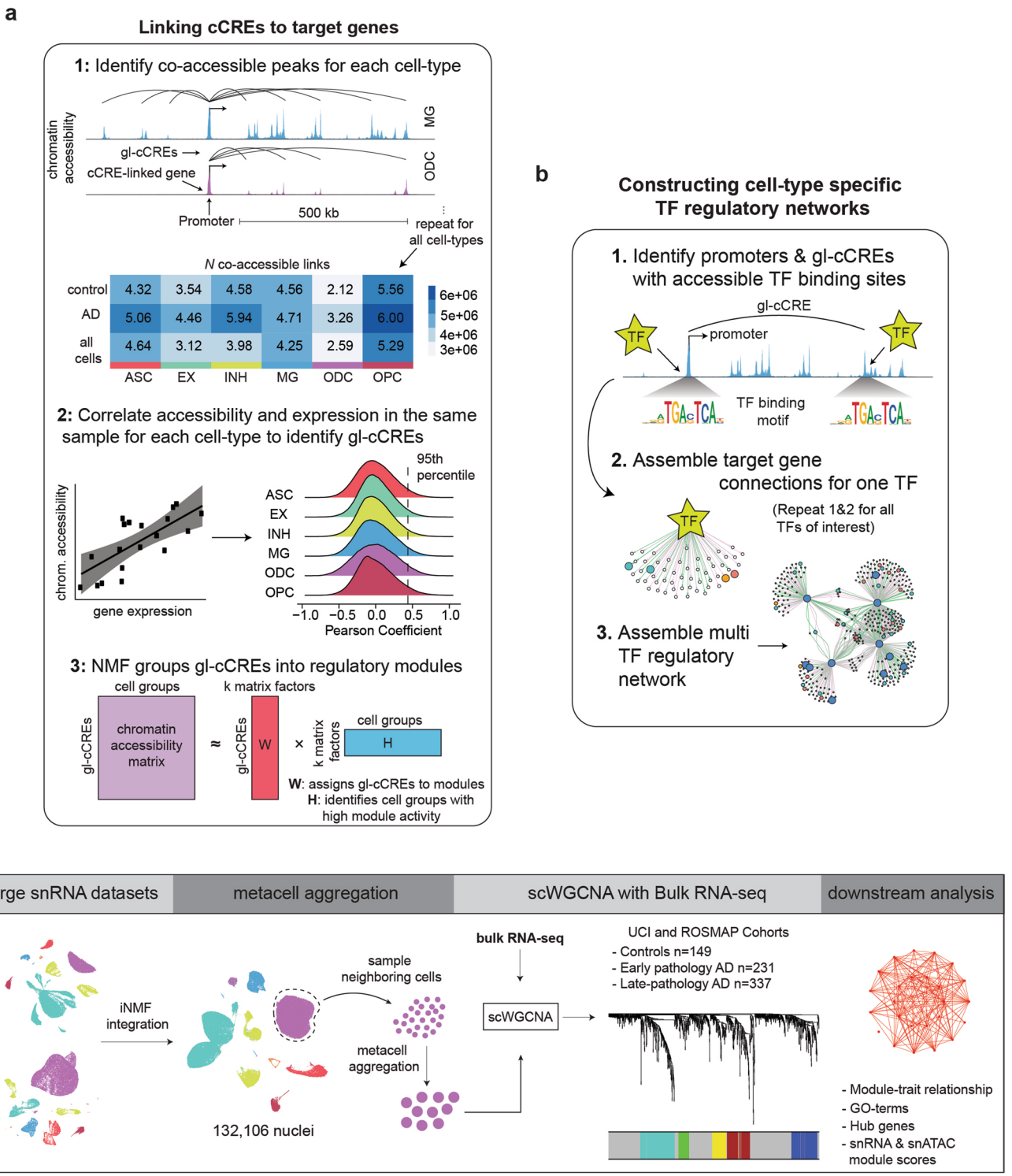
Extended Data Fig. 2 | Cell-type immunostaining and in situ hybridization. a–d, Representative immunofluorescence images from postmortem human brain tissue from control and late-stage AD cases for Iba-1 **(a)**, GFAP **(b)** MAP2 **(c)**, and 6E10 **(d)**. **e**, Quantification of 6E10-positive amyloid plaques. n = 3 cognitively healthy controls, 3 late-stage AD. Data is presented as the average of three different sections per sample. Linear mixed-effects model **** $p < 0.0001$. 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. **f**, Representative immunofluorescence images from postmortem human brain tissue from control and late-stage AD cases for OLIG2 with PDGFRA co-labeling. **g, h**, Representative RNAscope images from postmortem human brain tissue from control and late-stage AD cases for CNP **(g)** and PLP1 **(h)** with DAPI counterstain.



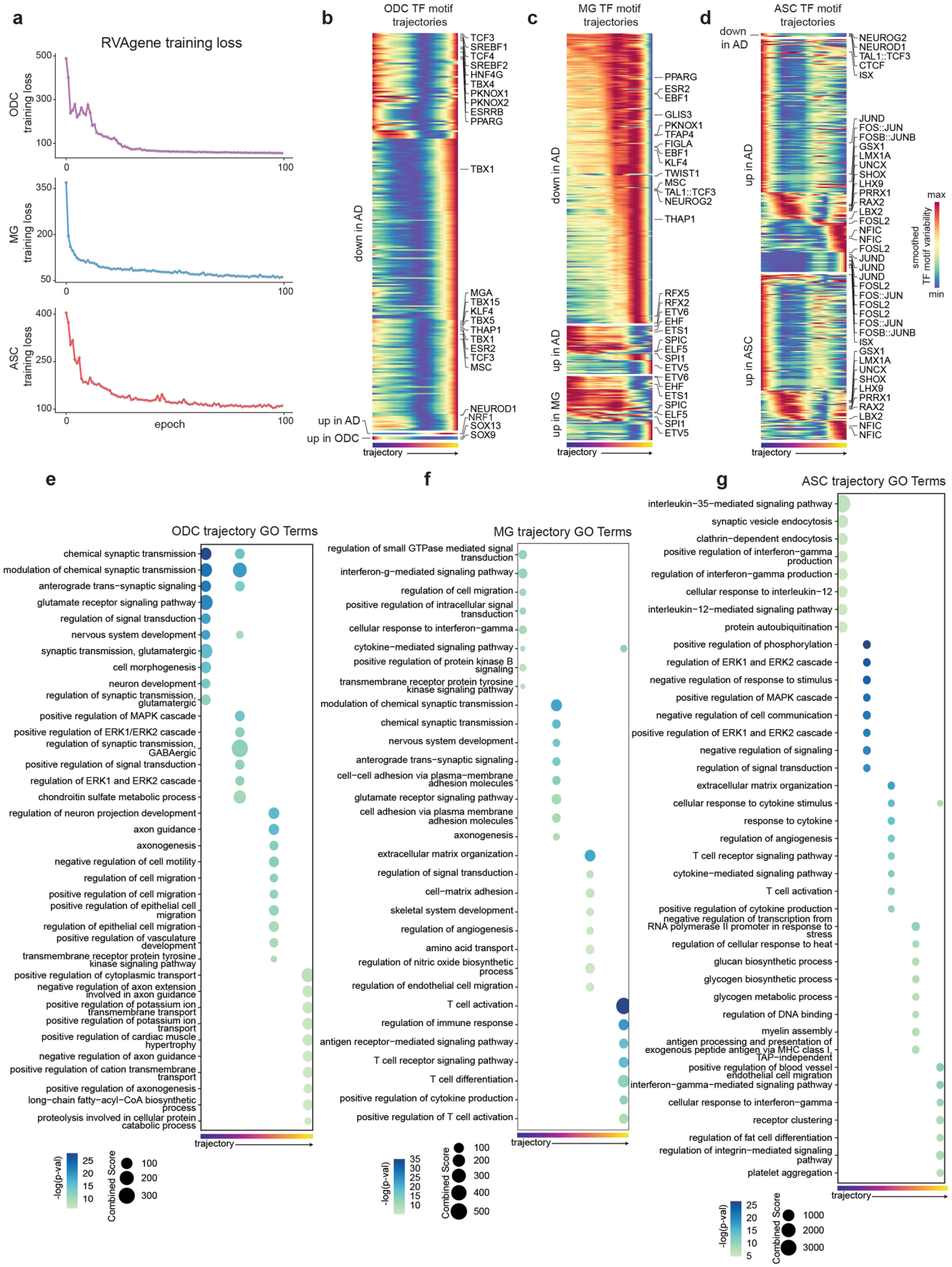
Extended Data Fig. 3 | Comparison of gene expression and gene activity. a, Scatter plot comparing average gene activity from snATAC-seq and average gene expression from snRNA-seq by each major cell-type, with accompanying Pearson correlation statistics and linear regression lines. **b**, Donut chart showing the percent of genes with high chromatin accessibility and low gene expression in grey for each major cell-type. High chromatin accessibility was defined as genes in the top 20% of gene activity, while low gene expression was defined as genes in the bottom 20% of gene expression. Percent of all other genes are colored by the cell-type.



Extended Data Fig. 4 | NEAT1 validation and neuronal TFs. a, b, Representative RNAscope images from postmortem human brain tissue for *NEAT1* and *AQP4* staining **(a)** and *NEAT1* and *MOG* staining **(b)** with DAPI nuclear counterstain. **c**, Boxplots showing quantification of *NEAT1* puncta per *AQP4* + astrocyte as in **a**. n = 4 cognitively healthy controls, 5 late-stage AD. **d**, Boxplots showing quantification of *NEAT1* puncta per *MOG* + oligodendrocyte as in **b**. n = 4 cognitively healthy controls, 4 late-stage AD. Data is represented as the mean of four equally sized regions per sample. Linear mixed-effects model **e**, Tn5 bias subtracted TF footprinting for JUN by snATAC-seq neuron cluster (top) and by AD diagnosis (bottom), with TF binding motif logo above and Tn5 bias insertions below. **f**, Left: Co-embedding UMAP colored by JUN motif variability (top) and JUN target gene score (bottom). Right: Violin plots of JUN motif variability (top) and JUN target gene score (bottom) in excitatory neuron clusters, split by diagnosis. Wilcoxon test (ns: $p > 0.05$, *: $p <= 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$, ****: $p <= 0.0001$). **g**, Tn5 bias subtracted TF footprinting for EGR1 by snATAC-seq neuron cluster (top) and by AD diagnosis (bottom), as in **e**. **h**, Left: Co-embedding UMAP colored by EGR1 motif variability (top) and EGR1 target gene score (bottom). Right: Violin plots of EGR1 motif variability (top) and *EGR1* target gene score (bottom) in excitatory neuron clusters, split by diagnosis, as in **f**. **i**, Violin plots of SREBF1 motif variability in oligodendrocyte snATAC-seq clusters, as in **f**. **j**, Violin plots of SREBF1 gene expression in oligodendrocyte snRNA-seq clusters, as in **i**. For boxplots, 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.

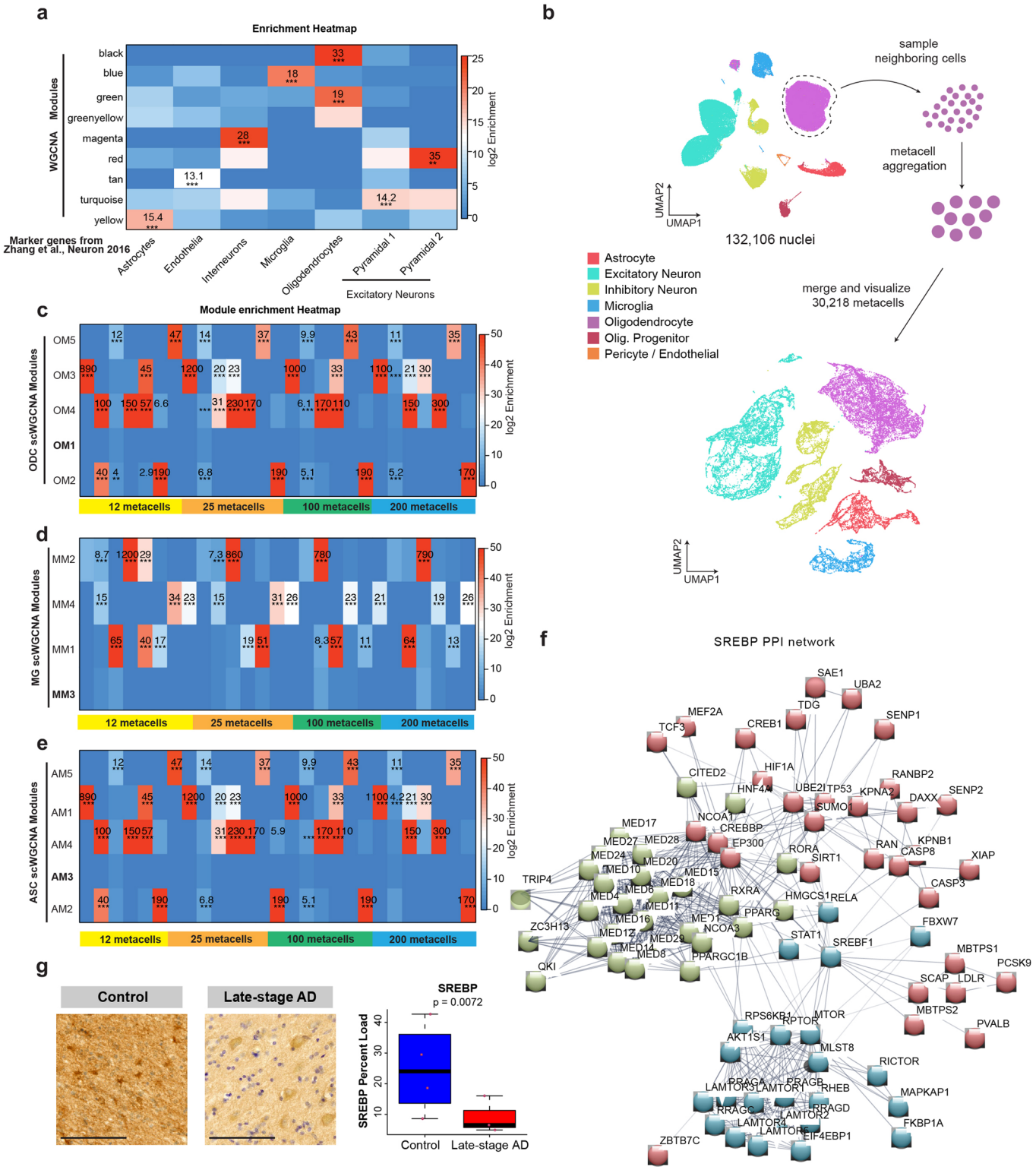


Extended Data Fig. 5 | Schematics of analyses. a, Schematic diagram linking cCREs to target genes and downstream analysis. First, we identify co-accessible chromatin peaks in each cell-type for control and late-stage AD. Second, we identify pairs of co-accessible peaks where one peak overlaps a gene promoter and correlate the expression of that gene with the chromatin accessibility of the other peak. Third, NMF is used to group gl-cCREs into functional modules. **b**, Schematic of construction of TF regulatory networks for each cell-type. **c**, Schematic representation of scWGCNA analysis, including iNMF integration with the Mathys *et al.* 2019 dataset, *metacell* aggregation, construction of co-expression networks, and downstream analysis of gene modules.

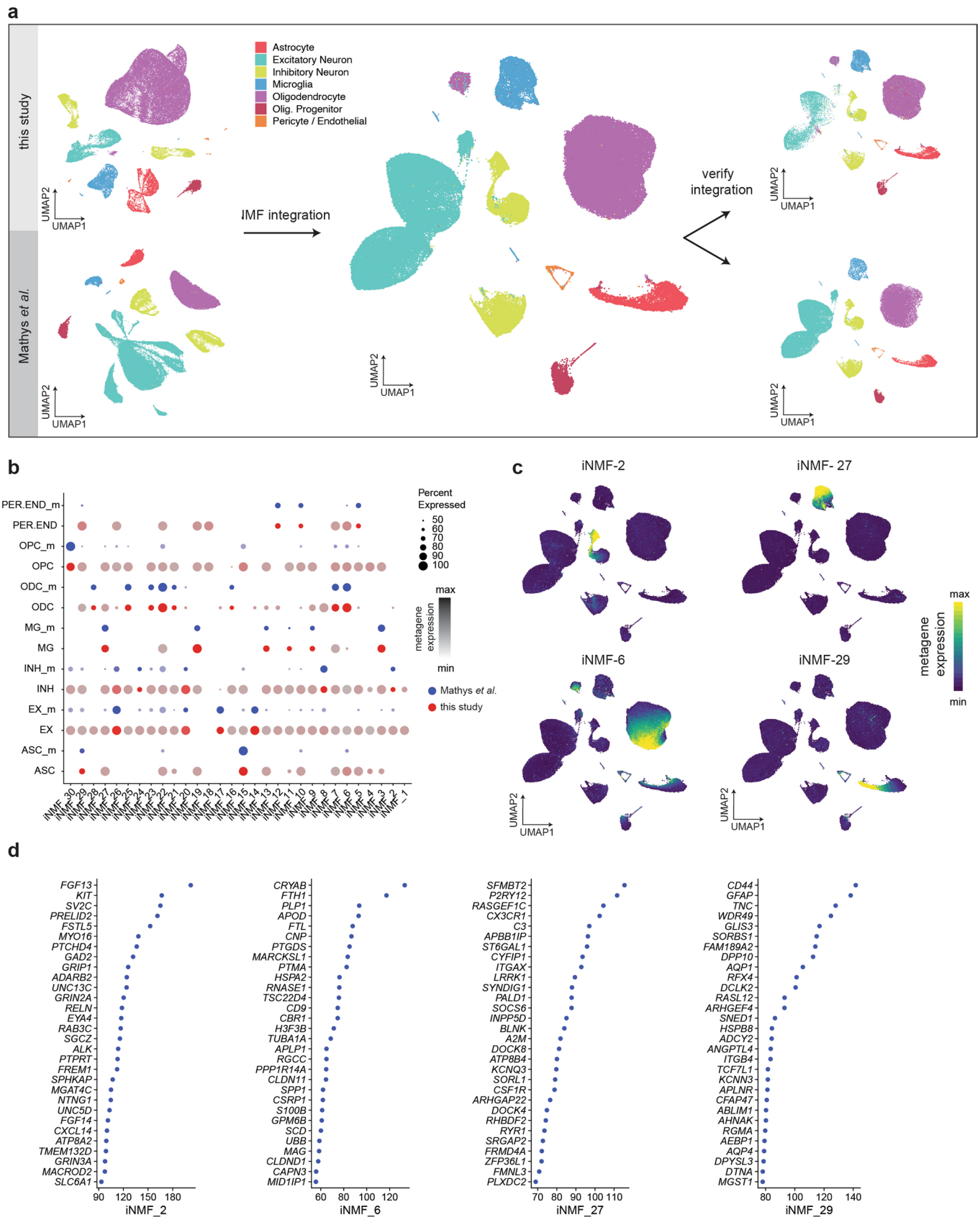


Extended Data Fig. 6 | See next page for caption.

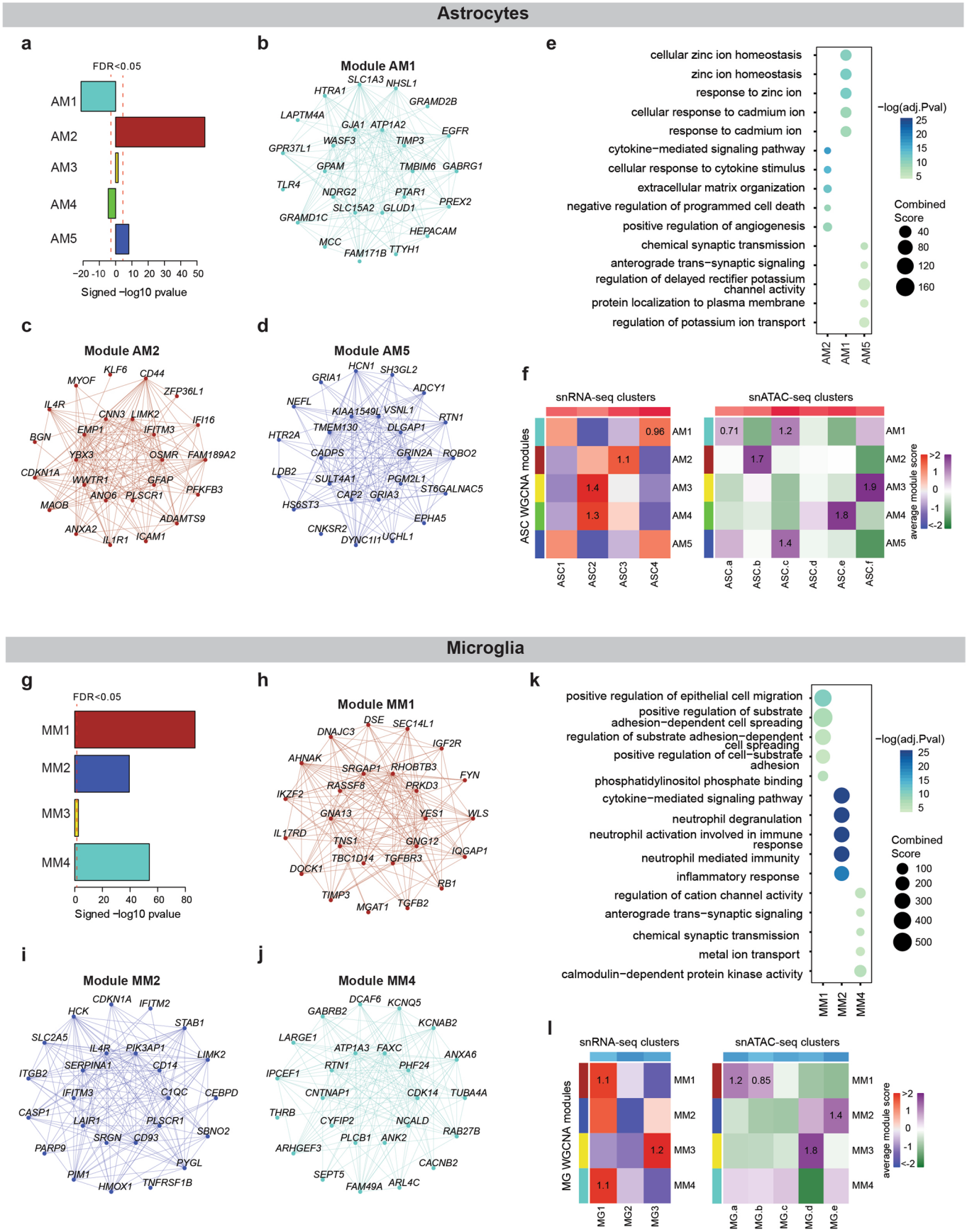
Extended Data Fig. 6 | Pseudotime trajectory analysis to identify dysregulated TFs and gene expression in glia. a, Line plot showing the RVAgene training loss at each epoch for oligodendrocyte (ODC), microglia (MG), and astrocyte (ASC) RVAE models. **b, c, d**, Heatmaps showing TF motif variability smoothed using loess regression and scaled to minimum and maximum values for TFs up- and down-regulated in AD as well as cell-type marker TFs along the oligodendrocyte trajectory **(b)**, microglia trajectory **(c)**, and astrocyte trajectory **(d)**. TFs are ordered by trajectory rank (point in trajectory where of 75% maximum value is reached). **e, f, g**, Dot plot showing the enrichR combined score for the top enriched GO terms in oligodendrocyte **(e)**, microglia **(f)**, and astrocyte **(g)** t-DEGs.



Extended Data Fig. 7 | Metacell aggregation and SREBP. a, Heatmap showing the enrichment of cell-type marker genes in standard WGCNA modules constructed from our snRNA-seq data. **b**, Schematic showing generation of 30,218 *metacells* from the integrated transcriptomic dataset of 132,106 nuclei from our snRNA-seq and Mathys *et al*. **c, d, e**, Heatmap showing enrichment of oligodendrocyte **(c)**, microglia **(d)**, and astrocyte **(e)** scWGCNA modules constructed with 12 metacells, 25 metacells, 100 metacells, and 200 metacells in the scWGCNA modules constructed with 50 metacells, as shown in Fig. 7 and Supplementary Fisg. 15 and 16. **f**, SREBP protein-protein interaction (PPI) network. Green circle denotes proteins involved in ribosome processing and transcription pathway, cyan circle for mTOR pathway, and red circle for lipid processing pathway. **g**, Left: Representative immunohistochemistry images from postmortem human brain tissue for SREBP with nuclear counterstain. Right: Quantification of SREBP staining. n = 4 pathological controls, 3 late-stage AD. Data is represented as the mean of four equally sized regions per sample. Scale bar represents 100 μm. Linear mixed-effects model ** $p < 0.01$. 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.

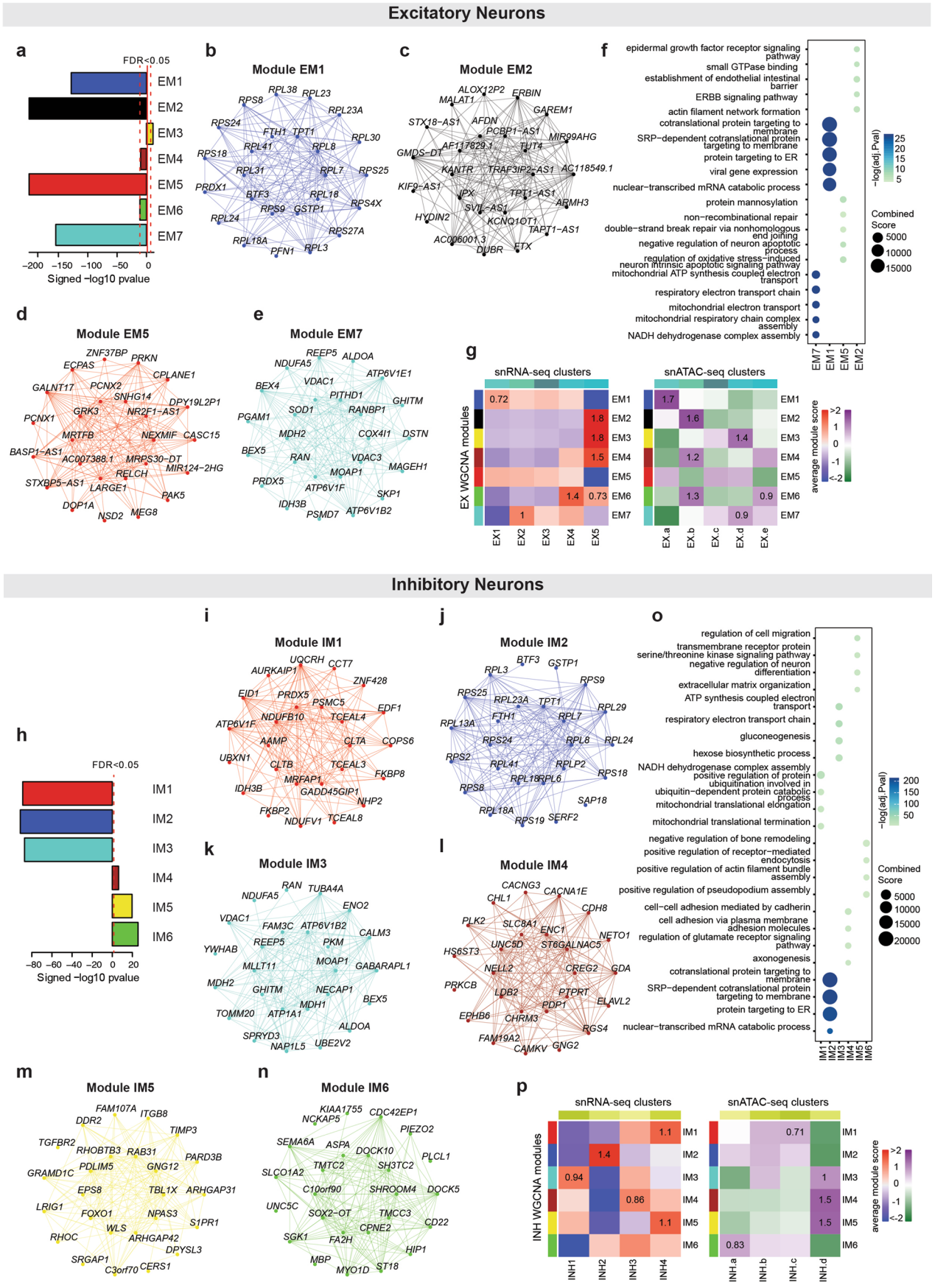


Extended Data Fig. 8 | iNMF integration of snRNA-seq with Mathys et al. snRNA-seq. a, Schematic representation of iNMF integration of snRNA-seq with Mathys et al. snRNA-seq. UMAP plots are colored by cell-type assignments. **b**, Dot plot of iNMF metagene expression in each cell-type, split by dataset of origin. **c**, UMAP plots of the integrated dataset colored by selected iNMF metagenes. **d**, Dot plots showing the iNMF loading for the top 30 genes for the same metagenes in **c**.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | scWGCNA in microglia and astrocytes. a, Signed correlation of astrocyte modules to AD diagnosis. **b, c, d**, Co-expression plots for modules AM1 **(b)**, AM2 **(c)**, and AM5 **(d)**. **e**, GO term enrichment of astrocyte modules. **f**, Heatmaps showing row-normalized Seurat module scores of astrocyte modules in snRNA-seq (left) and snATAC-seq (right) astrocyte clusters. **g**, Signed correlation of microglia co-expression modules with AD diagnosis. **h, i, j**, Co-expression plots for modules MM1 **(h)**, MM2 **(i)**, and MM4 **(j)**. **k**, GO term enrichment of microglia modules. **l**, Heatmaps showing row-normalized Seurat module scores of microglia modules in snRNA-seq (left) and snATAC-seq (right) microglia clusters.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | scWGCNA in neurons. a, Signed correlation of excitatory neuron modules to AD diagnosis. **b, c, d, e**, Co-expression plots for modules EM1 **(b)**, EM2 **(c)**, EM5 **(d)**, and EM7 **(e)**. **f**, GO term enrichment of excitatory neuron modules. **g**, Heatmaps showing row-normalized Seurat module scores of excitatory neuron modules in snRNA-seq (left) and snATAC-seq (right) excitatory neuron clusters. **h**, Signed correlation of inhibitory neuron modules to AD diagnosis. **i, j, k, l, m, n**, Co-expression plots for modules IM1 **(i)**, IM2 **(j)**, IM3 **(k)**, IM4 **(l)**, IM5 **(m)**, and IM6 **(n)**. **o**, GO term enrichment of inhibitory neuron modules. **p**, Heatmaps showing row-normalized Seurat module scores of inhibitory neuron modules in snRNA-seq (left) and snATAC-seq (right) inhibitory neuron clusters.

nature research

Corresponding author(s): Vivek Swarup

Last updated by author(s): 4/16/2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	We used publicly available softwares for all analysis and the data analysis was performed in R or python. These R and python packages are listed in the methods section with their appropriate citations and/or URLs. Custom codes used in the analysis can be accessed using this github link: https://github.com/swaruplab/Single-nuclei-epigenomic-and-transcriptomic-landscape-in-Alzheimer-disease

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the multi-omics raw and processed data are available here: https://www.synapse.org/#!Synapse:syn22079621/. Additionally, the data can be accessed through our online web app: https://swaruplab.bio.uci.edu/singlenucleiAD/ . Mathys et al., Nature 2019 data was accessed from https://www.synapse.org/#!Synapse:syn18485175 and ROSMAP bulk tissue RNA-seq data was accessed from https://www.synapse.org/#!Synapse:syn8456629. The proteomics data from AMP-AD was accessed from https://swaruplab.bio.uci.edu/ADpNet

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. Sample size for snATAC-seq (n=12 late-stage AD; n=8 control) and snRNA-seq (n=11 late-stage AD; n=7 control) was based on current standard in the field as defined by comparable studies in the field of neurology and Alzheimer's disease (PMIDs: 31768052, 30747918, 31932797) Sample size for RNAscope and immunohistochemistry experiments (n = 3 cognitively healthy controls, 5 late-stage AD) was considered sufficient to ensure reproducibility, and also consistent with comparable studies (PMID 31042697, 33795864)
Data exclusions	For snRNA-seq and snATAC-seq experiments, we excluded samples and nuclei based on quality control procedures, and these exclusion criteria were not pre-established. Details of our quality control procedures are provided in the methods section of the manuscript.
Replication	We did not perform a traditional replication analysis. We did however use a strategy to understand and validate the robustness of the findings - we performed correlation analysis between the samples as well as published snRNA-seq data (PMID: 31042697) and found great concordance. In addition, we performed the experiments in multiple batches, and differences between batches were negligible and were accounted for all downstream analysis.
Randomization	Samples were randomized by case and control status during RNA isolation or library preparation. For RNAscope and immunohistochemistry randomization was not possible as all the samples were processed in the same day.
Blinding	No blinding was performed during data collection or analysis. For RNAscope and immunohistochemistry studies, we did not have the personnel resources to blind this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Mouse anti-beta-Amyloid (Biolegend, 803001, Clone: 6E10) Rabbit anti-Iba1 (Wako, 019-19741) Rabbit anti-MAP2 (Abcam, ab32454) Mouse anti-GFAP (Sigma, G3893, Clone: G-A-5) Goat anti-PDGFRA (R&D Systems, AF-307) Rabbit anti-OLIG2 (Abcam, ab109186) Rabbit anti-SREBP (Novus Biologicals, NB100-2215)
Validation	Mouse anti-beta-Amyloid (Biolegend, 803001, Clone: 6E10) The manufacture's website shows validation that the antibody is specific for human Amyloid-beta protein and cites several studies in which the antibody was used for immunohistochemistry. Rabbit anti-Iba1 (Wako, 019-19741) The manufacture's website shows validation that the antibody is specific for mouse, rat and human Iba-1 protein and cites several studies in which the antibody was used for immunohistochemistry.

Rabbit anti-MAP2 (Abcam, ab32454)
The manufacture's website shows validation that the antibody is specific for mouse, rat and human MAP2 protein and cites several studies in which the antibody was used for immunohistochemistry.

Mouse anti-GFAP (Sigma, G3893, Clone: G-A-5)
The manufacture's website shows validation that the antibody is specific for pig, rat and human GFAP protein and cites several studies in which the antibody was used for immunohistochemistry.

Goat anti-PDGFRA (R&D Systems, AF-307)
The manufacture's website shows validation that the antibody is specific for human PDGFRA protein and cites several studies in which the antibody was used for immunohistochemistry.

Rabbit anti-OLIG2 (Abcam, ab109186)
The manufacture's website shows validation that the antibody is specific for mouse, rat and human OLIG2 protein and cites several studies in which the antibody was used for immunohistochemistry.

Rabbit anti-SREBP (Novus Biologicals, NB100-2215)
The manufacture's website shows validation that the antibody is specific for several species including mouse and human SREBP protein and cites several studies in which the antibody was used for immunohistochemistry.

Human research participants

Policy information about studies involving human research participants

Population characteristics	We have adjusted our analysis for age and gender. For details on human samples used in this study, please see Supplemental Table 1-2.
Recruitment	No donors were recruited, all participants gave prospective pro-mortem written consent for their brains to be banked and used for research
Ethics oversight	Human prefrontal cortex brain samples were obtained from UCI MIND's Alzheimer's Disease Research Center (ADRC) tissue repository and under UCI's Institutional Review Board (IRB, #HS-2018-4515). Because the project evaluates post-mortem brain samples, this activity is not considered to meet federal definitions under the jurisdiction of an IRB (IRB exemption to the project was approved), and falls outside the purview of the Human Research Protection Office (HRPO)

Note that full information on the approval of the study protocol must also be provided in the manuscript.